

Independent Evaluation of the Girls' Education Challenge Phase II - Evaluation Study 3: Aggregate impact of GEC-T projects between baseline and midline

Final Report



Client Name: Foreign, Commonwealth and Development Office (FCDO)

Project Name: Independent Evaluation of the GEC Phase II

Contract Number: PO 10019

Partners:

- Fab Inc
- Research and Equitable Access and Learning (REAL) Centre at the University of Cambridge

Authors: Poli, Florian; Ogando Portela, María José; Atherton, Paul; Kim, Youngjin; Koutecky, Tomáš; Tangpornpaiboon, Sirin.

September 2022

Tetra Tech International Development Europe Ltd

The Malthouse 1 Northfield Road Reading Berkshire RG1 8AH United Kingdom
T (+44) (0) 1189 566 066 F (+44) (0) 1189 576 066 www.tetratecheurope.com
Registered Office: 1 Northfield Road Reading Berkshire RG1 8AH United Kingdom
Registered in England No. 3799145 Vat Number: GB 724 5309 45

This document has been approved for submission by Tetra Tech International Development Project Director, based on a review of satisfactory adherence to our policies on:

- Quality management;
- HSSE and risk management;
- Financial management and Value for Money (VfM);
- Personnel recruitment and management;
- Performance Management and Monitoring and Evaluation (M&E)

Simon Griffiths

Project Director

Signature:



Disclaimer

This report is provided on the basis that it is for the use of the Foreign Commonwealth & Development Office (FCDO) only. Tetra Tech International Development will not be bound to discuss, explain or reply to queries raised by any agency other than the intended recipients of this report. Tetra Tech International Development disclaims all liability to any third party who may place reliance on this report and therefore does not assume responsibility for any loss or damage suffered by any such third party in reliance thereon.

Table of Contents

Acronyms.....	iv
Project Acronyms.....	vi
Executive Summary.....	vii
1. Introduction.....	1
2. Methodology.....	3
3. Changes in Girls' Learning.....	19
4. Changes in Girls' Transition Outcomes.....	36
5. Conclusions.....	40
6. Recommendations.....	43
7. Bibliography for Learning Standards and Benchmarks.....	45

List of Tables

Table 1: Sample size for Study 3 for Group A.....	6
Table 2: Learning sample size for Study 3 for Group B.....	6
Table 3: Comparison of means, treatment and comparison.....	7
Table 4: Attrition level by project.....	8
Table 5: Learning assessments subtasks.....	11
Table 6: Definition of successful and unsuccessful transition.....	15
Table 7: Learning improvements of the average GEC-T girl over and above comparison (beneficiary population weights).....	21
Table 8: Literacy improvements of GEC-T girls over and above comparison girls - % correct scores.....	21
Table 9: Numeracy improvements of GEC-T girls over and above comparison girls - % correct scores.....	22
Table 10: Average learning improvements of GEC-T projects over and above comparison groups (project-equal weights).....	23
Table 11: Change in the prevalence of zero scores over and above the comparison group.....	24
Table 12: Difference-in-difference by subgroup at baseline.....	27
Table 13: Learning progress of GEC-T girls between baseline and midline.....	30
Table 14: Change in intermediate outcomes of GEC-T girls.....	32
Table 15: Correlation between positive improvement in intermediate outcomes and learning change.....	33
Table 16: Transition outcomes of in-school girls, by key subgroups and treatment girls.....	37
Table 17: Transition outcomes of in-school girls, comparing between re-contacted girls only and accounting for attrition.....	39
Table 18: Transition outcomes of out-of-school girls.....	39

List of Figures

Figure 1: Projects included in Study 3.....	3
Figure 2: Comparison of sample compositions.....	5
Figure 3: Literacy progress of GEC-T girls, by age at baseline (% correct score).....	19
Figure 4: Numeracy progress of GEC-T girls, by age at baseline.....	20
Figure 5: Share of GEC-T girls scoring zero in literacy, at baseline and midline.....	24
Figure 6: Share of GEC-T girls scoring zero in numeracy, at baseline and midline.....	25
Figure 7: Average words-per-minute read correctly by age, at midline (words-per-minute capped at 100, project-equal weights, cross-sectional sample, treatment group).....	26
Figure 8: Share of girls able to read and understand what they read by age, at midline (project-equal weights, cross-sectional sample, treatment group).....	27

List of Boxes

Box 1: Key learning findings	19
Box 2: How to interpret percentage correct scores and percentage-point increases.....	20
Box 3: On the use of international standards and benchmarks for literacy and numeracy	26
Box 4: Key findings	36

Acronyms

ALP	Adult Learning Programme
BL	Baseline
CBE	Community Based Education
CS	Cross-sectional
DIBELS	Dynamic Indicator of Basic Early Literacy Skills
DID	Difference in differences
DRC	Democratic Republic of the Congo
EDT	Education Development Trust
EE	External Evaluators
EGMA	Early Grade Mathematics Assessment
EGRA	Early Grade Reading Assessment
ESWG	Evaluation Studies Working Group
FCDO	Foreign, Commonwealth and Development Office
FM	Fund Manager
GEC	Girls' Education Challenge
GEC II	Girls' Education Challenge Phase II
GEC-T	Girls' Education Challenge – Transition
GS	Girls Survey
GWD	Girls with Disabilities
HHS	Household Survey
IE	Independent Evaluation
IO	Intermediate Outcome
IP	Implementing Partner
ISG	In-School Girls
LA	Learning Assessment
LNGB	Leave No Girl Behind
MEL	Monitoring, Evaluation and Learning
M&E	Monitoring and Evaluation

Independent Evaluation of the Girls' Education Challenge Phase II – Aggregate Impact of GEC-T Projects Between Baseline and Midline Study – Final Report

ML	Midline
OLS	Ordinary Least Squares
OOS	Out-of-School
OPM	Oxford Policy Management
OSG	Out-of-School Girls
PCG	Primary Caregiver
PISA	Program for International Student Assessment
PwC	PricewaterhouseCoopers
RTI	Research Triangle Institute
RQ	Research Question
SeGMA	Secondary Grade Mathematics Assessment
SeGRA	Secondary Grade Reading Assessment
TIMSS	Trends in International Mathematics and Science Study
ToC	Theory of Change
ToR	Terms of Reference
TVET	Technical and Vocational Education and Training
UNESCO	The United Nations Educational, Scientific and Cultural Organization
USA	United States of America
WPM	Words per Minute

Project Acronyms

Project name	Acronyms used in report	Project location(s)
Aga Khan Foundation	AKF	Afghanistan
Avanti	Avanti	Kenya
Bangladesh Rural Advancement Committee	BRAC	Afghanistan
Camfed International	Camfed International	Tanzania, Zimbabwe, Zambia
Camfed Tanzania	Camfed (ex-BRAC)	Tanzania
CARE International	CARE	Somalia
ChildHope UK	ChildHope	Ethiopia
Cheshire Services Uganda	CSU	Uganda
Discovery Learning Alliance	DLA (Discovery)	Nigeria, Ghana, Kenya
Education Development Trust	EDT	Kenya
Health Poverty Action	HPA	Rwanda
I Choose Life-Africa	ICL	Kenya
Leonard Cheshire	LC	Kenya
Link Community Development	LCD (Link)	Ethiopia
Mercy Corps	MC Nepal/ MC Nigeria	Nepal and Nigeria
Opportunity International UK	Opportunity	Uganda
Promoting Equality in African Schools	PEAS	Uganda
Plan International	Plan	Sierra Leone
Relief International	Relief	Somalia
Save the Children (DRC)	STC DRC	DRC
Save the Children (Mozambique)	STC MOZ	Mozambique
Varkey Foundation	Varkey	Ghana
Viva	Viva	Uganda
Voluntary Service Overseas	VSO	Nepal
World Vision	WV	Zimbabwe
World University Service of Canada	WUSC	Kenya

Executive Summary

Introduction

This study is conducted as part of the independent evaluation (IE) of the Girls' Education Challenge Phase II (GEC II), an eight-year (2017-2025) programme supported by the UK Foreign, Commonwealth and Development Office (FCDO) that aims to improve the learning opportunities and outcomes of over one million of the world's most marginalised girls.

GEC Phase II is delivered through two funding windows: (1) GEC Transitions (GEC-T) Window, which supports marginalised girls participating in 27 GEC Phase I projects across 15 countries¹ in transitioning to the next stage of their education; and (2) Leave No Girl Behind (LNGB) Window, which supports 14 projects in 10 countries working with highly marginalised, adolescent girls who have never attended or have already dropped out of school.

This study focuses on the **GEC-T Window**, to enable the FCDO and the GEC II Fund Manager (FM) to respond to the **accountability objective** of whether, and to what extent the GEC-T is associated with changes in girls' learning and transition outcomes, as well as to provide **portfolio-level benchmarks for future IE studies** on what magnitude of impact the GEC-T has achieved, which subgroups were most affected by projects' interventions and which intermediate outcomes are correlated with improved learning.

Study 3 relies exclusively on the baseline and midline quantitative data² collected by the GEC-T projects' external evaluators (EE) and estimates the changes in learning and transition occurring before the disruptions caused by Covid-19.

The **study's objectives** are:

- **To create a portfolio-level dataset** of the quantitative baseline and midline data collected by GEC-T projects' external evaluations including their learning assessments, girl surveys, household surveys and primary caregiver surveys. This dataset will be used for research, further analysis and future studies; and
- **To quantify the aggregate changes** in girls' learning and transition outcomes associated with the GEC-T projects' interventions between their external baseline and midline evaluations.

Specifically, the study aims to answer the following **research questions**:

1. How did girls' learning levels change between baseline and midline?

Did GEC-T girls improve more than non-GEC girls? Are the observed changes different for literacy and numeracy?

2. Who has learned the most (and least), and how much did they learn?

How did changes in learning outcomes vary across contexts: by girls' individual characteristics? By household-level characteristics? By school status?

3. Why did some girls learn more than others?

What are the factors (including the project's intermediate outcomes) that are associated with learning across the GEC-T portfolio?

4. Did girls successfully transition between baseline and midline?

Who has not successfully transitioned in school?

Methodology

- This study is solely based on projects' quantitative data collected as part of project external evaluations (i.e., learning assessments, girl surveys (GS), household surveys (HHS), and primary caregiver surveys (PCG) data) for both baseline and midline.

¹ These countries are Afghanistan, Democratic Republic of the Congo (DRC), Ethiopia, Ghana, Kenya, Mozambique, Nepal, Nigeria, Rwanda, Sierra Leone, Somalia, Tanzania, Uganda, Zambia, and Zimbabwe.

² Endline projects are not included in this study because the adaptations to endline data collection activities caused by Covid-19 resulted in many project external evaluations not gathering learning data. Therefore, conducting portfolio-level analysis of the aggregate impact projects had on learning including endline data was not feasible.

- The GEC-T supports girls and young women in 15 countries through 27 projects³. Out of 27 GEC-T projects, seven were excluded from the study, four for lacking midline learning data, and three for not having a valid (i.e., conclusive) baseline data that could be compared with midline data. Therefore, **20 GEC-T projects have been used in the learning and transition analysis**, of which two (Relief International Somalia and World University Service of Canada (WUSC) Kenya) do not have a comparison group but are still included in the study.
- The learning analysis has been conducted twice: (1) on the cross-sectional sample (i.e., full sample of baseline and midline girls); and (2) on the panel sample (i.e., sample of girls followed over time). Due to a lack of common girls' unique IDs between the rounds, two projects are excluded from panel analysis (for learning and transition outcomes). As a result, 18 projects are included for panel learning analysis. In addition, one more project is excluded from the transition analysis due to inadequate matching between the baseline and midline using girls' unique IDs. As a result, 17 projects are included for the transition analysis. The **full sample at the portfolio level consists of 63,342 girls, including 35,047 recontacted girls**. For the learning analysis, the sample consists of 57,987 girls, including 29,890 recontacted girls.
- We analysed the difference between the baseline to midline progress achieved among the treatment group compared to that achieved among the comparison group, also called difference-in-difference. It is equal to the **progress achieved by girls from the treatment group over and above the learning progress achieved by girls from the comparison group**. The analysis suggests that although some level of attrition and replacement biases exist at the portfolio level, the characteristics of lost, recontacted and replaced girls are similar across the treatment and comparison groups. As a result, attrition and replacement biases do not cause concern regarding the difference-in-difference estimates for learning, while the panel and cross-sectional analysis generate broadly similar results.

Creation of the portfolio-level baseline-midline dataset

- The study created a **portfolio level dataset of quantitative baseline and midline data collected by projects' external evaluations**. Data from four instruments: (1) EGRA, EGMA, SeGRA and /or SeGMA Learning Assessments (LA); (2) Girl Surveys (GS); (3) Household Surveys (HHS); and (4) Primary Caregiver Surveys (PCG). The dataset contains nearly all variables included in the Fund Manager's questionnaire templates (that were cleaned and homogenised).
- This dataset will be **used for future evaluation or research studies** involving GEC-T projects. For example, it can be used to draw summary statistics on girls with disabilities based on the type of disability reported in Study 4 on Disability. In addition, the dataset and an accompanying codebook have been produced and made available to the FCDO to allow researchers, within the FCDO or the FM, to analyse this data. This dataset is of a publishable standard allowing open access to other researchers beyond the GEC should the FCDO wish to publish it.

RQ1. How did girls' learning levels change between baseline and midline?

To answer this question, we first calculated the *first-difference*, equal to the average change in the learning scores of girls in the treatment group between baseline and midline. This provides an estimate of whether and to what extent GEC-T girls have learned between baseline and midline. We then compared it to the first-difference estimated on the comparison group, by calculating the difference between the average progress in the treatment group and the average progress in the comparison group (this is the *difference-in-difference* estimator). In other words, the difference-in-difference is equal to the learning progress achieved by girls from the treatment group *over and above* the learning progress achieved by girls from the comparison group.

- The literacy and numeracy levels of GEC-T girls significantly improved between baseline and midline. Improvements are observed across all subtasks of the EGRA, EGMA, SeGRA and SeGMA, and for all age categories. In other words, **girls' learning improved across the GEC-T portfolio between baseline and midline, across a range of literacy and numeracy skills**.

³ Camfed International and DLA (multi-country projects) are counted as a single project.

- **These increases, while statistically significant, are relatively small⁴.** For oral reading fluency scores, which are expressed in words-per-minute rather than percentage correct, this means that girls correctly read 10 words-per-minute more than they did at baseline after 18 months of projects' interventions⁵.

Are the observed changes different for literacy and numeracy?

- **GEC-T girls improved slightly more in literacy than in numeracy:** their average percentage correct score increased by eight percentage points in EGRA and by ten percentage points in SeGRA, as opposed to four percentage points in EGMA and eight percentage points in SeGMA.

Did GEC-T girls improve more than non-GEC girls?

- When comparing GEC-T beneficiaries to comparison groups, it appears that **the average GEC-T girl has learned significantly more than the average comparison girl between baseline and midline.** This again is true across all four learning assessments. But the difference-in-difference effect is small: learning scores increased between two percentage points (EGRA and EGMA) and three percentage points (SeGRA and SeGMA) more than in the comparison group. On average, the progress of GEC-T girls is larger by only a couple of percentage points (or a couple of words-per-minute) more than that of comparison girls. This suggests that the **GEC-T is associated with a positive, but small, improvement in girls' learning.**
- When weighting each project equally in the portfolio, the GEC-T shows no statistically significant effect on learning. This is because **the average learning progress of GEC-T girls is driven by a small number of large projects:** three projects account for three quarters of the GEC-T target learning beneficiary population⁶. Many smaller projects are not associated with any significant improvement in learning over and above their comparison groups.

How do girls' midline learning levels compare with international standards and benchmarks?

- At midline, about one-fifth of GEC-T girls are still unable to read a familiar or invented word correctly. Among older girls who have been given SeGRA and SeGMA, more than one-fifth score zero in short essay construction, about one-third score zero in algebra and one-half in sophisticated word problems. This means that **many girls still do not have the most basic literacy and numeracy skills after several years of GEC intervention.**
- **Most GEC-T girls still fall short of international standards of oral reading and comprehension:** at age 10, GEC-T beneficiary girls only read 45 words-per-minute on average across projects. This is at the lower end of a contextually relevant reading fluency benchmark of between 45 to 60 words-per-minute⁷. Children in the USA would be expected to read at over 100 words per minute by the end of grade three⁸. At age 11, girls can read as many as 63 words-per-minute on average.
- **Only 8% of GEC-T girls aged 10 at midline can understand what they read.** Using the UNESCO-led Global Education standard that "all children should read and understand what they read by age 10, or at the latest by the end of primary school"⁹, this implies that the average GEC-T girl is still trapped in learning poverty at midline. The share of girls who can understand what they read reaches about 30% towards the end of primary school. This share increases steadily as girls grow older but remains capped at 40% of girls aged 18.

RQ2. Who has learned the most (and least), and how much did they learn?

To answer this question, we ran separate difference-in-difference regressions on key subgroups, for each of the standard EGRA/ EGMA and SeGRA/ SeGMA subtask and aggregate scores. The value of the difference-in-difference coefficients provides an estimation of the impact of the GEC-T on each subgroup between baseline and midline.

⁴ The average between the baseline and midline data collection is of 18 months, which is the average time girls have had to improve their learning. A ten percentage point increase means that girls who scored 50% correct at baseline scored 60% correct at midline, or that girls who scored 10% correct at baseline scored 20% correct at midline.

⁵ As a matter of comparison, and although they have been developed for the USA curriculum hence for a different context, the DIBELS benchmark expects an increase of about +45 words-per-minute during Grade 2, and of about +30 words-per-minute during Grade 3.

⁶ These are Camfed International in Tanzania, Zambia and Zimbabwe, EDT Kenya and Discovery Learning Alliance (DLA) in Ghana, Kenya and Nigeria.

⁷ See Developing Cross-Language Metrics for Reading Fluency Measurement (Abadzi, 2012).

⁸ See DIBELS benchmarks: http://oregonreadingfirst.uoregon.edu/downloads/assessment/dibels_benchmarks_3x.pdf

Note these benchmarks have been developed for the English language. Seven out of 27 GEC-T projects have assessed girls' learning in different languages of assessment than English.

⁹ See the official page of the Global Coalition for Education Data by the World Bank and UNESCO: <https://gaml.uis.unesco.org/learning-poverty/>

- The increase in girls' learning scores has been accompanied by a substantial decrease of the share of girls scoring zero, which implies that **even the most educationally marginalised girls learned between baseline and midline**. This improvement can be observed across all literacy and numeracy subtasks¹⁰.
- **Progress in learning over and above the comparison group is consistent across key subgroups** as defined by girls' enrolment status, school level, disability status, orphan status, family status (mother or father not living in the household¹¹), household poverty and education levels. In other words, the learning of girls belonging to these subgroups has improved slightly more than the learning of girls of the same subgroups who live in comparison areas.
- **However, focusing on the progress of GEC-T girls between baseline and midline, some subgroups appear to have learned at a significantly slower pace than other**. This is the case for girls who are overage for their grade, girls living in poor households and girls whose primary caregiver or head of household have no education (in literacy and numeracy); and girls whose father or mother does not live in the household (in numeracy only).

RQ3. Why did some girls learn more than others?

To answer this question, we examined whether girls in the project schools improved over and above comparison girls in a range of intermediate outcomes measured by projects. Then, we correlated any positive change in intermediate outcomes with changes in learning.

This was undertaken as GEC-T projects attempted to reduce barriers to learning that many marginalised adolescent girls confront as the result of barriers related to cultural, gender and social norms, education system, institutional and economic factors. Through a variety of approaches, including project-specific interventions and activities, projects contribute to: (a) intermediate outcomes such as increased attendance; improved teaching quality; progress toward safe, inclusive, conducive to learning, and empowering learning environments; improvement in girls' life skills, self-esteem and aspirations; and parental and community attitudes toward girls' education; and consequently to (b) learning and transition outcomes.

- **Intermediate outcomes have improved for the average GEC-T girl more than the average comparison girl between baseline and midline** in specific life skills areas (i.e., academic self-confidence, and leadership and communication), most teaching quality measures (i.e., gender-sensitive pedagogy, teacher attendance, child-centred practices, interactive learning, school corporal punishment), and parental attitudes to girls' education. GEC-T girls' intermediate outcomes have not improved more than the average comparison girl in relation to girls' attendance or reducing economic constraints to continue schooling.

What are the factors (including the projects' intermediate outcomes) that are associated with learning across the GEC-T portfolio?

- **Where improvements over and above the comparison group are observed, an improvement in the intermediate outcome is often associated with an increase in learning scores**. This is observed in relation to life skills, teaching quality indicators measuring interactive learning, reduction in school corporal punishment, and parental attitudes to girls' education.
- **While GEC-T improved intermediate outcomes, they are not always enough to raise learning levels alone, as hypothesized by projects' theories of change, and demonstrated by the small learning gains observed**. Girls largely remain constrained by the context they live in. Teacher quality, social norms, appropriateness of learning resources, adequacy of school and classroom facilities and infrastructure are all identified as additional necessary enabling factors to improved learning.

RQ4. Did girls successfully transition between baseline and midline?

To answer this question, we required information on the status of girls at two data points in time (baseline and midline). So, we examined transition outcomes with the sample of girls that were recontacted at midline from baseline – 59% of the baseline sample.

- **As most girls recontacted were found in school, the transition analysis offers limited value** – a successful transition is defined as a girl progressing one or two grades between baseline and midline (which more than 9 out of 10 girls achieved), and an unsuccessful transition is defined as a girl repeating a grade or

¹⁰ Except for oral reading fluency: only 2% of GEC-T girls could not read a single word at baseline.

¹¹ The number of married girls and mother girls was too small to

dropping out of school (which was the case for 5% of the re-contacted sample). **Girls in the treatment group were marginally more likely to progress a grade than girls in the comparison group.**

- This suggests a more optimistic picture for transition outcomes than reality because girls not re-contacted (41 per cent of the sample) are unaccounted for and some of these would most likely not have transitioned successfully. To address this, we adjusted the transition analysis to account for attrition and applied average reported successful, unsuccessful and unknown transition rates collated from projects' documentation to girls not re-contacted. Altogether, **we estimate that 63% of girls transitioned successfully, 23% did not transition successfully, and the transition status for 14% of girls is unknown.** This provides a more realistic and conservative estimate of transition outcomes for in-school GEC-T girls as many of those with unknown transition status most likely did not transition successfully.

Who has not successfully transitioned in school?

- Amongst recontacted girls, young girls (i.e., aged 6 to 8 and those who are underage for grade) are between 8 to 10 percentage points more likely to progress a grade than older girls (i.e., those who are overage for grade and who are aged 16 to 19). This is due to high dropout rates for girls aged 16 to 19 (8%) and girls overaged for grade (4%).
- Adjusting the transition analysis to account for attrition, the estimated successful school progression drops by about 30 percentage points. This shows that girls who were not re-contacted experienced the lowest progression rates while the specific transition pathways of girls not re-contacted remain largely unknown.

Limitations of the GEC-T evaluation design and effects on study findings

- Each GEC-T project operates in a different country and local context, with target populations of different age, marginalisation levels, social and economic characteristics, and intervenes through a range of different programme activities. **This makes it impossible when comparing the effect that each project has had on learning (or transition) outcomes, to attribute observed differences to any specific contextual factor, or to the effectiveness of specific projects' interventions.**
- As projects followed different targeting and sampling strategies, the project composition of sub-samples differ from one to another. It means that sub-samples are not representative of the overall portfolio-level sample. For the same reason, changes in outcomes at the subgroup level are highly prone to sample composition effects: for example, the average in-school girl and the average out-of-school girl do not only differ in their enrolment status, but in other key characteristics such as their age or socioeconomic backgrounds, and most importantly, may not come from the same project target populations¹². In other words, **cross-project comparisons cannot fully account for all these differences statistically, so any portfolio-level estimates need to be interpreted with these caveats in mind.**
- All GEC-T projects used the standardised and structured EGRA/ EGMA and SeGRA/ SeGMA tests. However, projects often used only a selection of subtasks, or gave different subtasks to different cohorts (typically grade levels). This means that **aggregate scores are not directly comparable not only across projects, but also across different girls within the same project, and across time for the same girl.** As such, learning assessments based on aggregate assessment scores need to be interpreted with caution.
- Due to sample imbalances between the treatment and comparison group that do not make them fully comparable within each project and of attrition and replacement bias, the estimated impacts are correlations and not causal effects. Therefore, a positive difference-in-difference result only suggests a positive impact of the GEC-T. Instead of 'impact', we therefore speak of **'correlations' or 'associations' between the GEC-T interventions and the observed changes in learning and transition outcomes.**
- **Reading wpm capped at 100 creates a ceiling effect.** For the reading fluency task, as per FM guidance, projects recorded raw scores, and words per min scores are capped at 100. The analysis uses the capped scores, as most datasets which were submitted by projects do not include the non-capped variables. However, the capped scores limit the ability to measure changes (and progress) in learning, as they include girls who have read more than 100 words per minute.
- **Transition analysis can only be examined with accuracy for girls who have been re-contacted at midline.** Girls in the recontacted sample are largely in school, and so the findings suggest a more optimistic

¹² Typically, more than 95% of the out-of-school girls of the GEC-T panel sample come from the same project, Mercy Corps Nigeria.

picture than reality because they do not account for girls who most likely dropped out of school. We conducted transition analysis that accounts for attrition based on project documents. Nonetheless, the figures are only as good as the assumptions that are behind them. Only nine projects reported estimates of girls lost to attrition between baseline and midline. For other projects, the status of lost girls is largely unknown.

Conclusions

- **GEC-T girls have significantly improved their learning between baseline and midline**, across all the literacy and numeracy subtasks. **Their learning improvements were significantly higher than those of comparison girls**, which demonstrates **an association between GEC-T interventions and improved learning**. The observed learning progress, however, is small, not only in absolute terms (when comparing girls' baseline to midline learning levels), but also over and above the comparison group: after about eighteen months of projects' activities, GEC-T girls can read only a few more words (10 additional words-per-minute), and complete a few more arithmetic operations correctly than comparison girls.
- **GEC-T girls' learning progresses as they get older and progress through the school grades. However, most of these girls still fall far short of international benchmarks for oral reading and comprehension. Girls do learn, but they start to do so later than what would be expected.** Indeed, it is only at age 10 that the average GEC-T girl is able to read 45 words-per-minute, which is at the lower end of a contextually relevant reading *fluency* benchmark of between 45 to 60 words-per-minute. At age 11, the average GEC-T girl can read 63 words-per-minute. Reading comprehension, however, remains very low: only 8% of girls aged 10, and no more than 30% of girls towards the end of primary school, are able to understand what they read. This suggests that most GEC-T girls are still trapped in *learning poverty* at midline, according to UNESCO Global Education Standards.
- The patterns of girls' learning progress hide **important disparities across projects**. Among the largest GEC-T projects, the (positive) difference between the learning progress of treatment girls and comparison girls is higher than on average. Indeed, **three projects with statistically positive learning increases over and above their comparison group account for three quarters of the GEC-T target learning beneficiary population**. Conversely, smaller projects (projects with relatively fewer learning beneficiaries) tend to show significantly less or no learning progress.
- **Key subgroups of girls show relatively consistent learning progress between baseline and midline.** This suggests that the GEC-T has consistently improved the learning of all girls, including the most marginalised in the sample. However, a few subgroups appear to have learned at a significantly slower pace than other. This is the case for girls who are overage for their grade, girls living in poor households and girls whose primary caregiver or head of household have no education and girls whose father or mother does not live in the household. Although projects targeted and sampled different groups, which means subgroup findings are very dependent on project-level performance, and one or two projects can account for the largest part of a subgroup sample. So, subgroup-level differences are for a substantial part a result of project-level differences.
- Although recognised as key enablers to improving learning and transition in projects' theories of change, **improvements in intermediate outcomes** over and above the comparison group in life skills, quality of teaching measuring interactive learning and school corporal punishment and parental attitudes, have **translated into small gains in learning**. Girls largely remain constrained by the context they live in and simultaneous improvement in other enabling factors such as teacher quality, social norms, learning resources, and school facilities is necessary to achieve improved learning.
- Transition outcomes can only be tracked for girls who were re-contacted at midline (from baseline) and more marginalised girls are lost to attrition by midline – this means the transition analysis is prone to attrition bias. So, any transition analysis that does not account for attrition would present an overly optimistic estimate of transition rates. Most recontacted girls were found in schools, with more than 9 out of 10 recontacted girls progressing at least one grade level between baseline and midline, and only five per cent of girls either repeating a grade (a majority in primary) or dropping out of school (a majority in secondary). **Girls in the treatment group were marginally more likely to progress a grade than girls in the comparison group.** However, after adjusting the analysis for attrition, we can only be confident that nearly 6 out of 10 girls progressed successfully.
- **The characteristics of the GEC-T design limit the study's ability to draw conclusions on the overall impact of the GEC-T and on the effectiveness of different interventions.** Because each GEC-T project operates in a different country and local context, with target populations of different characteristics (age, grade

level, enrolment status, socioeconomic factors) and intervenes through different programme activities, the effects of observable and non-observable project-level differences play an important role at the portfolio-level. As such, subgroup samples are very dependent on project composition effects. Besides, projects used different subtasks to assess learning and different definitions of transition. Direct comparisons between projects' progress need to be interpreted with these caveats in mind. Other aspects related to the GEC-T evaluation approach, such as the capping of words-per-minute scores at 100 words-per-minute, and the absence of transition data for girls lost between baseline and midline, have also limited the extent to which progress could be analysed.

Recommendations

The GEC-Transitions (GEC-T) Window is approaching the end of its lifetime with only four GEC-T projects remaining by October 2022. These recommendations are aimed at the FCDO and wider policy-makers to inform future girls' education programming, policy and measurement approaches. As a quantitative assessment of the GEC-T's impacts and associations with different variables (rather than causal relationships), these recommendations focus on high-level policy and programming objectives and measurement issues.

Policy

1. There is a global learning crisis and a need for continued support for girls' learning. GEC support has not sufficiently improved girls' literacy at an early enough age to enable them to learn as they progress through school, a stronger focus is needed on girls' literacy in early primary school grades.

While girls' reading fluency improves as they get older, their literacy is very low in the foundational years of primary school, which means that their progress is too slow as they do not have the reading skills needed to learn as they transition through school. At age 10, most GEC-T girls can hardly read (averaging just 45 words-per-minute) and less than 8% understand what they read. As reading fluently provides a foundation for learning other skills, a stronger focus is needed on understanding and addressing the factors preventing girls from learning the basics, to prevent them from being trapped in learning poverty. This could include greater attention to school readiness and girls' literacy in early primary school grades.

2. More targeted interventions are required to support those subgroups of girl learners who have fallen behind other GEC girls. Regularly assessing girls' learning and their learning needs is essential to identifying who is falling behind and what targeted support is needed to help those girls catch up.

Some groups of girls – those who are overage for their grade; are living in poor households; whose primary caregiver or head of household has no education; whose father or mother does not live in the household – learned at a significantly slower pace than other girls between baseline and midline, in both the treatment and comparison areas. These subgroups require more targeted interventions to enable them to catch up and keep up with their cohort.

3. Interventions to improve transition through education need well-defined pathways – especially for girls that drop out of school and /or migrate outside of project areas. This requires tracking strategies that allow projects to follow girls' progress across multiple pathways.

GEC-T was successful in supporting the girls that projects *could track* in their transition through education, but projects experienced high rates of attrition from their samples. When girls were sampled in their homes as opposed to schools, attrition rates were lower, but they were still high. Many projects were not designed to track girls outside defined project areas, meaning that when girls moved, they were lost.

This means the analysis on transition is biased as the "lost" girls are more likely to not have transitioned successfully. This is not just a statistical issue, but an education issue, as projects are losing beneficiaries early with the outcomes for these girls largely unknown.

It is crucial to invest in better tracking to identify what happens to girls who leave project areas (or move to pathways not supported by projects) but also to keep supporting girls in their transition to adulthood. Simple (though costly and time-consuming) activities can reduce attrition, such as keeping in touch with girls between evaluation points, collecting additional community contacts for the girl outside the household, or visiting girls at schools they have moved to.

Measurement

4. A more consistent measurement strategy is needed across the portfolio to achieve the right balance between project-level adaptability and portfolio-level consistency to enable aggregate impact assessment and to compare performance across the portfolio.

The GEC-T does not easily lend itself to aggregating project-level data to assess the overall impact of the portfolio and to compare performance across different girls, interventions and contexts – projects used different targeting and sampling strategies; undertook bespoke adaptations of learning assessments; and used different ways of tracking girls. The measurement requirements for GEC Phase II were less standardised than in Phase I to improve the flexibility and relevance of projects' evaluation designs. However, this limits a portfolio-level evaluation. The FCDO and other policy-makers need to consider the extent to which portfolio level aggregate assessments of impact and performance across a portfolio are important and the trade-offs involved when designing future girls' education programmes.

5. Intermediate outcomes need to be measured consistently using suitable scales, and they need to be linked to learning outcomes in the data. The same girls who are being assessed for changes in their intermediate outcomes should also be assessed for changes in their learning outcomes to establish these links.

The observed improvements in (some) intermediate outcomes have translated into small gains in learning. However, analysis has been limited by the inability to systematically link learning gains (captured through school assessments) to changes in intermediate outcomes (captured by household and girl surveys). As a minimum requirement, projects should measure the explicit link between changes in intermediate outcomes and their association with changes in the same girls' learning outcomes. Assessing such links requires tailoring evaluation designs at the portfolio- and project-level to ensure all outcomes are captured systematically so that they can be consistently linked together.

Further research

6. Further research would be beneficial to illuminate the factors driving the positive results achieved by the three large projects in the GEC-T that delivered most of the learning gains.

Larger projects (in terms of the number of beneficiaries) were more successful in increasing girls' learning outcomes compared to the comparison group. More investigation is required into the reasons why different types of interventions improve learning for different types of girls and under what conditions, including what effect the quality of a project's design has compared to the quality of its implementation and the role scale has on its success.

1. Introduction

This study is conducted as part of the independent evaluation (IE) of the Girls' Education Challenge Phase II (GEC II), an eight-year (2017-2025) programme supported by the UK Foreign, Commonwealth and Development Office (FCDO) that aims to improve the learning opportunities and outcomes of over one million of the world's most marginalised girls.

GEC Phase II is delivered through two funding windows: (1) GEC Transitions (GEC-T) Window, which supports marginalised girls participating in 27 GEC Phase I projects across 15 countries¹³ in transitioning to the next stage of their education; and (2) Leave No Girl Behind (LNGB) Window, which supports 14 projects in 10 countries working with highly marginalised, adolescent girls who have never attended or have already dropped out of school. Project activities are undertaken by locally based Implementing Partners (IPs), who are contracted to FCDO and managed by a Fund Manager (FM), a consortium led by PricewaterhouseCoopers (PwC).

This study focuses on the **GEC-T Window**, to enable the Foreign, Commonwealth and Development Office (FCDO) and the Fund Manager (FM) to respond to the **accountability objective** of whether, and to what extent, the GEC-T is associated with changes in girls' learning and transition outcomes, as well as to provide **portfolio-level benchmarks for future IE studies** on what magnitude of impact the GEC-T has achieved, which subgroups were most affected by interventions and which intermediate outcomes are associated with improved learning.

Study 3 relies exclusively on the baseline and midline quantitative data¹⁴ collected by the GEC-T projects' external evaluators (EE) and estimates the changes in learning and transition occurring before the disruptions caused by Covid-19.

The objective of the study is twofold:

- to **create a portfolio-level dataset of the quantitative baseline and midline data collected by GEC-T projects' external evaluations** through their learning assessments, girl surveys, household surveys and primary caregiver surveys. This dataset will be used for research, further analysis and future studies; and
- to **quantify the aggregate changes in girls' learning and transition outcomes associated with the GEC-T projects' intervention**¹⁵ between their external baseline and midline evaluations.

Specifically, the study aims to answer the following research questions:

1. How did girls' learning levels change between baseline and midline?
Did GEC-T girls improve more than non-GEC girls? Are the observed changes different for literacy and numeracy?
2. Who learned the most (and least), and how much did they learn?
How did changes in learning outcomes vary across contexts: by girls' individual characteristics? By household-level characteristics? By school status?
3. Why did some girls learn more than others?
What are the factors (including the projects' intermediate outcomes) that are associated with learning across the GEC-T portfolio?

¹³ These countries are Afghanistan, Democratic Republic of the Congo (DRC), Ethiopia, Ghana, Kenya, Mozambique, Nepal, Nigeria, Rwanda, Sierra Leone, Somalia, Tanzania, Uganda, Zambia, and Zimbabwe.

¹⁴ Endline projects are not included in this study because the adaptations to endline data collection activities caused by Covid-19 resulted in many project external evaluations not gathering learning data. Therefore, conducting portfolio-level analysis of the aggregate impact projects had on learning including endline data was not feasible.

¹⁵ Girls are the primary beneficiaries of GEC-T projects and the subject of this study. Most projects also work with boys (exceptions include BRAC Afghanistan and VIVA Uganda). However, boys typically receive a subset of interventions and benefit, directly or indirectly, from activities aimed at improving learning outcomes for marginalised girls but not from broader support. For example, boys receive support for boys' clubs, remedial learning classes, and life skills from Relief Somalia; still, only girls receive an expanded package of interventions including bursary support, payment of school fees, or cash grants, school uniforms, among other types of supports. As a result, data on boys, whether from direct or indirect beneficiaries of GEC-T projects, are not available across the whole portfolio of projects. Specifically, only nine projects (AKF, Camfed International, Camfed Tanzania, CARE, LINK, Opportunity, Plan Sierra Leone, Relief Somalia, and World Vision) collect quantitative data on boys, and what is collected varies across projects. Some projects do not track boys for learning and transition (e.g. Relief Somalia only tracks boys for learning), others do not track their learning longitudinally (e.g. AKF), or do not administer an extended boys' questionnaire which allows for subgroup analysis. Thus, boys are not included in the analysis.

4. Did girls successfully transition between baseline and midline?
Who has not successfully transitioned in school?

The Terms of Reference (ToRs) are presented in [Annex 1](#), together with a brief explanation of any deviations from the agreed-upon ToRs.

Section 2 of the report describes the methodology, including the process to create the unique baseline-midline dataset at the portfolio level and the structures and limitations of the analysis.

Section 3 of the report presents the findings from the quantitative analysis and provides answers to Research Questions 1, 2 and 3. **Section 4** of the report focuses on Research Question 4 on changes in transition outcomes, before presenting the conclusions and recommendations in **Section 5**.

2. Methodology

Projects included in the analysis

This study is solely based on projects' quantitative data collected as part of project external evaluations (i.e., learning assessments, girls survey (GS), household survey (HHS), and primary caregiver survey (PCG) data) for both baseline and midline. This section describes the availability, and validity, of these datasets for analysing learning and transition outcomes. Annex 2 provides further details on projects' data availability and the process followed to create the unique baseline-midline dataset used in this study.

Available projects for learning and transition analysis

The GEC-T supports girls and young women in 15 countries through 27 projects.¹⁶ Out of 27 GEC-T projects, seven were completely excluded from the study, four for lacking midline learning data, and three for not having a valid (conclusive) baseline data that can be compared with midline data. This leaves 20 projects that can be used to explore changes in learning and transition outcomes, of which two (Relief International Somalia and World University Service of Canada (WUSC) Kenya) do not have a comparison group but are still included in the study.¹⁷

The learning analysis is run twice, on the cross-sectional sample (full sample of baseline and midline girls) and the panel sample (sample of girls followed over time). **The cross-sectional learning analysis includes all 20 projects.**

However, due to a lack of common girl's unique IDs between the rounds, two projects are excluded from panel analysis (for learning and transition outcomes). As a result, **18 projects are included for panel learning analysis.**

In addition, one more project is excluded from the transition analysis due to inadequate matching between the baseline and midline using girl's unique IDs. As a result, **17 projects are suitable for transition analysis.**

Annex 3 (Table 24) provides a high-level summary of the availability of data by project for each round.

Figure 1: Projects included in Study 3

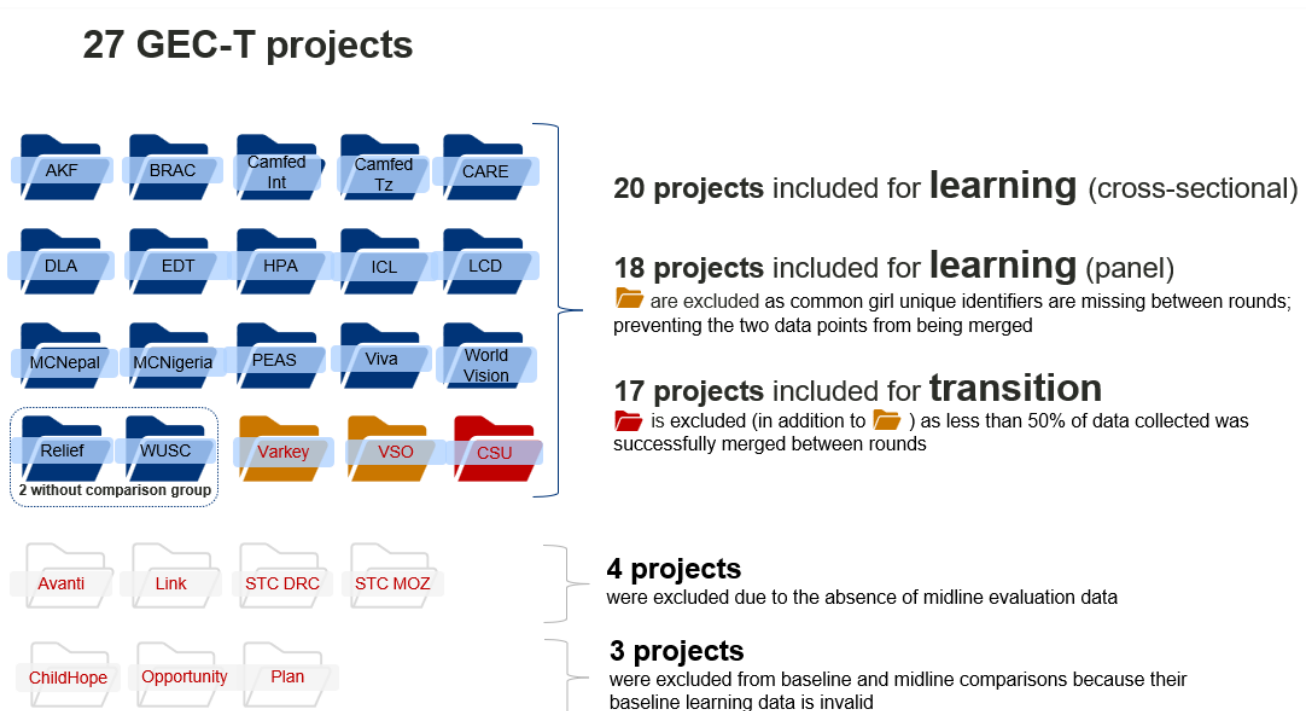


Figure 1 shows there are three projects with invalid baseline data (ChildHope, Opportunity, Plan) and four projects with no midline evaluation data (Avanti, Link, STC DRC, STC MOZ). These are excluded, leaving 20 projects. Of these, two

¹⁶ Camfed International and DLA multi-country projects are counted as a single project.

¹⁷ The projects without a comparison group will be included in *simple-difference* analysis, which is the analysis of changes between the baseline and midline data, for the treatment group only. They will be excluded from the *difference-in-difference* analysis, which compares the simple-difference over time of the treatment and comparison groups. See the sub-section of Analysis of learning outcomes for further detail on this analysis.

more projects (Varkey and VSO) have no common individual unique identifiers at baseline and midline and so are also excluded. Finally, Cheshire Services Uganda (CSU) was deemed unsuitable for the transition analysis as less than 50% of the data collected was successfully merged between baseline and midline.

Subgroup analysis

We use data from different sources (girls, caregivers and household surveys) to classify girls into key subgroups to answer research question 2 (“Who has learned the most and least”) and research question 4 (“Who has not successfully transitioned in school?”).

The following key subgroups were explored (on the basis of girls' *baseline* values):

- Subgroups derived using girls' surveys: girls aged 6 to 8/ 9 to 11/ 12 to 13/ 14 to 15/ 16 to 19; girls in school/ alternative education/ out-of-school; girls in primary/ secondary school; girls overaged/ underaged/ in the correct age for their grade; girls without disability / with one type of disability / with multiple types of disability; girls whose mother does not live in the household; girls whose father does not live in the household;
- Subgroups derived using caregivers' surveys: single orphans / double orphans; married girls; mother girls; girls whose primary caregiver has no education;
- Subgroups derived using household surveys: girls whose head of household has no education; and girls from poorer/ less poor households.

Table 25 in [Annex 3](#) documents the indicators used as key subgroups in the analysis, and their definition, or procedure of computation.

Each GEC-T project operates in a different country and local context, with target populations of varying age and marginalisation levels. For example, few projects work with out-of-school girls, girls in alternative education, or girls in lower primary (aged 6 - 8). As a result, any subgroup analysis is highly prone to sample composition effects, which means that conclusions are representative of the sample of projects included in that group.

Table 12 shows the *share of baseline sample* in each subgroup included for panel learning analysis. Key subgroups which are captured by the smallest number of projects include: out-of-school girls (3 projects: Care, HPA, and Mercy Corps Nigeria), which represents 2% of the baseline panel sample; girls in alternative education (5 projects: AKF, BRAC, Camfed International, LCD, and World Vision), which represents 5% of the baseline panel sample; and girls aged 6 - 8 (7 projects: AKF, BRAC, Camfed International, CSU, DLA, Viva, and World Vision), which represents 2% of the baseline panel sample.

Sample sizes

All GEC-T projects had to longitudinally track a cohort of beneficiary girls in the baseline sample for both learning and transition outcomes. A cohort of girls (referred to as learning cohort) was tracked in school for learning outcomes by some projects and given learning assessments and girl surveys. Another cohort of girls (referred to as transition cohort) was sampled in the household and given girl surveys and household surveys which included questions tracking transition outcomes.

The rationale for this was that measuring transition outcomes requires a different sampling approach to capture the relevant information on the many pathways that girls can take. On these grounds, the Fund Manager's GEC-T MEL Guidance recommended projects link the learning and transition cohorts but also allowed external evaluators to treat them separately.

As a result, there is a mix of strategies across projects. Some projects de-linked the transition and learning cohorts; others combined them and administered all tools to all of the girls in their sample (such as Aga Khan Foundation (AKF)); while others combined these two strategies (i.e., administered learning assessment to the transition cohort but not household questionnaire to the learning cohort, and vice-versa). Given this, we do not use the original groupings in this study (learning/transition cohort) as there is no one-size-fits-all definition at the portfolio level, but we use all available data in our analysis.

Figure 2 depicts how the sample composition varies when we combine the learning data with data collected from the girls' surveys and household and primary caregiver questionnaires. For visual simplicity, it only displays the sample at baseline (in light blue) and those girls re-contacted from baseline (in dark blue).

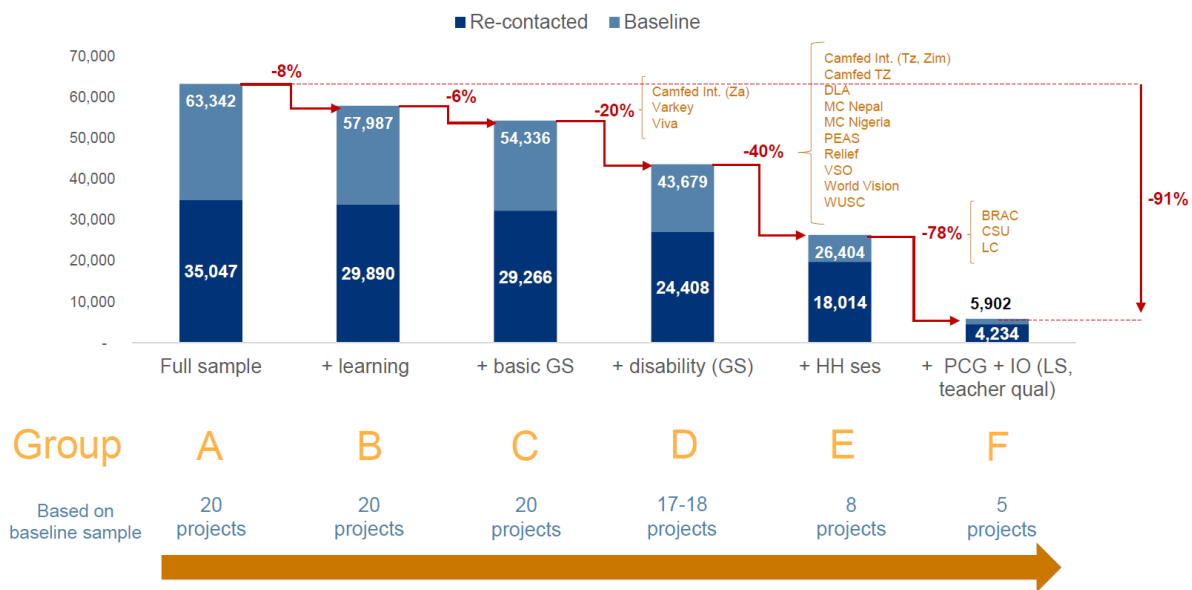
The full sample (described first from left to right) shows the total number of girls (treatment/ comparison group) in the dataset used for analysis, referred to as Group A (Figure 2Table 1). **This sample (light blue) is used as the basis for our analysis on attrition; while the sample of girls re-contacted from baseline (in dark blue),** minus two

projects with no standard unique identifiers between baseline and midline and missing information in the grade variable across projects, **is the basis for the transition analysis in schools.**

Group B describes the total number of girls in the sample who were given learning assessments at baseline and midline. **This sample is the basis for the learning analysis: cross-sectional (when using full baseline and midline samples) and panel (when limiting the sample to girls re-contacted from baseline).**

Groups C to F in Figure 2 depict other samples (with variations of these used in analysis)¹⁸ to demonstrate that GEC-T external evaluators administered the learning assessments, girl surveys, and household and primary caregiver questionnaires to different samples of girls and their households within projects. **More importantly, it shows that when these sources are combined, our sample is reduced and is no longer representative of the portfolio of projects.**

Figure 2: Comparison of sample compositions



GS = Girl survey; HH = Household survey; ses = socioeconomic status; PCG = Primary Caregiver survey; IO= Intermediate Outcomes (from GS, life skills and teacher quality). Multi-country projects can fall in different groups categories (i.e. Camfed International split between Group D and E).

The reduction in sample size is caused by missing data for girls within projects and whole projects. We observe a small reduction in sample size when we restrict the sample to those girls with learning data available (Group B), and girls with learning and girls' schooling characteristics (Group C)¹⁹. This reduction is partly attributable to girls being tracked for transition but not learning outcomes (from A to B) or from having incomplete data on age or grade status. When we add girls' disability status (Group D), reported by girls themselves, the sample reduces by 31% (from starting point in Group A) and three projects get dropped.²⁰

As discussed, many projects did not link learning data to household data. Surveys administered to the household heads and caregivers include information on household socioeconomic situation (such as poverty indicators and household level of education). Figure 2 shows 11 projects are dropped when the sample of girls is restricted to girls with learning data, schooling characteristics, and household surveys at baseline or at midline.²¹

Finally, when we include other girls' characteristics reported by the primary caregiver, such as orphan status and whether the girl is married, mother, or disabled, and questions mapped to intermediate outcomes, such as life skills

¹⁸ The samples (described from right to left) show the total number of girls in the sample (treatment/ comparison group) – Group A; the total number of girls in the sample who were given learning assessments in both rounds and have complete data on age, grade, and enrolment status at baseline (Group C); the total number of girls in Group C who were also administered the Washington Group questions on disability at baseline (Group D); the total number of girls in Group B with additional data on household economic situation at baseline (Group F); the total of girls in Group B with additional data asked to primary caregivers and measures of life skills and teacher quality.

¹⁹ Includes in-school girls only.

²⁰ VIVA and Camfed International Zambia did not collect disability data from the girls at baseline, only from their caregivers. Varkey dropped the disability data collected from girls at baseline for anonymisation purposes. In addition, this reduction includes girls who had missing data on disability across different projects.

²¹ For simplicity, Figure 2 only shows the sample size at baseline and for those recontacted. However, the list of projects includes those dropped either because the data is not available at baseline or at midline.

and teaching quality indicators, the sample size of girls drops significantly (by 91%) and the number of projects reduces to just six (out of 20). The limitations this presents to the analysis are described at the end of this section.

We describe Group A and Group B next.

Full sample – Group A

Looking at the 20 projects with valid and complete data, in Table 1, the total sample of girls in the treatment and comparison group at baseline is 63,342 and at midline is 56,729. The total sample of girls re-contacted at midline is 35,047. These samples consist of girls only assessed in learning, only tracked for transition, or both.

Table 1 summarises the sample size at baseline and midline, and the percentage of recontacted at the portfolio level. The midline sample includes girls re-contacted from baseline and 'top up' girls at midline (i.e., 'new girls' who replaced those who were not found at midline)²² – projects often 'topped up' their midline sample to account for the loss of sample between baseline and midline and to ensure an appropriate sample at midline.

Table 26 in Annex 3 disaggregates this by project.

Table 1: Sample size for Study 3 for Group A

Sample	Sample at baseline	Sample at midline	Girls recontacted at midline	
	N	N	N	N
All	63,342	56,729	35,047	59%
Treatment	38,364	35,916	21,068	59%
Comparison	24,978	20,813	13,979	61%

* Varkey Foundation and VSO are included in the sample at baseline and midline but are not included in the percentage of girls recontacted reported in Table 1. Therefore, the share of recontact is calculated excluding these two projects samples from the denominator – the sample at baseline.

Learning sample – Group B

As in Group A, looking at the 20 projects with valid and complete data, the total sample of girls in the treatment and comparison group at baseline is 57,987 and at midline is 51,545. The total sample of girls re-contacted at midline is 29,890. These samples consist of girls assessed in learning (Table 2).

Table 2: Learning sample size for Study 3 for Group B

Sample	Sample at baseline	Sample at midline	Girls recontacted at midline	
	N	N	N	N
All	57,987	51,545	29,890	52%
Treatment	34,306	32,934	18,220	53%
Comparison	23,681	18,611	11,670	49%

Comparability of treatment vs comparison samples

To assess the balance of the sample between the treatment and the comparison groups at baseline, we restrict the sample to Group B, C and D (see Table 3) in the table below.

²² These girls often received a different unique identifier to the girls they replaced.

Table 3 shows that the treatment vs comparison samples are relatively well balanced in terms of age, grade, enrolment status and percentage of underage-for-grade girls, and disability.^{23,24} The differences in means for grades, overage and having a disability are statistically significant but the differences are small, with the exception for the percentage of children who are overage-for-grade which are more frequently represented in the treatment groups.

Table 3: Comparison of means, treatment and comparison

		Group B	Group C					Group D
		Enrolment status	Age	Grade	Enrolment status	Being overage	Being underage	Girl with disability
Baseline	Treatment mean	97%	14.02	7.32	100%	41%	8%	10%
	Comparison mean	96%	14.02	7.40	100%	36%	9%	11%
	Difference	1%	0.00	-0.07	0%	4%	-1%	-2%
	<i>P</i> -value	0.00	0.92	0.00	-	0.00	0.00	0.00

Note: Group B (N treatment = 34,151; N comparison = 23,611); Group C (N treatment = 32,244; N comparison = 22,092); Group D (N treatment = 26,280; N comparison = 17,399). The discrepancies in the difference is due to rounding. *P*-value of 0.00 does not mean the *P*-value is zero but less than 0.001. The actual *P*-value is reported in the table.

Attrition

Attrition level

We define attrition as the percentage of girls whose data were collected in the baseline but were not re-contacted in the midline. We used matching girls' unique identifiers between baseline and midline to define if a girl is re-contacted at midline.²⁵ Therefore, the base in the attrition calculation includes all girls sampled in the baseline (18 projects).²⁶

Attrition at the portfolio level is high – 41% of girls at baseline were not re-contacted by midline. *This rate also applies to girls in the treatment group.*



Table 4 summarises the level of attrition for both the treatment and comparison groups. Attrition varies considerably by project (2–75%); however, the bulk of projects have medium (21-40%) to high (41-60%) attrition rates. Camfed (ex-BRAC), DLA, and (to a lesser extent) MC Nepal have the highest re-contact rates for both treatment and comparison groups. Of these, Camfed (ex-BRAC) attributed its low attrition rates to the short amount of time between data collection rounds. At the portfolio level, the data seems to confirm this statement. We find a negative and significant association between the number of months between baseline and midline data collection rounds and recontact status. One additional month between the data collection rounds is associated with a 2.6% lower probability of being recontacted at midline (*P*-value = 0.039).

²³ When breaking down the sample by age, we also observed a balanced sample across ages and grades.

²⁴ Disability classification followed guidance of Washington Group disability definitions and utilised short scale of questions asked to girls. Girls are classified as having a functional disability whenever they express difficulty in any one domain. See **Error! Reference source not found.** for definitions of all the variables.

²⁵ Projects replaced girls from the baseline or topped up the sample with new girls to account for the problem of attrition and ensure an adequate sample at midline. Most projects assigned new unique identifiers to these added girls (e.g. BRAC and Relief added an R to the unique identifier to identify replaced girls). For a few projects, however, replacements (new girls) were assigned the same unique identifiers as baseline girls. We conducted extensive checks of the unique identifiers to identify possible cases where this was done, using projects' re-contacted variable and triangulated with reported figures in baseline and midline reports. We found that three projects used the same identifier for replaced girls: AKF replaced 69 identifiers due to questionable data, and parts of EDT and Mercy Corps Nepal's added samples used the same identifiers as in baseline. For these projects, we assigned a different unique identifier to these girls and did not include them in the sample of re-contacted girls at midline.

²⁶ Attrition was calculated using all girls sampled at baseline (Group A in light blue in Figure 2) from 18 projects whose re-contact status can be assessed using matching girls' unique identifiers. This excludes VSO and Varkey (see Figure 1).

Table 4: Attrition level by project

Level of attrition	Number of projects	Projects
Very low (less than 10%)	2	Camfed (ex-BRAC), DLA.
Low (11-20%)	1	MC Nepal.
Medium (21-40%)	8	AKF, BRAC, Camfed International (Zambia), HPA, PEAS, Relief, Viva, World Vision.
High (41-60%)	5	Camfed International (Tanzania and Zimbabwe), CSU, EDT, LC, MC Nigeria.
Very high (61-100%)	3	CARE International, ICL, WUSC.

Note: Level of attrition for both the treatment and comparison groups. Table 26 in Annex 3 shows the sample size at baseline, the percentage of girls re-contacted from baseline, and the percentage lost to attrition for each project and treatment and comparison girls separately.

The reasons for attrition reported in the midline external evaluation reports are impacted by the context in which they operate and the population of girls they support and thus vary by project. Projects were unable to track girls as a result of them dropping out of school (due to reasons such as marriage) or migrating to different locations outside of the project's area. Transfers to schools outside the project area were also mentioned as a factor – for example, when transitioning from one grade level to another, such as from primary to secondary. Table 28 in Annex 3 summarises these reasons.

However, project external evaluations recognised that the status of girls not re-contacted is largely unknown. Many projects do not track girls who moved outside the project area and cannot verify if the girls remain in education. A few projects shifted to other sources of evidence to track girls not found in midline, such as asking schools, parents, and neighbours. Nonetheless, this information may not be verified, and sometimes are only estimates of what the girls were doing. This is demonstrated by the fact that projects provide reasons but not proportions of girls lost as a result of the cited reasons.

We examine how project attrition rates reported in external evaluation reports differ from ours, alongside looking into the specific causes of attrition cited by each project (in footnotes). Table 27 in Annex 3 documents the main difference in reporting between this study and the external evaluations.

For almost half of the projects (8 projects: AKF²⁷, DLA²⁸, EDT Kenya²⁹, HPA Rwanda³⁰, MC Nepal³¹, MC Nigeria³², PEAS Uganda³³, and VIVA Uganda³⁴), our attrition rates (treatment and comparison groups together) are comparable to those reported by external evaluators (+/- 0-5%).³⁵

External evaluators also placed BRAC³⁶, Relief³⁷, and World Vision³⁸ in the “medium” attrition level, despite higher differences (>5%) in the rate reported.

We categorise seven projects differently (Camfed International³⁹, Camfed (ex-BRAC)⁴⁰, CARE International⁴¹, CSU⁴², ICL⁴³, LC⁴⁴, and WUSC⁴⁵), and the reasons are explained as follows.⁴⁶ Camfed International and CARE International have high attrition rates (and higher than those reported by their EE) because their external evaluators decided not to track all girls from baseline. Camfed (ex-BRAC) reported higher attrition rates than us because they categorised girls who moved schools (but whose data was collected) as attrition, while we put them as re-contacted. LC has high re-contact rates for the treatment group (88%) but due to changes in sampling protocol, lost all of their comparison group from baseline, placing them in the high attrition group for the overall measure. A few projects, CSU⁴⁷, ICL, and WUSC, are placed in a higher category than that reported by external evaluations because their data presented issues with the unique identifiers. Portfolio attrition, however, remains significantly high when CSU, ICL, and WUSC are excluded (38%).

²⁷ AKF Afghanistan considered that migration was the key reason for girls not tracked at midline. Part of attrition is also due to excluding whole communities because of security reasons and excluding girls with questionable data.

²⁸ DLA Ghana and Nigeria visited every Junior Secondary School that girls had transitioned to and administered the learning assessments to those girls in an attempt to reduce attrition. Attrition rates, however, were greater in DLA Kenya, partly because transfers between schools are very common in Nairobi and partly as a result of dropping two schools in the comparison group.

²⁹ EDT Kenya struggled to support the primary to secondary pathway. The midline EE report acknowledges that tracing project target girls who had completed Class 8 was a challenge.

³⁰ HPA operates in Rwanda where girls are assigned to schools based on their performance or area of residence. Therefore, the biggest reason for attrition identified by the project was girls migrating or moving to schools outside the project.

³¹ Mercy Corps Nepal reported that nearly all attrition was from Grade 10 who emigrated for work after completing the project intervention cycle for In-School Girls. Those who could be tracked after Grade 10 remained in the cohort and were re-categorised as School Graduates; presumably receiving a different intervention from the In-School Girls. Another reason for attrition among In-School Girls is because of fieldwork timing that coincided with girls finishing examinations and starting school break. Hence, some moved to be with their families and could not be tracked. Lastly, some OOS girls and School Graduates were found to have migrated outside the community for work.

³² For Mercy Corps Nigeria, there were changes in the programme design and some schools present in baseline are no longer part of the programme. In addition, there was high attrition from girls completing senior secondary education who were not tracked as they already completed school. Other reasons for unsuccessful re-contact are migration and dropouts due to marriage.

³³ For PEAS Uganda, many of the girls who were lost between baseline and midline were reported to have moved to another school, dropped out of education completely due to marriage, pregnancy, illness, lack of school fees, or completed lower secondary.

³⁴ VIVA reported that attrition in comparison group is very high because they are not obligated to take part. Additionally, many girls transitioning from primary to secondary were not tracked, but they are assumed to remain in education. Girls already in secondary level, by contrast, are assumed to have dropped out of education.

³⁵ We estimated the same attrition rates as external evaluators for AKF, DLA, MC Nigeria, and PEAS Uganda; and 1-5% differences for EDT Kenya, HPA Rwanda, MC Nepal, and VIVA Uganda. The reason for the discrepancies is from differences in reported number of girls tracked in baseline and/or re-contacted girls. In most cases, we were able to merge fewer girls than reported. DLA and MC Nepal adopted a joint cohort approach, where the same girls are re-contacted for both learning and transition. Attrition is higher for learning cohort as some girls could be tracked at home but were not administered learning assessments. The numbers used to compare with our data correspond to transition cohort attrition, as the girl has been successfully re-contacted, even though she had not taken learning assessments.

³⁶ BRAC highlighted three main reasons girls did not continue in the programme schools. These are migration, changing school, and marriage. Other less common reasons include dropping out of school because of illness or financial situation.

³⁷ For Relief International, all schools in Banadir region were removed from the midline sample. Some schools (in other regions) were replaced due to inaccessibility. Additionally, the project estimated that most grade 8 students who could not be re-contacted are likely to drop out, rather than continuing in a secondary school.

³⁸ World Vision mentioned that migration outside project area is the main reason for attrition. It is unknown if these girls remain in education.

³⁹ Camfed International identified migration as a reason for attrition. In Zimbabwe, it is hypothesised that in the time of economic decline, girls who are less marginalised had left education for work in other communities.

⁴⁰ Camfed (ex-BRAC) ascribed its low attrition rate to a short time between data collection rounds.

⁴¹ In CARE International, the most common reason for attrition is migration out of the community, which applies to half of lost girls. In other cases, the family remained in the community, but the girl had left home. A small proportion of attrition is from girls not being at home or refusing to participate.

⁴² CSU mentioned that girls moving to schools outside CSU support is the main reason for attrition.

⁴³ For ICL, attrition is from girls dropping out of school or completing primary or secondary levels of education (and transitioned out of education or moved to a different school).

⁴⁴ In LC, attrition in grade 8 is high as girls either moved to a new secondary school or dropped out. The project assumes most dropped out rather than continuing education. Other reasons include death, relocation, and school movement outside project area.

⁴⁵ WUSC operates in a complex refugee setting, which serves as barrier for girls continuing education in the same school including poverty and stigma. The project mentioned the transient nature of refugees, making them difficult to track over time. This is from many factors, including relocation due to conflict, disease, camp relocation, drought, terrorism, etc.

⁴⁶ As our rates differ, the reasons cited by projects in footnotes apply to the percentage of attrition reported by external evaluators documented in **Error! Reference source not found.** in Annex 3; with the remaining difference attributed to changes in computation where our denominator includes all girls sampled at baseline.

⁴⁷ For example, CSU presented data quality issues as less than 50% of data was successfully merged between rounds using girl's unique identifiers.

When we average the attrition rates of the 18 projects with equal weighting, portfolio attrition falls to 37%.⁴⁸ This shows that large projects such as Camfed International and EDT drive overall attrition.

Attrition bias

Attrition can bias estimates of baseline-to-midline changes in learning. In particular, this would happen if:

- Girls that dropped out of the sample were significantly different to the ones that remained in the sample – i.e., a selective attrition;
- AND if the selective attrition in the treatment groups is different from the comparison groups.

There is a risk that girls who were not recontacted at midline may on average be more marginalised relative to girls who were recontacted (see details in Table 29 in Annex 3). In this case, the estimation strategy risks attributing to the GEC-T effects that are actually due to midline sample being different from the baseline one. In other words, this means that we could no longer claim that the sample used for analysis was well-balanced before the programme interventions took place and threaten our identification strategy.

We framed this problem as a simple prediction one and investigated it using regression techniques. We adopted this approach to test the differences in attrition determinants, including across treatment and comparison groups. In particular, we regressed binary indicator for girl attrition (at midline) on treatment status and a set of baseline characteristics of the girls, clustering standard errors at the project level. These characteristics include age, being overage for grade, girl reported disability, overall literacy and numeracy, orphan status, being married, and being a mother. We do this in a stepwise regression model, adding girl characteristics to the model. Results are shown in Table 30 in Annex 3.

As shown by the absence of statistically significant difference for the *Treatment* variable, the **overall attrition rate is not predicted by treatment status**. Some control variables come out significant, meaning that less marginalised girls are somewhat more likely to be recontacted. However, the coefficients are low in magnitude. This finding may then be interpreted as the midline recontacted sample being somewhat less marginalised relative to the baseline sample of girls. As a result of that, it needs to be recognised that the GEC-T association with improvement in learning may partially be explained by the change in the girls' marginalisation status. We test this next.

Given the differences in characteristics of lost and recontacted girls identified above, we next checked whether the levels of the bias are different between the treatment and comparison groups. For this, we deployed a difference-in-difference estimator which uses the difference between the treatment groups (first difference) and recontact status (second difference). As we did not find any statistically significant differences in any of the girl characteristics,⁴⁹ we conclude that the above finding of more marginalised girls not being recontacted **does not present a concern regarding the difference-in-difference estimates in learning and validates the use of the panel dataset** (Table 31 in Annex 3). This is because while there is some form of selective attrition present (less marginalised girls are somewhat more likely to be recontacted), it does not differ between the treatment and the comparison group. See attrition bias section of Annex 3 for more details.

Replacement bias

Similar to attrition bias, we examined a possibility of replacement bias (i.e., if girls replaced in midline differ in their characteristics from girls lost in baseline). This source of bias, caused due to attrition, may affect estimates for the cross-sectional learning analysis. While there is some evidence of differing characteristics between replaced and lost girls (meaning not a 1:1 replacement), the differences in characteristics within treatment groups are low in magnitude apart from being overage for grade (i.e. replaced girls are somewhat more overaged for their grades relative to lost girls) and the statistical significance is primarily driven by the large sample sizes. Importantly, we find a similar pattern across all characteristics between the treatment and comparison group, which supports the validity of the cross-sectional difference-in-difference analysis (Table 32 in Annex 3).

⁴⁸ This is 32% when excluding CSU, ICL, and WUSC.

⁴⁹ No statically significant differences were identified for the non-weighted model or when using project level weights. When beneficiary population weights are deployed, a difference is identified for age and grade in favour of the comparison group girls (i.e. the comparison group girls who were lost were about half year older and attended higher grade, on average).

Analysis of learning outcomes

Learning subtasks and scores used

All GEC-T projects used the standardised and structured EGRA/ EGMA and SeGRA/ SeGMA tests. Table 5 shows the standard subtasks included in these assessments.

Not all subtasks were used by each GEC-T project. Some projects only used EGRA/ EGMA, some only used SeGRA/ SeGMA, and some used both. Within each assessment, projects sometimes only used a selection of subtasks, or adapted them for their own means. Non-standard subtasks that could not be mapped to any of the subtasks shown in Table 5 were excluded from our analysis. The full mapping of the learning assessments and subtasks for each project at baseline and midline can be found in [Annex 8. Mapping of baseline and midline learning assessment data](#).

Table 5: Learning assessments subtasks

Assessments	Subtasks	Subtask names
EGRA	Subtask 1	Letter sound identification (phonological awareness, mapping sounds to letters)
	Subtask 2	Familiar word (phonics. i.e. recognition of words)
	Subtask 3	Invented word (phonics. i.e. decoding new words using knowledge of sounds)
	Subtask 4	Oral reading fluency (passage reading), expressed in words-per-minute (wpm)
	Subtask 5	Comprehension
SeGRA	Subtask 1	Short reading comprehension
	Subtask 2	Longer reading comprehension
	Subtask 3	Short essay construction
EGMA	Subtask 1	Number identification
	Subtask 2	Quantity Discrimination
	Subtask 3	Missing Numbers
	Subtask 4a	Addition Level 1
	Subtask 4b	Addition Level 2
	Subtask 5a	Subtraction Level 1
	Subtask 5b	Subtraction Level 2
	Subtask 6	Word Problems
SeGMA	Subtask 1	Advanced multiplication, division, fractions, percentages, geometry, measurement
	Subtask 2	Algebra

Assessments	Subtasks	Subtask names
	Subtask 3	Sophisticated word problems

The learning subtask scores are expressed in **percentage correct scores**, which is the mark achieved by students out of the total possible mark reachable⁵⁰. The EGRA oral reading fluency score is the only exception as it is expressed in words-per-minute capped at 100 wpm maximum, following the FM guidance.⁵¹

These subtask scores are combined at the portfolio-level, which means that a score ranging from 0 to 100 was created for each subtask, across all projects for which the subtask is available. In addition, four aggregate EGRA, EGMA, SeGRA and SeGMA scores were created by averaging the percentage correct scores across the relevant subtasks.

In addition to percentage correct scores, **zero scores** (referring to the proportion of girls who score zero, i.e., who cannot perform even one item correctly for a given subtask) are also explored.



Projects often dropped subtasks between baseline and midline. Besides, many of them gave specific subtasks to specific groups (usually grade levels). This means that aggregate scores are not directly comparable not only across projects, but also across different girls within the same project, and across time for the same girl. As such, learning analysis based on aggregate assessment scores need to be interpreted with caution. More details are given in the section on limitations below.

Weighting

The number of girls who have been given the learning assessment at baseline and at midline differs widely across projects, and to a smaller extent, across subtasks within the same project (see [Annex 4. Learning samples](#) for full sample sizes by project and learning subtask).

For this reason, two different sets of weights were calculated for each round, project and subtask, and used in the learning analysis:

- **Beneficiary-population weights** are based on the target number of learning beneficiaries for each project and subtask. They are **used by default** for the learning analysis, which means they are the weights used in the report's body unless explicitly stated otherwise. Beneficiary-population weights provide estimates of the **effect of the GEC-T on the average (learning beneficiary) girl**; and
- **Project-equal weights, or “inverse sample weights”**, were created with a view to give the same weight to each project in the portfolio-level analysis. They are equal to the inverse of the actual learning sample size of each project (number of girls who were given the learning assessments). Analysis based on project-equal weights estimates the **average effect of the GEC-T across projects** and is usually shown in annex.

The values of each set of weights are detailed in [Annex 5. Weights for learning analysis](#).

Learning standards and benchmarks

While this study has not involved a detailed literature review, multiple reports and guidelines were reviewed with a view to identify international standards and benchmarks that could be used to compare girls' learning levels and progress to expected literacy and numeracy levels at specific ages and grades (see [Annex 11](#) for full bibliography for learning benchmarks). A starting point of the review was *Developing Cross-Language Metrics for Reading Fluency Measurement* by Abadzi which was used as part of the GEC I Step Change Window Evaluation Reports⁵².

Typically, EGRA and EGMA are assessments for 'early grade' students, therefore targeting grade 2-3 students. There are no international standards or benchmarks in place for grades 4 and above. Similarly, because the SeGRA and

⁵⁰ In EGRA and EGMA, the total number of marks typically corresponds to the number of items (e.g., letter sounds read, words read) or to the number of questions, marked 1 when correct and 0 when incorrect. In SeGRA and SeGMA, questions often have different markings, with some questions being marked out of 1, 2 or 3 depending on their perceived importance and difficulty.

⁵¹ While this has practical benefits, this also introduces an artificial ceiling into the scores.

⁵² See https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/676532/GEC-EM-SCW-Endline-Evaluation-Report-final.pdf

SeGMA were designed specifically for the purpose of the GEC evaluation, no international benchmarks and standards do exist (and none have been developed along with the assessments).

For oral reading fluency, RTI (who designed the EGRA and EGMA) suggest using different benchmarks by country and language⁵³. The list provided does not include all GEC-T countries, but the standard for oral reading fluency lies between 45 and 60 words-per-minute (wpm) read correctly in the language of instruction. This corresponds to the benchmarks that were used during Phase I of the GEC evaluation⁵⁴. The benchmark of 45-60 wpm was therefore used to define oral reading fluency in this study, though more research is needed in this area, as this is much lower than the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) benchmarks for the USA (which are 100+ wpm for grade three children, which is above the cap imposed on the GEC-T data).

For literacy in general, we use the standard defined by the UNESCO-led Global Coalition for Education Data⁵⁵ which defines learning poverty as *being unable to read and understand a short, age-appropriate text by 10*. We put this in relation to the scores obtained by girls in EGRA reading comprehension, which is composed of a handful of questions that aim to assess the students' understanding of a short passage that they have read orally⁵⁶.

First difference, difference-in-difference and “*impact*” of the GEC-T

We first analysed the learning progress of GEC-T girls, which was estimated as the midline score minus the baseline score – on average for the cross-section; and for the same girl using panel data. We did this just for the treatment group. We then look at the ‘difference-in-difference’ to estimate the effect of GEC-T projects.

These are equal to the difference between the baseline-midline progress achieved among the treatment group compared to that achieved among the comparison group. In other words, the difference-in-difference is equal to the learning progress achieved by girls from the treatment group *over and above* the learning progress achieved by girls from the comparison group.

For the subgroup analysis, the difference-in-difference regressions were run separately on key subgroups, for each of the standard EGRA/ EGMA and SeGRA/ SeGMA subtask and aggregate scores. The value of the difference-in-difference coefficients provide an estimation of the impact of the GEC-T on each subgroup between baseline and midline.

The statistical significance of the difference-in-difference coefficients are reported for *P*-values of below 0.05 and below 0.10. Calculations were made in Stata, using the *diff* command for descriptive statistics and cross-sectional difference-in-difference regressions, and the *xtreg* command for panel regressions, with standard errors clustered at the project level.



*By extrapolation, we describe a positive (and statistically significant) difference-in-difference as ‘a positive impact of the GEC-T’. Because of sample imbalances between the treatment and comparison group that do not make them fully comparable within each project and of attrition and replacement bias, **we cannot say the impacts are causal but rather correlational**. Therefore, a positive difference-in-difference result only suggests a positive impact of the GEC-T.*

*In the rest of the report, instead of the term ‘impact’, we will speak of **correlations** or **associations** between the GEC-T interventions and the observed changes in learning and transition outcomes.*

Cross-sectional vs. panel approach

Where possible, the learning analysis is run twice for the cross-sectional sample (full sample of baseline and midline girls, including girls lost between baseline and midline, and new girls added as ‘top-up’ to the midline samples) and on the panel sample (sample of recontacted girls only, whose baseline and midline data has been matched using unique identifiers). The analysis from the cross-sectional and panel samples leads to different estimates, but the overall picture and key takeaways are largely the same across the two. The main section report estimates from the panel with results from the cross-sectional are either included alongside or reported in [Annex 9](#).

We use panel analysis because it compares the same girls across time, while cross-sectional estimates report the group average differences. As such, the panel analysis provides more accurate estimates of girls' learning progress, both across time and over and above the comparison group.

⁵³ See [All Children Reading – Asia \(ACR-Asia\) EGRA Benchmarks and Standards Research Report \(RTI International, 2017\)](#).

⁵⁴ See [Developing Cross-Language Metrics for Reading Fluency Measurement \(Abadzi, 2012\)](#).

⁵⁵ See <https://gaml.uis.unesco.org/learning-poverty>

⁵⁶ The reading comprehension subtask directly follows the EGRA oral reading fluency subtask and is based on the same passage.

However, there is a trade-off, as the panel analysis excludes the girls who have not been tracked between baseline and midline, or whose identifiers have not been recorded consistently, such as in the Varkey and VSO evaluation samples and reduces the working sample by about 37%⁵⁷.

Running the same analysis on the two samples allows us to check the robustness of our findings. The analysis of attrition and replacement biases (see previous section) shows that biases exist for some variables, although they are small and similar for the treatment and comparison groups.

Association of Intermediate Outcomes (IO) with learning

Each project's Theory of Change (ToC) was grounded in the rationale that girls' abilities to enrol, regularly attend school, learn, and continue their education are greatly constrained by: (1) cultural, gender and social norms, (2) education system and institutional constraints (such as the lack of education facilities, of girl-friendly environments, of female teachers, low quality of education and unequal treatment of girls and boys in the classroom), and (3) economic constraints (including lack of access to learning materials, lack of affordable transport, etc.).

GEC-T projects attempted to reduce these barriers through a range of approaches, project-specific activities and school inputs. According to project ToCs, these activities contribute to: (a) intermediate outcomes, including improved attendance, quality of teaching, learning environments, girls' life skills, self-esteem and aspirations, economic empowerment (such as income-generation and asset building skills), parental and community attitudes and behaviours, and school leadership and management; which contribute to (b) the GEC's overarching outcomes: learning, transition, and sustainability.

We mapped all girl and primary caregiver survey questions to intermediate outcome areas tracked by projects; and selected those tracked by most projects. The following questions mapped to intermediate outcomes are explored:

- *Attendance*: Since the start of the most recent school year, has [GIRL] attended her (main) school on most days that the school was open?⁵⁸

- *Economic empowerment*: Is it difficult to afford for [GIRL] to go to school?⁵⁹

- *Life skills*: In the absence of a standard, robust and reliable scale for assessing specific life skills across the portfolio, we mapped survey questions to several life skills constructs.⁶⁰ Among them are:

- *Academic self-confidence*: I get nervous when I have to read in front of others; I get nervous when I have to do maths in front of others; I feel confident answering questions in class; I ask the teacher if I don't understand something; when I succeed at school it is because I worked hard; If I do well in a test it is because I am lucky; I can describe my thoughts to others when I speak

- *Self-efficacy*: I can stay focused on a goal despite things getting in the way, I can put a plan in place and stick with it

- *Motivation and goal orientation*: I would like to continue studying/ attending school after this year

- *Decision-making*: I recognise when choices I make today about my studies can affect my life in the future

- *Leadership and communication*: I can work well in a group with other people; When I have the opportunity, I can organise my peers or friends to do an activity

- *School-based loneliness*: I often feel lonely at school.

- *Teaching quality*: we mapped survey questions asked to girls against factors that determine the quality of teaching for marginalised adolescent girls according to the Quality Teaching Framework⁶¹, including learning environments that are safe, inclusive, conducive to learning, and empowering.⁶² Specifically, we mapped the following questions:

- *Welcoming environment*: My teachers make me feel welcome in the classroom

- *Gender-sensitive pedagogy*: My teachers treat boys and girls differently in the classroom; Does your teacher(s) ask more questions to: boys, girls, or both; Does your teacher(s) ask harder questions to: boys, girls, or both

⁵⁷ This is calculated using the full sample.

⁵⁸ Derived using caregivers' surveys.

⁵⁹ Derived using caregivers' surveys

⁶⁰ Derived using girls' surveys

⁶¹ [gec_qtf_december_2020.pdf \(girlseducationchallenge.org\)](#)

⁶² Derived using girls' surveys

- *Attendance*: My teachers are often absent for class
- *Child-centred practices*: If you don't understand something, do your teachers use a different language to help you understand?; Does your teacher(s) suggest ways you can continue to study after school/at home?
- *Interactive learning*: Does your teacher(s) encourage students to participate during lessons, for example by answering questions?
- *School corporal punishment*: Do your teachers discipline or punish students who get things wrong in a lesson?; In that week, did you see a teacher use physical punishment on other students?; In that week, did the teacher use physical punishment on you?.

- *School related gender-based violence*⁶³: Do you feel safe travelling to and from school?; How safe or unsafe is it for girls to travel to schools in this area?.

- *Parental attitudes*: What level of schooling would you like [GIRL] to achieve?; Do you listen to the views of [GIRL] when you make decisions about her education or are these decisions made by adult members of the family only?; To what extent do you agree that "even when funds are limited it is worth investing in [GIRL]'s education"; To what extent do you agree "a girl is just as likely to use her education as a boy".⁶⁴

We then examined whether girls in the project schools improved over and above comparison girls with respect to these intermediate outcomes.

Analysis of transition outcomes

Transition is defined as progression through key stages of education including training, vocational education or employment. Table 6 summarises how projects defined transition for primary and secondary school girls and for out-of-school girls.

To track transition, we need information on the status of girls (e.g., whether they are in school, employment or Technical and Vocational Education and Training (TVET)) at two data points in time (baseline and midline). For this, we examine transition outcomes with the sample of girls that were re-contacted at midline from baseline.⁶⁵

We define success as defined by projects (in bold in Table 6) - a successful transition is defined as the girl progressing one or two grades⁶⁶ between baseline and midline, and an unsuccessful transition is when a girl repeats a grade or drops out of school.

Table 6: Definition of successful and unsuccessful transition

	Successful transition	Unsuccessful transition
Primary	<ul style="list-style-type: none"> - In-school progression - Moves into secondary school 	<ul style="list-style-type: none"> - Drops out of school - Repeats same grade - Enrolled in informal school (e.g., madrasa) - Moves into employment
Secondary	<ul style="list-style-type: none"> - In-school progression - Enrols in TVET - Moves into age-appropriate, safe and fairly paid employment 	<ul style="list-style-type: none"> - Drops out of school - Repeats same grade - Enrolled in informal school (e.g. madrasa) - Moves into employment, but not age-appropriate, fairly paid or safe

⁶³ Derived using girls' and caregivers' surveys

⁶⁴ Derived using caregivers' surveys

⁶⁵ Irrespective of whether girls were sampled by projects for learning or transition cohorts. As discussed before, we do not distinguish between original groupings, but use all available data.

⁶⁶ We distinguish between progression and partial progression, with partial progression achieved when girls transition one grade ahead of their baseline grade when they were intended to progress two or more grades.

	Successful transition	Unsuccessful transition
Out of School (OOS)	<ul style="list-style-type: none"> - Re-enrol in school - Re-enrol in school with conditions (married, working, training) - Enrols into technical & vocational education & training (TVET) - Started/continued business/ working as a wage labourer - Positive Migration (e.g. migration for a better job, education, social opportunities) 	<ul style="list-style-type: none"> - Remains out of school - Drop-out (If the OOS/SG girl was enrolled in school last year) - Engaged in unpaid domestic work only - Has been doing nothing - Negative migration (e.g. trafficking, due to natural disasters etc.) - Early marriage/childbirth under 20 years

However, in Study 3, as most girls re-contacted were found in school, the transition analysis offers limited value. Inclusion of only in-school girls implies that the transition outcomes are restricted to only include progression in school (i.e., for girls who were enrolled in school at baseline) and progression into school (i.e., for girls who were out of school at baseline)⁶⁷. This disregards girls who may not be re-contacted because they dropped out and could not be traced. **As a result, transition rates using only re-contacted girls will overestimate the portfolio success.**

With this, in addition to transition outcomes for re-contacted girls, we adjusted the transition analysis to account for attrition, or girls who could not be re-contacted at midline. To account for attrition, some assumptions had to be made, since the status of girls not re-contacted at midline is largely unknown⁶⁸. Based on project documents, Table 33 in Annex 3 provides the status of girls for some projects for which data are available⁶⁹.

Transition rates of those who could not be recontacted vary by project. Based on projects' documentation, on average, the reported successful transition rate is 20%, followed by a 47% unsuccessful transition rate, and 33% unknown transition rate (see Table 33 in Annex 3). By applying these percentages to lost girls, we calculated a separate transition rate (adjusted for attrition)⁷⁰. First, we calculated the number of lost girls for each project. The percentages indicated above were then applied to these girls to approximate the number of girls who were not re-contacted by each transition outcome – i.e., *successful*, *unsuccessful*, and *unknown*. Then, to estimate attrition-adjusted transition rates, we added together the number of girls re-contacted and lost (based on the assumptions discussed) with successful and unsuccessful transition and divided by the total number of girls in the sample. This provides a more realistic and conservative rate, as the percentage of girls lost to attrition with unsuccessful transition may be larger due to those classified as unknown.

It is important to note that attrition for some projects stems from their decision to not track a subset of girls⁷¹. As the hypothesised transition rates may not apply to them, we calculated transition rates excluding these girls for comparison purposes.

Limitations

GEC evaluation design

1. Each GEC-T project operates in a different country and local context, with target populations of different age, marginalisation levels, social and economic characteristics, and intervenes through a range of different programme activities. **This makes it impossible when comparing the effect that each project has had on learning (or transition) outcomes, to attribute observed differences to any specific contextual factor, or to the effectiveness of projects' interventions.**
2. We are limited by the data collected, and there are several constraints from combining data from different sources. As projects followed different targeting strategies and sampling strategies, the samples differ (different girls and different projects) when we look at different factors (variables). This means there is a

⁶⁷ Transition pathways outside sample schools (into TVET or employment) are mostly documented in midline reports but are not available in the datasets for the percentage of girls that were not re-contacted.

⁶⁸ It would not be realistic to assume that all girls not re-contacted have unsuccessful transition as the most common reason for attrition is migrating outside project area (see **Error! Reference source not found.**). Girls could be continuing education after they had migrated.

⁶⁹ To understand status of lost girls, project documents were examined if there are any indications of what the lost girls were doing. Some projects, such as WUSC and Relief International, surveyed schools and neighbours of the reasons girls could not be recontacted. This offers some insights into girls' statuses. Other projects provided rationale and potential pathways for the lost girls (for example, if they are likely to have dropped out of school or transferred to a new school outside project area).

⁷⁰ Attrition-adjusted transition is calculated for only in-school girls, as out-of-school girls only account for 2% of the sample.

⁷¹ These include 4 projects. Camfed International Tanzania and Zimbabwe do not track Form 4 (grade 11) girls. CARE tracked only girls who are in school in baseline, and a subset of schools (excluding 13 schools due to various reasons). LC tracked only treatment group. Relief International excluded all schools in Banadir region from tracking.

trade-off between the breadth of indicators and the consistency of estimations. It also means that when an extensive set of factors is combined, our sample is not representative at the portfolio level.

3. For similar reasons, changes in outcomes at the subgroup level are highly prone to sample composition effects: for example, the average in-school girl and the average out-of-school girl do not only differ in their enrolment status, but in other key characteristics such as their age or socioeconomic backgrounds, and most importantly, may not come from the same project target populations⁷². In other words, **cross-project comparisons cannot fully account for all these differences statistically, so any portfolio-level estimates need to be interpreted with these caveats in mind.**

Learning analysis

4. **Different assessments and subtasks that are administered in different projects make projects not directly comparable.** For example, Camfed Tanzania administered only SeGRA and SeGMA at both baseline and midline while Camfed Zambia additionally conducted EGRA and EGMA. As another example, Care International conducted EGRA and EGMA, but only a few subtasks for EGRA (i.e., familiar word, oral reading fluency, reading comprehension) were developed. Sometimes, within a project, different subtasks were used for different groups of girls depending on their grade levels or enrolment status. Therefore, it is important to interpret aggregate literacy and numeracy scores with caution, since the number and type of subtasks included in the aggregate scores varies from project to project and from girl to girl.
5. **Some projects do not have timing variables for reading fluency, such as time remaining and time taken, or have incomplete variables (e.g., only a few girls have records of time taken).** When such cases were identified, the following steps were taken to maximise sample sizes. First, the wpm score was checked to see if the variable had valid data. If so, it was assumed that the girls had spent the whole time given and recalculated the wpm scores based on the raw scores (number of words that girls have read) and time taken.
6. **Reading wpm capped at 100 creates a ceiling effect.** For the reading fluency task, as per FM guidance, projects recorded their raw scores, words per minute scores, and words per minute scores that are capped at 100. The analysis uses the capped scores, as most datasets which were submitted by projects do not include the non-capped variables. However, the capped scores limit the ability to measure changes (progress) in learning, as they include girls who have read more than 100 words per minute.
7. Due to the loss of sample experienced when different sources of data are combined, the analysis of predictors of change in learning was limited to examining the change in intermediate outcomes over and above the comparison group and looking at individual correlations between positive changes in intermediate outcomes and learning. The data was not fit for multivariate model that remains representative at the portfolio level.

Correlation analysis between Intermediate Outcomes and Learning

8. **The correlation analysis between intermediate outcomes and learning is limited to the available data. As such, any portfolio level conclusions need to be treated with caution.**

In the TORs we proposed conducting an analysis of predictors of change in learning, whereby predictors were integrated into a multivariate regression model. However, as detailed in Annex 1 under the heading 'Deviation from the TORs', this was not feasible due to data limitations. To examine how changes in intermediate outcomes correlate with changes in learning, both IO data and learning data need to be available for the same girl in the two evaluation rounds. However, as shown in Methodology and discussed in the Limitation No. 2, we lose more than 90% of the sample and various projects when we try to fit such a multivariate model with information on various intermediate outcomes.

Instead, it was possible to examine the change in intermediate outcomes over and above the comparison group and we conduct this analysis with the panel sample of treated girls – same girls in baseline and midline. In addition, we examined how a positive change in IO between baseline and midline correlates with a change in learning. This analysis, however, still led to projects dropping out from the analysis (e.g. those who did not collect IO data and learning data for the same girls) and sample size reduction for projects kept (e.g. when both types of data is only available for a subset of girls).

⁷² Typically, more than 95% of the out-of-school girls of the GEC-T panel sample come from the same project, Mercy Corps Nigeria.

There are 20 projects with learning information available for secondary school level Maths (SeGMA) and 17 projects with learning information available in the two evaluation rounds for treated girls for reading (EGRA, SeGRA), and primary school level Maths (EGMA). When we combine learning data with IO data, to correlate the improvements in IO factors with changes in learning, There are 20 projects with learning information available for secondary school level Maths (SeGMA). the number of projects is reduced depending on the learning test used and the IO being inspected – it reduces to 3-14 projects for EGRA, 4-15 projects for SeGRA, 3-14 projects EGMA, and 4-16 projects for SeGMA, respectively.

As noted above, there is a further drop in the sample size as the IO factor may not be linked with learning data within projects for all girls. As such the share of girls with information available for both IO factor and learning for projects that are included in the correlation analysis may be as low as 2% (SeGRA, EDT). Conversely, there are projects which link IO factors and learning data for all or most of the girls (e.g., AKF, BRAC). Importantly, while some projects display a high share of girls with available data, the sample size of girls with learning data is very low to begin with (HPA, ICL).

Therefore, the sample size available for analysis of the correlation between positive changes in intermediate outcome factors and changes in learning between baseline and midline is limited in terms of drawing portfolio-level conclusions.

Transition analysis

9. Transition analysis can only be examined with accuracy for girls who have been re-contacted at midline. As a result, it suffers from attrition bias because girls who were not re-contacted are more marginalised, and thus likely to have higher unsuccessful transition rates than those re-contacted. Projects often reported transition figures based on the number of girls they could re-contact. BRAC, for example, considered girls lost to attrition as having an unsuccessful transition.
10. Girls in the recontacted sample are largely in school, and as such, we separated the analysis on transition to progression through school (when girls were sampled in schools), and progression into school (when girls were sampled out-of-school). Overall, the findings suggest a more optimistic picture than in reality because girls who likely dropped out of school are unaccounted for. We conducted transition analysis that accounts for attrition based on project documents. Nonetheless, the figures are only as good as the assumptions that are behind it. Only nine projects reported estimates of lost girls. For other projects, status of lost girls is largely unknown.

3. Changes in Girls' Learning

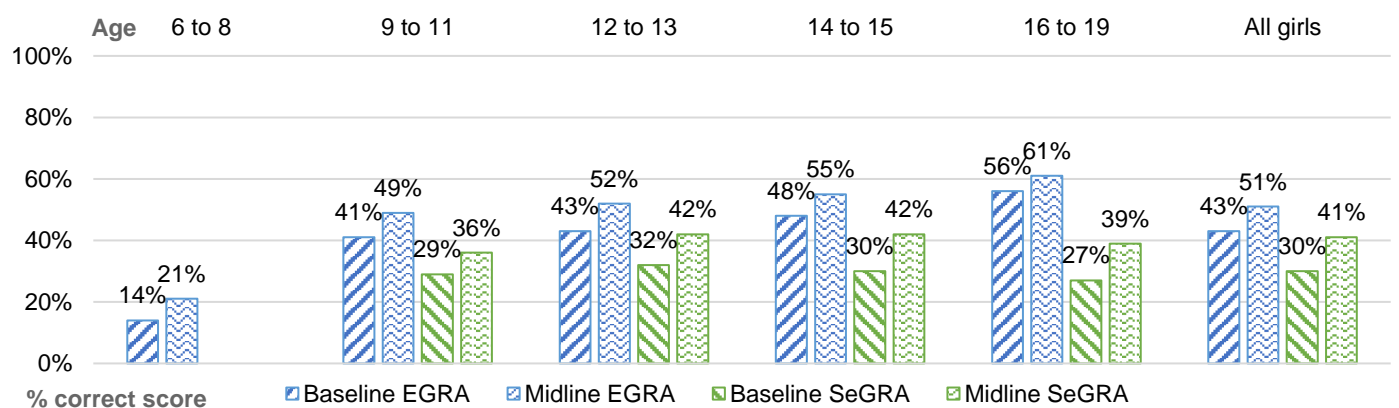
Box 1: Key learning findings

1. The literacy and numeracy levels of GEC-T girls statistically significantly improved between baseline and midline. They did so in each of the four assessments: EGRA and SeGRA (literacy), EGMA and SeGMA (numeracy)⁷³.
2. The average GEC-T girl has learned significantly more⁷⁴ than the average comparison girl between baseline and midline (when weighting projects by the size of their beneficiary population). However, the average increase achieved by GEC-T girls *over and above the comparison group* is small.
3. The observed increase in learning is driven by large projects – three projects with a positive change in learning over and above comparison groups cover three quarters of beneficiaries. When weighting projects equally, the GEC-T did not improve learning more than in comparison areas, with many GEC-T projects showing similar increases in treatment and comparison groups.
4. At midline, about one-fifth of GEC-T girls are still unable to read a familiar or invented word correctly. Among older girls who have been given SeGRA and SeGMA, more than one-fifth score zero in short essay construction, about one-third score zero in algebra and one-half in sophisticated word problems. This means that many girls still do not have the most basic literacy and numeracy skills after several years of GEC intervention.
5. Despite the observed progress in girls' literacy scores, most GEC-T girls still fall short of international standards of reading fluency and comprehension benchmarks at midline⁷⁵.
6. Changes in learning are consistent across key subgroups. However, some have progressed less than others between baseline and midline, such as girls who are overage for their grade, girls living in poor households and girls whose primary caregiver or head of household has no education (for literacy and numeracy), as well as girls whose father or mother do not live in the household (for numeracy only).

1. The learning levels of GEC-T girls improved in literacy and numeracy between baseline and midline. They did so in each of the four assessments: EGRA and SeGRA (literacy), EGMA and SeGMA (numeracy).

GEC-T girls' learning levels (of girls in the treatment group), expressed as the percentage correct score across the EGRA/ SeGRA and EGMA/ SeGMA improved across all four assessments and all age categories between baseline and midline.

Figure 3: Literacy progress of GEC-T girls, by age at baseline (% correct score)



⁷³ The baseline to midline increase for GEC-T girls is statistically significant at the 95% level for all four assessments, with P-values of 0.000.

⁷⁴ P-values for panel analysis: EGRA = 0.037, SeGRA = 0.056, EGMA = 0.021 and SeGMA = 0.069.

P-values for cross-sectional (CS) analysis: EGRA = 0.022, SeGRA = 0.042, EGMA = 0.036 and SeGMA = 0.116 (not significant).

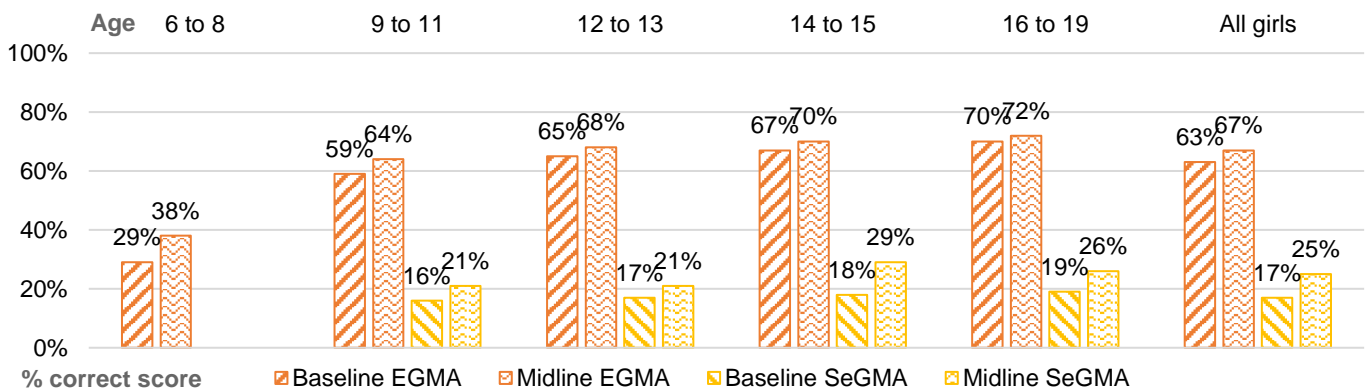
⁷⁵ For oral reading fluency in English, we use the RTI-defined threshold of 45-60 wpm in international contexts. For reading comprehension, we use the UNESCO Global Coalition standard that "all children should be able to read and understand a short, age-appropriate text by age 10". See Box 3 for more details on the benchmarks used in this study. It is worth noting that the majority of children in low and middle income countries fall behind international benchmarks, as shown in UNESCO's latest Global Education Monitoring Report (See <https://www.unesco.org/gem-report/en>).

Note: EGRA bars are shown in blue, SeGRA bars in green. The left bar shows the average baseline score, the right bar shows the average midline score. There are no SeGRA bars for girls aged 6 to 8 because not enough of them have been given the assessment (see Box 7 for further details on the population base of learning assessments).

As shown in Figure 3, GEC-T girls' literacy scores have improved consistently by 5 percentage points to 12 percentage points across all age categories (see Box 2 below for further details about how to interpret percentage scores and increases).

Similarly, girls' numeracy scores have improved across all age categories, although to a slightly lesser extent than literacy scores as shown in Figure 4: by 2 to 9 percentage points in EGMA, by 4 to 11 percentage points in SeGMA. The smaller increase in EGMA scores is due to a ceiling effect in early subtasks, with many girls already scoring close to maximum scores at baseline, which limited their ability to increase their scores at midline (see Table 45 in Annex 9 for the full baseline and midline learning levels of girls by subtask).

Figure 4: Numeracy progress of GEC-T girls, by age at baseline



Note: EGMA bars are shown in orange, SeGMA bars in yellow. The left bar shows the average baseline score, the right bar shows the average midline score. There are no SeGMA bars for girls aged 6 to 8 because not enough of them have been given the assessment (see Box 3 for further details on the population base of learning assessments).

Box 2: How to interpret percentage correct scores and percentage-point increases

Learning score levels are expressed as *percentage correct* and their changes are expressed in *percentage points*⁷⁶. For example, in EGRA:

- The typical *letter sound identification* subtask includes 100 items (sounds), marked 1 when correct, 0 when incorrect. The percentage correct score is therefore equal to the number of sounds correctly identified. An increase of 2 percentage points means that GEC-T girls' increase in score was of 2 additional letter sounds read correctly.
- The typical *familiar word reading* subtask includes 50 items (words), marked 1 when correct, and 0 when incorrect. The percentage correct score is therefore equal to twice the number of familiar words correctly read. An increase of 2 percentage points means that GEC-T girls' increase in score was of 1 additional word read correctly.

The *oral reading fluency* subtask is scored in words-per-minute (wpm), which is the total number of correct words read in a minute out of a short story/passage written in the language of instruction. This score has no maximum in theory, but projects capped it to 100 wpm following FM guidance. The percentage correct score is therefore equal to the number of correct wpm, capped at 100⁷⁷. An increase of 2 percentage points means that GEC-T girls' increase in score was 2 additional words-per-minute (wpm).

⁷⁶ This is equal to the percentage of correct marks out of the total number of marks (maximum possible score). An increase of 2 percentage points therefore corresponds to going from a 18% correct score to a 20% correct score, from a 50% correct score to a 52% correct score, or from a 95% correct score to a 97% correct score.

⁷⁷ The capping of the wpm score, along with ceiling effects on other subtasks, may lead to partly underestimate the overall learning gains of girls between baseline and midline, and to underestimate the gains of treatment girls as opposed to comparison girls. At midline, 36% of treatment girls (among those who were given the EGRA oral reading fluency) scored the maximum score of 100 wpm, as opposed to 24% of comparison girls. At baseline, the share was 19% in both groups.

2. The average GEC-T girl has learned significantly more than the average comparison girl between baseline and midline (when weighting projects by the size of their beneficiary population). However, the average increase achieved by GEC-T girls over and above the comparison group is small.

The average GEC-T girl has improved in literacy (EGRA and SeGRA) and numeracy (EGMA) more than the average comparison girl. These improvements, measured by difference-in-difference (DID), are statistically significant at the 95% confidence level for EGRA (Panel *P*-value = 0.037 and CS *P*-value = 0.022) and EGMA (Panel *P*-value = 0.021 and CS *P*-value = 0.036), at the 90% confidence level for SeGRA in the panel sample (*P*-value = 0.056) and at the 95% confidence level in the CS sample (*P*-value = 0.042). For SeGMA, only the panel sample shows a statistically significant improvement (*P*-value = 0.069).

As shown in Table 7, the scale of girls' improvement is +2 pp (percentage points) in literacy and +3 pp in numeracy for the panel sample of recontacted girls. It is of similar scale across the cross-sectional sample (all girls), although slightly higher for EGRA (+ 4 pp instead of +2 for the panel sample) and slightly lower for SeGMA (+2 pp instead of +3 for the panel sample).

Table 7: Learning improvements of the average GEC-T girl over and above comparison (beneficiary population weights)

Difference-in-difference <i>in percentage points</i>		Literacy		Numeracy	
		EGRA	SeGRA	EGMA	SeGMA
Panel sample (recontacted girls only)	% correct score	+2**	+2*	+3**	+3*
	<i>P</i> -value	0.037	0.056	0.021	0.069
Cross-sectional sample (all girls)	% correct score	+4**	+2**	+3**	+2
	<i>P</i> -value	0.022	0.042	0.036	0.116

Key: DID coefficients with two asterisks are statistically significant at the 95% confidence level (*P*-value lower than 0.05 = 5%). Those with one asterisk are statistically significant at the 90% level (*P*-value lower than 0.1 = 10%). The actual *P*-value is reported in the table.

The observed increase in learning over and above the comparison groups, however statistically significant, is relatively small in size⁷⁸. GEC-T girls have improved by only two to three percentage points more than girls in the comparison group over eighteen months⁷⁹, the average between the baseline and midline data collection timepoints across the GEC-T.

Aggregate increases at the test level hide disparities in the way GEC-T girls have improved in specific subtasks. Table 8 breaks down girls' increases over and above the comparison group across all literacy subtasks.

Table 8: Literacy improvements of GEC-T girls over and above comparison girls - % correct scores

Difference-in-difference <i>in percentage points</i>		EGRA						SeGRA			
		Letter sound ident.	Familiar word	Invented word	Oral reading fluency	Reading comp.	Overall EGRA	Short reading comp.	Longer reading comp.	Short essay constr.	Overall SeGRA
Panel sample	% correct	+5	+1	+2**	+3**	+1	+2**	+2	+3*	+3**	+2*
	<i>P</i> -value	0.146	0.525	0.046	0.042	0.350	0.037	0.349	0.056	0.001	0.056
	N (per round)	11,777	15,327	13,553	12,362	13,960	16,215	18,579	16,408	13,399	19,178

Key: DID coefficients with two asterisks are statistically significant at the 95% confidence level (*P*-value lower than 0.05 = 5%). Those with one asterisk are statistically significant at the 90% level (*P*-value lower than 0.1 = 10%). The actual *P*-value is reported in the table.

⁷⁸ Statistical significance is as much a reflection of the magnitude of an effect as of the sample base on which this effect is observed. Working with large samples increases the likelihood of finding more statistically significant effects, which does not mean that the magnitude of these effects is large. Because these difference-in-difference have been calculated on relatively large samples (30,000 girls at baseline for the panel sample, 57,000 girls at baseline and 48,000 at midline for the cross-sectional sample), they are statistically significant, even though they are small.

⁷⁹ Eighteen months is the average length between the baseline and midline data collection timepoints across the GEC-T.

The overall EGRA increase reported in Table 8 is therefore mostly a result of the increase in two EGRA subtasks: invented word and oral reading fluency, in which GEC-T girls improved by +2 pp and + 3 pp over and above comparison girls. The increase in letter sound identification is also large, although not statistically significant. For SeGRA, the largest effects are observed in the subtasks of longer reading comprehension and short essay construction.

Table 9: Numeracy improvements of GEC-T girls over and above comparison girls - % correct scores

Difference-in-difference in percentage points		EGMA									SeGMA			
		Number ident.	Quantity Discrim.	Missing Num.	Addition Level 1	Addition Level 2	Subtract. Level 1	Subtract. Level 2	Word Prob.	Overall EGMA	Mult./div., fractions, geometry	Algebra	Sophist. word problems	Overall SeGMA
Panel sample	% correct	+6*	+4	+1	+3	+1	+2	-1	+2	+3**	+5**	+5**	+4**	+3*
	P-value	0.053	0.107	0.649	0.239	0.409	0.282	0.684	0.300	0.021	0.026	0.001	0.000	0.069
	N (per round)	9,401	10,586	15,933	13,206	6,520	14,948	7,667	16,396	16,458	22,242	15,980	12,559	23,011

Key: DID coefficients with two asterisks are statistically significant at the 95% confidence level (P -value lower than 0.05 = 5%). Those with one asterisk are statistically significant at the 90% level (P -value lower than 0.1 = 10%). P -value of 0.000 does not mean the P -value is zero but less than 0.001. The actual P -value is reported in the table.

The increase in EGMA is mostly explained by improvements in early subtasks. This suggests that GEC-T girls who were given the EGMA tests (girls of younger age and in early grades, as opposed to SeGMA which has been given to older girls and girls in higher grades, see Box 7 in Annex 9) have improved more in the most basic numeracy skills compared to girls in the comparison group. Conversely, girls who have been given the SeGMA test have significantly improved (P -values = 0.026, 0.001, 0.000 for each SeGMA subtask) more than comparison girls across the three subtasks.



*It is possible that the DID coefficient of the aggregate (or overall) assessment score is higher, or lower, than all its subtask scores, as shown with SeGMA above. This comes from a **subtask composition effect**. Indeed, girls have not been given the same subtask at baseline and at midline: some have been dropped due to ceiling or floor effects, or girls progressing through higher grades have been given more difficult subtasks. The aggregate score of a girl may therefore not be directly comparable between baseline and midline, meaning greater weight should be put on the DiD*

estimates compared to the first difference (which might not see a change even when children are learning). When possible, the learning findings are displayed at the subtask level. When this is not possible to show all subtasks separately, only overall scores are described, and subtask-level findings are reported in annex.

The effect of the GEC-T across the portfolio hides important project-level differences in the baseline levels of girls and the midline levels they reached (see Figure 12 in Annex 9 for the average midline scores of the worst and best-performing project in each subtask). The next point explores project-level disparities across the GEC-T portfolio.

3. The observed increase in learning is driven by large projects – three projects with a positive change in learning over and above comparison groups cover three-quarters of beneficiaries. When weighting projects equally, the GEC-T did not improve learning more than in comparison areas, with many GEC-T projects showing similar increases in treatment and comparison groups.

Given the differences in sample size between projects, we report the estimates using project level weights. Difference-in-difference calculations are rerun by attributing ‘inverse sample size’ weights to each GEC-T project so they are all weighted the same in portfolio averages, irrespective of their sample sizes. This allows us to calculate the average learning improvement across GEC-T projects (so calculating the average change for each project; then averaging across the projects) rather than the improvement of the average GEC-T girl.

As shown in Table 10, this gives a very small, insignificant increase in learning scores over and above the comparison group. For the panel sample, the overall scores from the four assessments have increased by 1 percentage-point or less. The cross-sectional sample shows similar patterns, despite a higher (but still not statistically significant) increase in EGRA scores.

Table 10: Average learning improvements of GEC-T projects over and above comparison groups (project-equal weights)

Difference-in-difference (DID) in percentage points		Literacy		Numeracy	
		EGRA	SeGRA	EGMA	SeGMA
Panel sample (recontacted girls only)	% correct score	0	0	+1	0
	<i>P</i> -value	0.969	0.964	0.645	0.934
Cross-sectional sample (all girls)	% correct score	+4	0	0	0
	<i>P</i> -value	0.259	0.789	0.938	0.972

Key: Difference-in-difference (DID) coefficients with two asterisks are statistically significant at the 95% confidence level (*P*-value lower than 0.05 = 5%). Those with one asterisk are statistically significant at the 90% level (*P*-value lower than 0.1 = 10%). The actual *P*-value is reported in the table.

The absence of statistically significant improvement over and above the comparison girls when weighting each GEC-T project equally suggests that learning improvements are driven by a few large projects, while in other projects GEC-T girls have not improved more than comparison girls on average.

This is confirmed by the difference-in-difference coefficients shown in Table 48 in Annex 9, which shows substantial variation across projects. The largest increases in learning scores over and above the comparison group are indeed observed among some of the projects with the largest number of GEC-T beneficiaries.

Three projects are prominent in this respect: Camfed International (26% of the total number of GEC-T learning beneficiaries); DLA Ghana-Kenya-Nigeria (which supports 43% of the beneficiaries); and EDT Kenya which supports 7%. These three implementing partners provide three quarters of the beneficiaries and have relatively consistent performance: they all positively performed in EGRA, SeGRA and EGMA (except Camfed Zambia for SeGRA, which has a negative result but who represents only 1% of all GEC-T beneficiaries). The GEC-T girls of DLA Nigeria, which make up 19% of the portfolio beneficiaries, improved by +4 pp for EGRA and +8 pp for EGMA over and above comparison girls. The same is true of Camfed Zimbabwe, which has the second-largest share of beneficiaries (16%), and whose treatment girls in the project improved by +2 pp for SEGRA and +8 pp for SEGMA.

Some other projects, such as AKF Afghanistan and Mercy Corps Nepal and Nigeria show substantial increases in +3 pp or more in at least one of the four assessments. But each of them only account for 2% or less of all the GEC-T beneficiaries.

A range of projects of relatively small size shows no effects or negative effects compared to their comparison group. This is the case for BRAC Afghanistan, Cheshire Services Uganda, CARE Somalia, HPA Rwanda, ICL Kenya, PEAS Uganda, Viva Uganda and World Vision Zimbabwe. None of these projects account for more than 5% of the overall population of the GEC-T learning beneficiaries.

Two projects do not have a comparison group and two projects have not tracked girls between baseline and midline and have therefore been excluded from the panel sample. The same table based on the cross-sectional sample, which includes Varkey Foundation Ghana and VSO Nepal, is shown in Table 49 in Annex 9.



As mentioned in the methodology, two types of weights have been used in the analysis of learning outcomes: beneficiary-population weights, which weight projects proportionally to the size of their learning beneficiary population; and project-equal weights, which weight each project equally. By default, i.e., when not explicitly mentioned, the learning findings are calculated on the panel sample (recontacted girls only), using beneficiary-population weights.

Where relevant, we have cross-checked the robustness of our findings by running the same calculations on the cross-sectional sample (as shown in Table 10) or by using project-level weights. When not included in the main body, findings from these robustness checks are reported in annex.

4. At midline, many GEC-T girls are still unable to read a familiar or invented word correctly. Among older girls who have been given SeGRA and SeGMA, more than one-fifth score zero in short essay construction, about one-third score zero in algebra and one-half in sophisticated word problems. This means that many of the girls still do not have the most basic literacy and numeracy skills after several years of GEC intervention.

As shown in Table 11 below, the share of girls scoring zero (i.e., the girls who scored all the items wrong or where unable to provide a response) has not decreased significantly across the GEC-T when compared to the comparison group. The pattern is the same across the panel and the cross-sectional samples. The prevalence of EGRA zero scorers has decreased (however not significantly) by 4 percentage points, mostly through decreases in the zero scores of the most basic 'easiest' subtasks of letter sound identification, familiar word and invented word. However, the EGMA score shows a slight increase in the share of zero scorers compared to comparison areas, again under the effect of the most basic subtasks of number identification and quantity discrimination. Table 46 in Annex 9 shows the difference-in-difference of zero scores by subtask.

Table 11: Change in the prevalence of zero scores over and above the comparison group

Difference-in-difference (DID) in percentage points		Literacy		Numeracy	
		EGRA	SeGRA	EGMA	SeGMA
Panel sample (recontacted girls only)	% zero score	-4	-1	+1	-1
	<i>P</i> -value	0.267	0.221	0.181	0.660
Cross-sectional sample (all girls)	% zero score	-4	0	+1	+2
	<i>P</i> -value	0.287	0.457	0.096	0.276

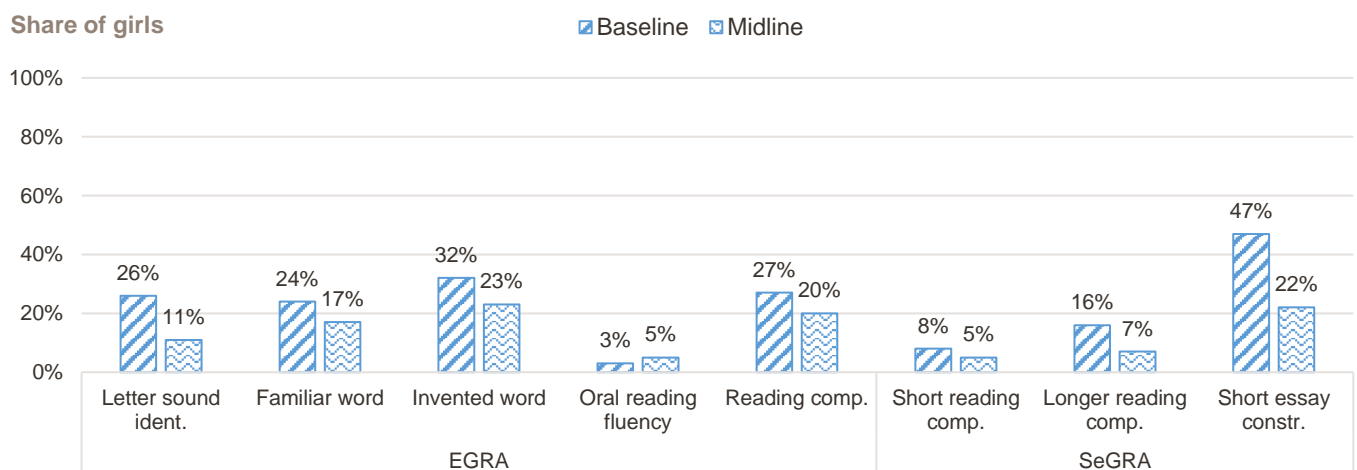
Key: The actual *P*-value is reported in the table.

Focusing on GEC-T girls only and their progress between baseline and midline, the share of girls scoring zero has decreased across almost all subtasks (see Figure 5). The only exception is oral reading fluency whose prevalence of zero scores *increased* but was already very low at baseline: very few girls scored zero words correctly within the allowed timeframe of sixty seconds.

The largest decreases are measured for the EGRA letter sound (minus 15 percentage points) and for SeGRA short essay construction (minus 25 percentage points, with almost half of girls scoring zero at baseline).

The smallest decreases (excluding oral reading fluency) are for the EGRA reading comprehension and familiar word subtasks (minus 7 percentage points). While the prevalence of girls scoring zero in SeGRA short reading comprehension is low at midline (5% of tested girls, minus 3 percentage points), more than one-fifth of girls tested in EGRA reading comprehension continue to score zero. This will be explored in further detail as part of the next key finding and put in relation to international benchmarks and standards of literacy and reading fluency.

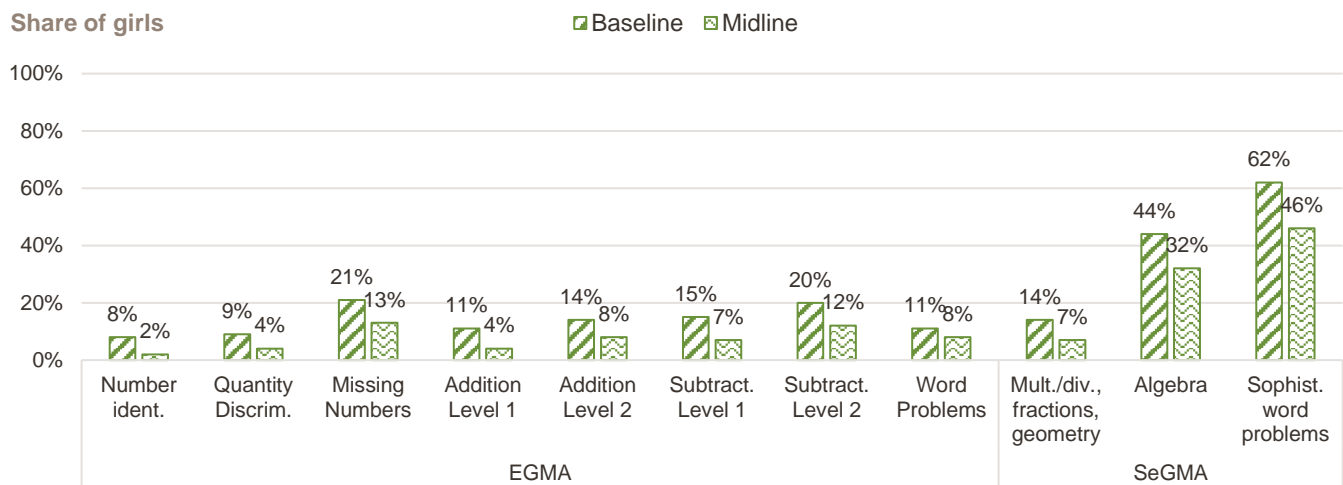
Figure 5: Share of GEC-T girls scoring zero in literacy, at baseline and midline



For numeracy, the pattern is similar overall with a decrease in zero scores across all EGMA and SeGMA subtasks (see Figure 6). For most EGMA subtasks, and the first SeGMA subtask (multiplication, division, fractions and geometry), the prevalence of zero scores was relatively low at baseline, with one-fifth of girls or less scoring only zero. Still, these subtasks show substantial decreases, with the prevalence of zero scores being cut by two or more for most of them.

Conversely, about half of girls were unable to score anything at the second and third standard SeGMA subtasks, namely algebra and sophisticated word problems at baseline. The prevalence of zero scorers has decreased at midline, by 12 percentage points for algebra and 16 percentage points for sophisticated word problems. However, between one-third and half of girls are still unable to score anything at midline. This suggests that, although GEC-T girls master early numeracy skills at least partially and are able to calculate basic operations (addition, subtraction, multiplication and division), the average GEC-T girl is still far from being able to understand and solve algebra or word problems.

Figure 6: Share of GEC-T girls scoring zero in numeracy, at baseline and midline



5. Despite the observed progress in girls' literacy scores, GEC-T girls still fall short of international standards for reading fluency and comprehension benchmarks.

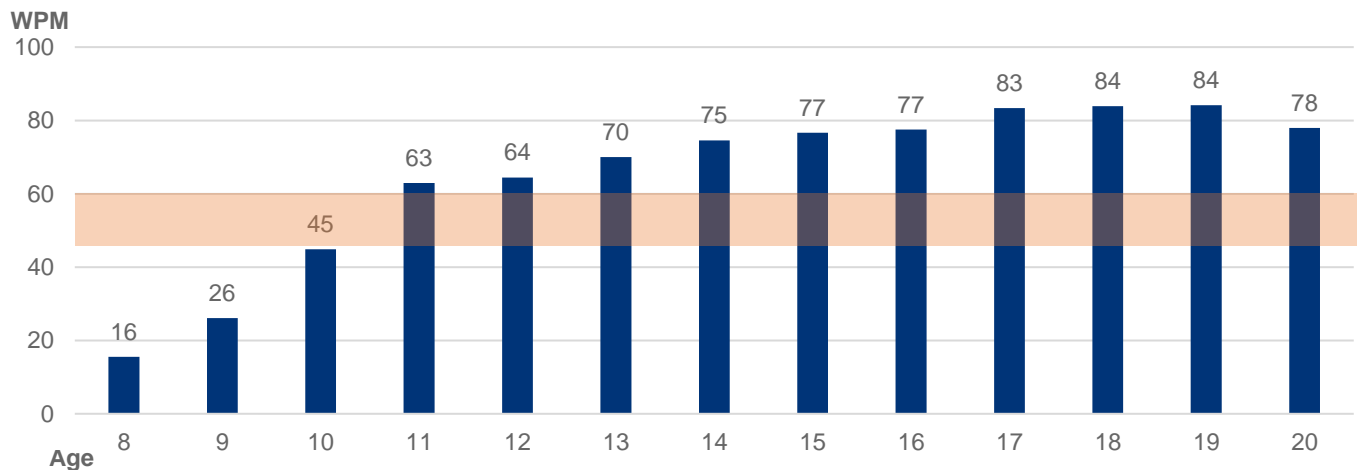
GEC-T girls aged 10 read 45 wpm on average across projects, which is the lower limit of the 45 to 60 wpm oral reading proficiency band in the RTI studies (see Box 3 for a discussion on international benchmarks of literacy) that cover GEC-T countries. This is noticeably below benchmarks for the USA, where the DIBELS scale suggests children should be reading at 100+ WPM by the end of the third grade.⁸⁰

The number of correct words-per-minute read by girls increases at age 11 to reach 63 wpm on average, which exceeds this benchmark. This age corresponds to the first year of secondary school in most systems. Reading gains are substantial between 8 and 11 (from 16 wpm at 8 to 63 wpm at 11), but this progress comes late, and would rather be expected to happen between 5 and 8, at the start of primary school.

Girls' wpm scores then slowly increase until they are 17, after which it stays constant, as shown in Figure 7. This plateauing is however largely due to the capping of wpm scores at 100: by age 12, 31% of tested GEC-T girls attain the maximum score of 100 wpm, a share that reaches 66% at age 18.

⁸⁰ Considered applicable to about half of GEC-T projects who administer tests in English language.

Figure 7: Average words-per-minute read correctly by age, at midline (words-per-minute capped at 100, project-equal weights, cross-sectional sample, treatment group)



Note: the orange band denotes the oral reading proficiency band of 45-60 wpm defined by RTI (see Box 3 for further details). The same graph with average calculated on the *uncapped* words-per-minute score can be found in Figure 13 in Annex 9.

Box 3: On the use of international standards and benchmarks for literacy and numeracy

The EGRA and EGMA are *early grade* assessments that are meant to be given to girls up to grade 3. For this reason, no international benchmarks have been developed for later grades, although many girls of grades 4 to 6 or higher have been assessed using the EGRA and EGMA.

The SeGRA and SeGMA was developed specifically for Phase II of the GEC to avoid the ceiling effects that were observed on some EGRA /EGMA subtasks during Phase I⁸¹. They are not associated with any standard benchmarks or target levels.

The Fund Manager (FM) uses proficiency bands based on the percentage correct scores, defined as follows:

- 0% = Non-Learner / 0-5 wpm = Non-Reader
- 1%-40% = Emergent Learner / 6-44 wpm = Emergent Reader
- 41-80% = Established Learner / 45-80 wpm = Established Reader
- 81%-100% = Proficient Learner / 80 wpm or more = Proficient Reader

These bands, however, are only statistical and do not correspond to real-life literacy or numeracy skills.

Several EGRA studies by the Research Triangle Institute (RTI) have shown that students who correctly read 45-60 wpm are able to understand a simple text, which suggests it is an appropriate benchmark of oral reading fluency in international contexts given the language differences across contexts⁸².

In parallel, the UNESCO-led Global Coalition for Education Data has defined *learning poverty* as “being unable to read and understand a short, age-appropriate text by age 10”. In other words, “all children should read by age 10 – or at the latest, by the end of primary school”⁸³. We apply this benchmark to the EGRA reading comprehension subtask score, assuming that girls who correctly answer strictly more than 3 reading comprehension questions out of 5 are able to understand the text they have read.

Figure 8 shows the share of GEC-T girls who were given the reading comprehension subtask who scored 70% correct or more. This corresponds to correctly answering at least four out of five or six comprehension questions after reading a short passage or text⁸⁴, hence demonstrating students have *understood* what they read.

⁸¹ These tests should be considered as a part of wider work to improve measurement of learning outcomes for children in systems where PISA/TIMSS are evidenced as being too advanced.

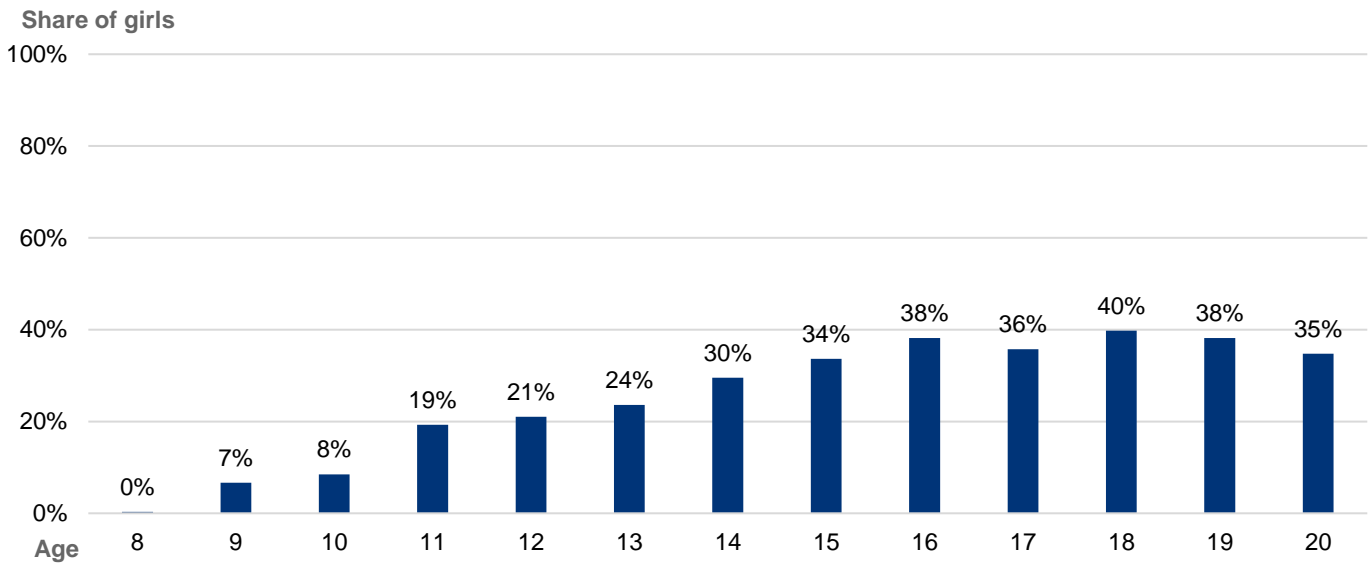
⁸² See [Abadazi \(2011\)](#) and more recently, [USAID RTI « All Children Learning » \(2017\)](#) for oral reading fluency benchmarks in international contexts for the English language. Note that seven out of 27 GEC-T projects have assessed girls' learning in different languages of assessment than English: AKF (Afghanistan) and BRAC (Afghanistan) used Dari and Pashto, CARE International (Somalia) and Relief International (Somalia) used Somali, HPA (Rwanda) used Kinyarwanda, Mercy Corps (Nepal) and VSO (Nepal) used Nepali. Relief International (Somalia) and HPA (Rwanda) also assessed some of the girls in English, but these were not considered as part of the learning analysis.

⁸³ See the official page of the Global Coalition for Education Data by the World Bank and UNESCO: <https://gaml.uis.unesco.org/learning-poverty/>

⁸⁴ Girls who were not able to read a single word correctly were not given the reading comprehension subtask. These were attributed a score of zero to both subtasks and have been included in the sample base for reading comprehension scores.

The share of girls able to understand what they read is correlated to the number of words-per-minute they read, and the sharp increases between age 8 and 11 noted earlier is still observable. The share of girls understanding what they read then increases steadily as they age, but remains small: before the age of 14, less than one-quarter of girls can understand what they read. This is at odds with the World Bank and UNESCO benchmark that children should be able to read and understand what they read by age 10, and shows that at midline, most GEC-T girls are still trapped in 'learning poverty', as defined by the Global Coalition for Education Data.

Figure 8: Share of girls able to read and understand what they read by age, at midline (project-equal weights, cross-sectional sample, treatment group)



Looking at reading comprehension by grade level, only 17% of GEC-T girls are able to understand what they read six years after starting primary school, and only 31% do so seven years after starting primary school (see Figure 14 in Annex 9). This again does not meet the UNESCO standard that “all children should read at the latest by the end of primary school”.

6. Changes in learning are consistent across key subgroups: most show statistically significant learning improvements over and above the comparison groups. However, focusing on GEC-T girls only, some subgroups have progressed less than others between baseline and midline.

The difference-in-difference in learning of key subgroups is similar to that of the overall sample ('the average girl'), i.e., generally positive and of relatively small size. In other words, the GEC-T is associated with slightly improved learning for most subgroups, for at least one of the four assessments, as shown in Table 12 below.

Some subgroups show a better improvement over and above the comparison group than the average beneficiary girl. This is the case for girls in alternative education (except in SeGMA), for girls who are in the grade they should be with respect to their age, and for girls whose mother or father does not live in the household.

Conversely, the GEC-T is associated with the smallest progress, when compared to girls in comparison areas, for the following four key subgroups: girls aged 16-19 at baseline, overage girls (girls who are too old for their grades), girls living in poor households⁸⁵, and girls whose head of household has no education.

Table 12: Difference-in-difference by subgroup at baseline

Difference-in-difference by subgroup, in percentage points		EGRA	SeGRA	EGMA	SeGMA	Share of baseline sample
All ('average girl')	% correct score	2**	+2*	+3**	+3*	100% (N = 29,997)
	P-value	0.037	0.056	0.021	0.069	

⁸⁵ This was defined as households meeting at half or more of the following poverty criteria, among those available from the data: household does not have land, household has gone without cash income most times over the past twelve months, household has gone to sleep at night feeling hungry, household is unable to meet basic needs, household has difficulty paying to send girl to school.

Independent Evaluation of the Girls' Education Challenge Phase II – Aggregate Impact of GEC-T Projects Between Baseline and Midline Study – Final Report

Difference-in-difference by subgroup, in percentage points			EGRA	SeGRA	EGMA	SeGMA	Share of baseline sample	
Age	6 to 8	% correct score	+2**	n/a	+3**	n/a	2%	
		P-value	0.024	n/a	0.045	n/a		
	9 to 11	% correct score	+2**	-1	+5**	0	19%	
		P-value	0.012	0.196	0.011	0.914		
	12 to 13	% correct score	+2*	+3*	+2**	+2	25%	
		P-value	0.077	0.056	0.016	0.212		
	14 to 15	% correct score	+2	+2**	+1	+5**	34%	
		P-value	0.283	0.001	0.639	0.004		
	16 to 19	% correct score	0	+2	-1	+3*	19%	
		P-value	0.920	0.244	0.767	0.052		
	Enrolment	In-school	% correct score	+2	+2**	+3	+3*	93%
			P-value	0.172	0.004	0.104	0.059	
Alternative education		% correct score	+9*	+4**	+5**	0	5%	
		P-value	0.068	0.010	0.020	0.945		
OOS		% correct score	+6	n/a	+3**	n/a	2%	
		P-value	0.144	n/a	0.037	n/a		
School level	Primary	% correct score	+2**	-1	+3**	0	55%	
		P-value	0.027	0.707	0.014	0.740		
	Secondary	% correct score	-1	+3**	-4	+6**	43%	
		P-value	0.145	0.000	0.170	0.000		
Age-grade status	Overage	% correct score	+2	+1	+2	+1	35%	
		P-value	0.189	0.677	0.102	0.326		
	Underage	% correct score	+3**	+4	+3	+4**	10%	
		P-value	0.008	0.100	0.201	0.032		
	Correct age	% correct score	+2**	+2**	+5**	+4**	53%	
		P-value	0.007	0.000	0.005	0.021		
Disability (girl)	Not disabled	% correct score	+2**	+2	+3**	+3*	75%	
		P-value	0.026	0.165	0.015	0.094		
	Disabled (girl - one type)	% correct score	+3	+3*	+3**	+3**	6%	

Independent Evaluation of the Girls' Education Challenge Phase II – Aggregate Impact of GEC-T Projects Between Baseline and Midline Study – Final Report

Difference-in-difference by subgroup, in percentage points			EGRA	SeGRA	EGMA	SeGMA	Share of baseline sample
	Disabled (girl - multiple)	<i>P</i> -value	0.236	0.075	0.006	0.014	2%
		% correct score	n/a	+4	n/a	+6**	
		<i>P</i> -value	n/a	0.169	n/a	0.001	
Disability (PCG)	Not disabled	% correct score	+2**	+1	+4**	+2	47%
		<i>P</i> -value	0.034	0.436	0.021	0.253	
	Disabled (PCG)	% correct score	n/a	n/a	n/a	+4*	2%
		<i>P</i> -value	n/a	n/a	n/a	0.068	
Orphan status	Single orphan	% correct score	+1	+3*	+4**	+4**	9%
		<i>P</i> -value	0.311	0.067	0.015	0.034	
	Double orphan	% correct score	n/a	n/a	n/a	n/a	2%
		<i>P</i> -value	n/a	n/a	n/a	n/a	
Family status	Girl is married	% correct score	n/a	n/a	n/a	n/a	0%
		<i>P</i> -value	n/a	n/a	n/a	n/a	
	Girl is a mother	% correct score	n/a	n/a	n/a	n/a	0%
		<i>P</i> -value	n/a	n/a	n/a	n/a	
	Mother not in HH	% correct score	+3**	+3**	+3*	+2	12%
		<i>P</i> -value	0.005	0.008	0.058	0.134	
	Father not in HH	% correct score	+3**	+2**	+3**	+3*	20%
		<i>P</i> -value	0.007	0.005	0.005	0.087	
Household poverty indicators	Less than 50% (less poor)	% correct score	+3**	+1	+4**	+2	42%
		<i>P</i> -value	0.010	0.584	0.015	0.112	
	More than 50% (poor)	% correct score	+2*	+1	+3*	+1	19%
		<i>P</i> -value	0.086	0.570	0.092	0.413	
Household education	PCG has none	% correct score	+2*	-1	+4**	+1	22%
		<i>P</i> -value	0.086	0.706	0.034	0.346	
	HOH has none	% correct score	+2	-1	+3*	+1	18%
		<i>P</i> -value	0.107	0.715	0.078	0.601	

Key: Difference-in-difference (DID) coefficients with two asterisks are statistically significant at the 95% confidence level (*P*-value lower than 0.05 = 5%). Those with one asterisk are statistically significant at the 90% level (*P*-value lower than 0.1 = 10%). Green cells show positive and statistically significant DID (GEC-T girls have increased significantly more than comparison girls), orange shows negative and statistically significant DID (GEC-T girls have increased significantly less than comparison girls).

Note: subgroups marked “n/a” do not have sufficient sample sizes in the GEC-T portfolio, i.e., too few GEC-T girls belong to them to generate meaningful estimates.



Changes at the subgroup level are very dependent on sample composition effects. Indeed, projects weight differently in each subgroup category: in some, one or two projects sometimes account for all or almost all the subgroup sample. Project-level differences may therefore play an important role in the observed progress of some subgroups. Table 50 in Annex 9 helps contextualise subgroup findings by showing the share of subgroup samples that come from each GEC-T project.

We look at how subgroups progressed in GEC-T areas between baseline and midline, irrespective of what happened in non-GEC areas. This may point to difficulties for specific subgroups to learn, or at least to keep the same learning pace as the average girl. Indeed, as shown in Table 13, some subgroups have progressed more than others in the four learning assessments.

The progress of EGRA scores shows relatively small variations across subgroups, although girls in poor households have only progressed by 6 percentage points, compared to 8 percentage points for the average GEC-T girl. Secondary girls show no progress in EGRA, but this finding is not generalisable to all secondary girls as most of them have not been given the EGRA and EGMA.

The SeGRA progress shows more substantial variations, with girls in poor households and whose primary caregiver or head of household has no education having progressed the least, while underage girls have progressed by 16 percentage points, substantially more than the average girl (11 percentage points).

The progress in EGMA shows even greater variations, with poor households only progressing by one percentage point, as much as girls whose father does not live in the household. Girls who are overage, with disability and whose mother does not live in the household only progressed by 2 percentage points, which is twice less than the average girl who progressed by 4 percentage points.

The SeGMA progress is largely correlated with SeGRA, with the smallest learning progress observed for overage girls and girls whose primary caregiver or the head of household has no education.

Table 13: Learning progress of GEC-T girls between baseline and midline

Learning progress of GEC-T girls, in percentage points		EGRA	SeGRA	EGMA	SeGMA
All ('average girl')		+8	+11	+4	+8
Enrolment	OSG	+8	n/a	+7	n/a
	ISG	+8	+13	+4	+8
School level ⁽¹⁾	Primary	+8	+8	+4	+4
	Secondary	0	+12	-1	+12
Age - grade status	Overage	+8	+8	+2	+5
	Underage	+9	+16	+9	+9
	Correct age	+8	+11	+5	+10
Disability (girl)	One type	+9	+12	+2	+8
	Multiple	n/a	+11	n/a	+12
Disability (PCG)	Disabled	+8	+11	+1	+8
Orphan status	Single orphan	+8	+10	+4	+8
	Double orphan	n/a	n/a	n/a	n/a
Family status	Girl is married	n/a	n/a	n/a	n/a

Learning progress of GEC-T girls, <i>in percentage points</i>		EGRA	SeGRA	EGMA	SeGMA
	Girl is a mother	n/a	n/a	n/a	n/a
	Mother not in HH	+7	+10	+2	+6
	Father not in HH	+8	+10	+1	+8
Household poverty indicators	Less than 50 (less poor)	+11	+9	+8	+7
	More than 50 (poorer)	+6	+7	+1	+6
Household education	PCG has none	+9	+7	+5	+3
	HOH has none	+7	+6	+4	+3

Key: As opposed to the previous table, the colours in this table are unrelated to statistical significance. Simply, darker shades demonstrate a larger progress between baseline and midline, while lighter shades demonstrate a smaller progress.

Note: (1) Only a few primary girls have received SeGRA or SeGMA, and only a few secondary girls have received EGRA or EGMA. The corresponding learning progress is therefore based on small samples, not representative of all primary/secondary girls, and needs to be interpreted with caution.

7. Intermediate outcomes have improved for the average GEC-T girl more than the average comparison girl between baseline and midline. However, these improvements have translated into small gains in learning.

Changes in intermediate outcomes of GEC-T girls

Alongside learning and transition, GEC-T projects also tracked 'intermediate outcomes'. These intermediate outcomes were included to measure key steps more accurately in the programmes' theories of change and were recognised as key enablers to improving learning, transition, and sustainability. The intermediate outcome categories include attendance, economic empowerment, life skills, quality of teaching, gender-based violence, and parental attitudes.

To analyse the change in intermediate outcomes, we use difference-in-difference regressions for the panel sample (recontacted) girls, controlling for age and overage for grade status, and using beneficiary-population weights. These correspond to the difference between the baseline-midline change among the treatment group compared to that among the comparison group. In other words, the difference-in-difference is equal to the change in intermediate outcome for girls from the treatment group *over and above* the change for girls from the comparison group. The difference-in-difference regressions were run separately for each intermediate outcome factor (see Methodology section for more details).

The GEC-T is associated with a statistically positive change in most intermediate outcomes relative to the comparison group. The difference in intermediate outcome improvements between the treatment and comparison group is statistically significant at the 95% confidence level across numerous outcomes as shown in Table 14 below. In particular, GEC-T is associated with improved girls' confidence as approximated by lower nervousness, girls' leadership, and communication skills as well as numerous categories of teaching quality including improvements in gender-sensitive pedagogy, teacher attendance, and reduction in school corporal punishment.

Surprisingly, GEC-T is statistically negatively associated with three factors: girls' school attendance in the most recent school year, girls' perception of doing well on test because of luck, and perception of unsafe travel to school (reported by primary caregivers). Portfolio level change in any intermediate outcome factor was generally driven by three projects at any time.⁸⁶

⁸⁶ Meaning that a change in any one intermediate outcome was typically driven by three different projects at a time. The specific projects driving the change in any intermediate outcome vary by factor considered for the analysis.

Table 14: Change in intermediate outcomes of GEC-T girls

Difference-in-difference for panel sample (recontacted girls), in <i>percentage points</i>		Variable	DID coefficient
Attendance		Did not attend school on most days since the start of the most recent school year	+5** (<0.001)
Economic empowerment		Difficult to afford to go to school	-2
Life skills	Academic self-confidence	Nervous when reading in front of others	-4** (<0.001)
		Nervous when doing maths in front of others	-4** (0.001)
		Not confident answering questions	-1
		Does not ask teacher if does not understand	0
		Does not do well on test because of hard work	0
		Does well on test because of luck	+6** (0.011)
		Cannot describe thoughts when speaking	0
	Self-efficacy	Not focused on a goal	-2
		Cannot put a plan in place and stick with it	-2
	Motivation / goal orientation	Does not want to continue studying	-1
	Decision making	Does not recognise how choices made today affect future life	+1
	Leadership & communication	Cannot work well in a group	-2** (<0.001)
		Cannot organise peers to do an activity	-5** (<0.001)
	School-based loneliness	Feels lonely at school	-8
Teaching quality	Welcoming environment	Not made to feel welcome in the classroom	-1
	Gender-sensitive pedagogy	Boys and girls are treated differently in the class	-4** (0.002)
		Teacher asks more questions to one gender	-2** (0.001)
		Teacher asks harder questions to one gender	-2** (0.008)
	Attendance	Teacher often absent from the class	-5** (<0.001)
	Child-centered practices	Teachers does not use a different language to help understand	-3** (0.002)
		Teacher does not suggest ways to continue study after school	-1
	Interactive learning	Teacher does not encourage students to participate	-1** (0.044)
		Teacher punishes students who get things wrong	-4** (0.001)

Difference-in-difference for panel sample (recontacted girls), in <i>percentage points</i>		Variable	DID coefficient
	School corporal punishment	Teacher used physical punishment on other students in the past week	-6** (<0.001)
		Teacher used physical punishment on girl in the past week	0
School related gender-based violence		Unsafe for girls to travel to schools (PCG)	+6** (0.001)
		Feels unsafe travelling to and from school	-6** (<0.001)
		Feels unsafe at school	0
Parental attitudes		Girls should not stay at school until tertiary education	-10** (<0.001)
		Does not listen to views of girl about her education	-4** (0.027)
		Not worth investing in girl's education when funds are limited	-1
		Girl not as likely to use her education as a boy	0

Note: All regressions control for girls' age and being overaged and use girls' fixed effects and robust standard errors.

Key: Difference-in-difference (DID) coefficients with two asterisks are statistically significant at the 95% confidence level (P -value lower than 0.05 = 5%). Those with one asterisk are statistically significant at the 90% level (P -value lower than 0.1 = 10%). P -values for statistically significant coefficients are included in the brackets.

Green cells show positive and statistically significant DID (GEC-T girls' intermediate outcomes improved significantly more than comparison girls), orange shows negative and statistically significant DID (GEC-T girls' intermediate outcomes worsened significantly more than comparison girls).

Association between learning and intermediate outcomes

Given the positive improvement across intermediate outcomes shown above, the next step is to verify whether these changes translated into learning gains. We conduct this analysis with the panel sample of treated girls – same girls in baseline and midline. The analysis constitutes correlating the change in IO factors with the change in learning, controlling for age and being overage for grade, using beneficiary-population weights, and clustering standard errors by project. The dependent variable is the learning change, while the independent variables include the change in intermediate outcomes between baseline and midline (omitting the negative no change between the rounds – i.e., never doing) and the controls specified above.

The results are shown in Table 15 below which displays the coefficients of the positive change in intermediate outcomes. Note that depending on the variable having a positive or negative meaning, this is either a change from 'no' to 'yes' (e.g. for school attainment in the past week), or from 'yes' to 'no' (e.g. difficulty of school affordability). The remaining coefficients of the regression output are not displayed.

Table 15: Correlation between positive improvement in intermediate outcomes and learning change

Correlation between improvement in IO factor and learning change (in percentage points) for panel sample (recontacted girls)	Overall EGRA	Overall SeGRA	Overall EGMA	Overall SeGMA
Attained school	0	1	-8* (0.059)	5
It is difficult to afford for girl to go to school	0	1	1	-1
Girl gets nervous when she has to read in front of others	2* (0.092)	3* (0.086)	0	1
Girl gets nervous when she has to do maths in front of others	3** (0.008)	2** (0.049)	1	2** (0.033)
Feels confident answering questions in class	3** (0.016)	-2	5** (0.020)	1
Girl asks the teacher if she does not understand something	2	2	4	1

Independent Evaluation of the Girls' Education Challenge Phase II – Aggregate Impact of GEC-T Projects Between Baseline and Midline Study – Final Report

Correlation between improvement in IO factor and learning change (in percentage points) for panel sample (recontacted girls)	Overall EGRA	Overall SeGRA	Overall EGMA	Overall SeGMA
Thinks that succeeded at school because of hard work*	27	-17	32	-10
Feels that girl did well on a test because of luck	1	-1	-1	-4
Can describe thoughts to others when speaking	2	2	2	3
Can stay focused on a goal despite things getting in the way	-4	1	3	7
Can put a plan in place and stick with it	3	-2	2	1
Would like to continue studying/ attending school after this year	10	-4	-7	3
Recognises when choices made today can affect future life	-7	-7** (0.001)	-3	1
Can work well in a group with other people	5* (0.057)	7	5	7
Can organise peers or friends to do an activity	2	4	2	2
Feels often lonely at school	8	9	15** (0.006)	-1
Teachers make girl feel welcome in the classroom	8	3	15** (0.035)	3
Teachers treat boys and girls differently in the classroom	1	1	2	2
Teacher asks more questions to one gender	-9** (0.019)	-1	-4	3
Teacher asks harder questions to one gender	-6** (0.020)	-1	-4	2
Teachers are often absent for class	1	0	1	-3
Teacher uses a different language to help understand	6	-5** (0.013)	2	1
Teacher suggests ways you can continue to study after school	-11	2	-5	-1
Teacher encourages students to participate during lessons	19** (0.002)	17	13* (0.059)	5
Teachers discipline or punish students who get things wrong in a lesson	0	-2	2	-1
Teacher used physical punishment on girl	0	-2	2	1
Saw a teacher using physical punishment on other students	2	1	3	4* (0.077)
Safe for girls to travel to schools [PCG]	0	-2	1	-2
Feels safe travelling to and from school	0	4** (0.006)	1	0
Feels safe at school	5	3	2	0
Would like for girl to achieve tertiary education	4	8** (0.002)	-1	5** (0.041)
Listens to the views of girls when making decisions about her education	5	10** (<0.001)	4** (0.041)	14** (<0.001)
Thinks it is worth investing in girl's education even when funds are limited	3	11	0	2
Thinks that a girl is just as likely to use her education as a boy	13	9	5	10** (<0.001)

Independent Evaluation of the Girls' Education Challenge Phase II – Aggregate Impact of GEC-T Projects Between Baseline and Midline Study – Final Report

Note: All regressions control for girls' age and being overaged and use beneficiary population weights and cluster standard errors at the project level.

Key: Ordinary least squares (OLS) coefficients with two asterisks are statistically significant at the 95% confidence level (P -value lower than 0.05 = 5%). Those with one asterisk are statistically significant at the 90% level (P -value lower than 0.1 = 10%). P -values for statistically significant coefficients are included in the brackets.

Green cells show positive and statistically significant associations (improvement of intermediate outcomes is positively and significantly associated with learning change), orange shows negative and statistically significant associations (improvement of intermediate outcomes is negatively and statistically significantly associated with learning change).

*Positive improvement in IO factor '*Thinks that succeeded at school because of hard work*' is not found to be statistically significantly correlated with learning change despite large coefficients as shown in the table due to extremely low number of observations in the omitted category (disagree in baseline – disagree in midline) which consists of 6 girls for EGRA and EGMA, 48 girls for SeGMA, and 51 for SeGRA.

Factors found positively correlated with learning change include improved girls' self-confidence, classroom positive atmosphere, decrease in self-perceived school-based loneliness, and various factors of parental attitudes.

Importantly, **where there are improvements in intermediate outcomes over and above the comparison group, an improvement in the intermediate outcome is often associated with an increase in learning scores.**

However, several factor improvements are in fact found to be negatively associated with a learning change. Namely, gender bias of teacher asking more / harder questions to one gender, use of different language to help students understand, recognition of impact of choices made today on future girls' outcomes, or school attainment in the current school year. As such, while GEC-T improved intermediate outcome factors, these are necessary, but sometimes insufficient to raise learning levels.

4. Changes in Girls' Transition Outcomes

Box 4: Key findings

1. More than 9 out of 10 re-contacted girls progressed at least one grade level between baseline and midline,⁸⁷ with only five per cent of girls repeating a grade or dropping out of school. Girls in the treatment group were marginally more likely to progress a grade than girls in the comparison group.
2. When accounting for attrition due to girls not being recontacted (41% of the sample), successful transition rate between baseline and midline is reduced to 63% (69% when excluding girls not followed over time by projects' external evaluations).
3. Most girls that did not progress in primary school repeated a grade, while at the secondary school level girls that did not progress to the next grade dropped out.
4. Large variation in transition figures is observed across projects. This is due to projects targeting different age groups (i.e., girls in secondary schools who drop out in the transition to higher secondary grades), girls with disabilities (who were more likely to repeat grades), and the armed conflict in Somalia (which resulted in girls losing a grade due to school closures).

GEC-T external evaluators administered household surveys to capture the many pathways girls can take; then tracked these girls for transition outcomes (see Table 6 for a definition of successful and unsuccessful transition). Besides BRAC and CSU, which considers girls who dropped out since baseline to have failed to transition, most external evaluators excluded girls not re-contacted from their reported transition rates. No data exists for these girls at midline as external evaluators could not track them.

Transition outcomes, thus, are examined in this section for girls who were re-contacted between baseline and midline – 59% of the baseline sample. This may paint a more optimistic picture for transition outcomes than the reality as girls not re-contacted (41 per cent of the sample), of which some are likely to have had unsuccessful transitions, are unaccounted for.⁸⁸ On the other hand, it is worth noting that transition beyond formal cycle of compulsory schools is difficult to categorise. As such, in the next section we present findings derived after consulting project documents and estimating a progression rate which accounts for girl attrition. In addition, we conducted further analysis excluding the sample of girls who were said not to be tracked for transition.

Because the majority of re-contacted girls were sampled in school and are still in school (98% of baseline sample), transition outcomes are limited to schooling outcomes: progression, repetition, and drop-out.

For the out-of-school girls sampled at baseline (2% of baseline sample), the analysis of transition is limited to re-enrolment in school or remaining out of school. Table 16 and Table 17 summarise the transition outcomes of in-school girls by treatment and comparison status and key subgroups⁸⁹, while Table 18 outlines the analogy for out-of-school girls.

In-school girls

More than 9 out of 10 recontacted girls progressed at least one grade level between baseline and midline (Table 16). Girls in the **treatment group were marginally more likely to progress a grade than girls in the comparison group** (1.5 percentage points difference). This difference is the lower bound if the 'unknown transition' were to be added to this measure. This is also because the difference between the treatment and comparison group for the 'unknown transition' is 1.5 percentage points.⁹⁰ Therefore, it is likely that the difference in successful progression between the treatment and comparison group is in fact larger.

The average repetition rate is 3% for both the treatment and comparison group. In comparison, the average dropout rate is 2% and is slightly lower for treatment girls compared to girls in the comparison group.

When we examine school transition data for girls in the treatment group, we find that the percentage of girls who successfully transition through school decreases with age. Young girls (aged 6 to 8 & underage for grade) are

⁸⁷ The average time between baseline and midline data collection is 18 months. The minimum difference between the data collection is 12 months, while the maximum is 24 months.

⁸⁸ Reasons for attrition, which include dropout, are given in **Error! Reference source not found.**

⁸⁹ Treatment girls only.

⁹⁰ Considering 'unknown' transition means unsuccessful transition for a subset of girls. Some projects such as BRAC and CSU define girls lost to attrition (unknown transition rate) as unsuccessful transition.

between 8 to 10 percentage points more likely to progress a grade than older girls (overage for grade & aged 16 to 19, respectively). This is due to young girls' low grade repetition (1%) and dropout rates (0%), and high dropout rates in girls aged 16 to 19 (8%) and overaged (4%).

Unsuccessful progression in primary grades is primarily due to grade repetition (5%), whereas it is primarily due to dropout in the secondary grades (5%). Girls whose household head and primary caregiver did not complete high school have the highest rate of grade repetition.

Treated girls with single and multiple self-reported disabilities are more likely to progress grade relative to the overall treatment sample. This is due to the fact that girls with disabilities have remained in school. Girls who are married and mothers had among the highest transition rates (99% and 98%, respectively) because they only included girls who did not drop out of project schools and could be re-contacted. Girls who were re-contacted were less likely to be married and mothers than girls lost due to attrition (see Table 29 in Annex 3).

Table 16: Transition outcomes of in-school girls, by key subgroups and treatment girls

Transition outcomes of in-school girls <i>in percentage</i>		Successful progression	Partial progression	Any progression (either one or two years)	Grade repetition	Drop-out	Transition between formal ed. and ALP	Unknown transition
All		92	1	93	3	2	0	2
Treatment status	Treatment	92	1	93	3	2	0	2
	Comparison	91	1	92	3	3	0	3
GEC-T girls (Treatment)								
Age	6 to 8	96	0	97	1	0	0	2
	9 to 11	95	1	96	3	0	0	1
	12 to 13	95	1	95	3	0	0	1
	14 to 15	93	1	94	3	1	0	2
	16 to 19	86	1	87	2	8	1	2
School level	Primary	93	1	93	4	1	0	2
	Secondary	91	1	93	1	5	0	1
Age-grade status	Overage	89	1	90	4	5	0	2
	Underage	96	1	97	2	0	0	1
	Correct age	95	1	95	2	1	0	1
Disability (girl)	One type	93	1	95	3	0	0	1
	Multiple	96	1	96	2	0	0	1
Disability (PCG)	One type	89	2	90	3	2	1	2
Orphan status	Single orphan	94	1	95	2	0	0	2

Transition outcomes of in-school girls <i>in percentage</i>		Successful progression	Partial progression	Any progression (either one or two years)	Grade repetition	Drop-out	Transition between formal ed. and ALP	Unknown transition
	Double orphan	94	1	95	2	0	1	2
Family status	Girl is married	99	0	99	1	0	0	0
	Girl is a mother	97	0	97	0	0	1	1
	Mother not in HH	92	2	94	2	1	0	3
	Father not in HH	93	1	95	3	1	0	2
Household poverty	Less than 50% PI	94	1	94	3	1	0	1
	More than 50% PI	94	1	95	2	1	0	2
Household education	PCG has none	94	0	95	4	0	0	1
	HOH has none	93	1	93	4	1	0	1

Note: row columns (any progression, grade repetition, dropout, transition between formal education and ALP, and unknown transition) sum 100; where they sum 99 it is an expected result of adding to the nearest whole number

There is a large variation in successful grade progression among the treatment girls by project (see [Table 51 in Annex 10](#)), ranging from about 60% (PEAS and Viva) to 100% (AKF). This variation is partly explained by the level of education projects target, with those targeting grades 8 onwards showing lower progression rates (i.e., girls in secondary schools) due to high dropout rates. PEAS Uganda, for example, targets girls in secondary school, and we observe that enrolment decreases significantly from S1 to S4. 90% of dropouts occurring between baseline and midline take place in the transition from S3 to S4 (lower secondary) and S4 to S5 (lower to upper secondary). The same is observed for Mercy Corps Nepal, where most girls drop out of school transitioning from Grade 10 to 11 (secondary to higher secondary). In addition, projects targeting disabled girls (CSU and LCD) show lower grade progression (84% and 83%, respectively) due to higher repetition rates (16% and 12%). Lastly, a subset of girls in CARE project schools was affected by armed conflict in Somalia. Girls were held back or asked to repeat a grade due to school closures which in some cases caused a loss of a full academic year.

Accounting for attrition

Transition rate above is restricted to only re-contacted girls, which may present a more optimistic picture of transition outcomes. It is likely that many girls who are not re-contacted did not transition successfully as most girls were tracked through schools. Based on the project documents, we estimate that 20% of lost girls transitioned successfully, 47% transitioned unsuccessfully, and 33% had an unknown transition (more details in Methodology section). Accounting for attrition, we estimate that 63% of girls transitioned successfully, 23% transitioned unsuccessfully, and 14% have unknown transition ([Table 52 in Annex 10](#) shows transition rates by project). This provides a more realistic but still conservative estimate of transition outcomes for in-school GEC-T girls as many of those with unknown transition likely did not transition successfully.

As some projects changed protocol in tracking girls from baseline to midline, some girls were not tracked at all by external evaluators. These girls may have different trajectories than the reasons given by the projects. The third row of [Table 17](#) excludes girls who were not tracked due to changes in programme design. This brings up the estimate of successful transition to 69%.

Table 17: Transition outcomes of in-school girls, comparing between re-contacted girls only and accounting for attrition

Transition outcomes of in-school girls <i>in percentage</i>	Successful progression	Unsuccessful progression	Unknown transition
Re-contacted girls only	92	6	2
Accounting for attrition	63	23	14
Accounting for attrition (excluding girls not tracked by programme design)	69	19	12

Out-of-school girls

Out-of-school girls represent two per cent of the overall sample.⁹¹ Most out-of-school girls at baseline (98%) remained out of school at midline – only eight girls (1%) transitioned to formal education between baseline and midline (Table 18). This, however, may indicate that girls received other support – to transition into TVET, beginning employment in a business or as a wage labourer, or having a positive migration (see Table 6). For example, Mercy Corps Nigeria reported that 16% of out-of-school girls in the treatment group were actively involved in business and 43% received vocational training.

Table 18: Transition outcomes of out-of-school girls

Transition outcomes of out-of-school girls <i>in percentage</i>		Transition to formal education from OOS	Remained OOS	Unknown transition
All		1	98	1
Treatment status	Treatment	0	99	1
	Comparison	2	96	2

⁹¹ Due to small sample size, we do not conduct subgroup analysis for out-of-school girls.

5. Conclusions

Objective 1

The study created a portfolio-level dataset of quantitative baseline and midline data collected by projects' external evaluations. Data was combined from four instruments: (1) the EGRA, EGMA, SeGRA and/or SeGMA Learning Assessments (LA); (2) the Girl Surveys (GS); (3) the Household Surveys (HHS); and (4) the Primary Caregiver Surveys (PCG). The dataset contains nearly all variables included in the Fund Manager questionnaire templates (cleaned and homogenised).

This dataset will be used for future evaluation studies involving GEC-T projects. This dataset, for example, will be used to draw summary statistics on girls with disabilities based on the type of disability reported in Study 4 on Disability. In addition, the dataset and an accompanying codebook have been produced to allow other researchers, within FCDO or the FM, to analyse it.

Objective 2

RQ1. How did girls' learning levels change between baseline and midline?

- The literacy and numeracy levels of GEC-T girls have significantly improved between baseline and midline. They did so across all subtasks of the EGRA, EGMA, SeGRA and SeGMA, and for all age categories. In other words, **girls' learning improved across the GEC-T portfolio between baseline and midline, across a range of literacy and numeracy skills.**
- **These increases, while statistically significant, are relatively small⁹².** For oral reading fluency scores, which are expressed in words-per-minute rather than percentage correct, this means that girls correctly read 10 words-per-minute more than they did at baseline after 18 months of intervention⁹³.

Are the observed changes different for literacy and numeracy?

- **GEC-T girls improved slightly more in literacy than in numeracy:** their average *percentage correct score* increased by eight percentage points in EGRA and by ten percentage points in SeGRA, as opposed to four percentage points in EGMA and eight percentage points in SeGMA.

Did GEC-T girls improve more than non-GEC girls?

- When comparing GEC-T beneficiaries to comparison groups, it appears that **the average GEC-T girl has learned more than the average comparison girl between baseline and midline.** This is also true across all four learning assessments. But the *difference-in-difference* effect is small: learning scores increased between two percentage points (EGRA and EGMA) and three percentage points (SeGRA and SeGMA) more than in the comparison group. On average, the progress of GEC-T girls is larger by only a couple of percentage points (or a couple of words-per-minute) more than that of comparison girls. This suggests that **the GEC-T is associated with a positive, but small, increase of girls' learning.**
- When weighting each project equally in the portfolio, the GEC-T shows no statistically significant effect on learning. This is because **the average learning progress of GEC-T girls is driven by a small number of large projects:** three projects account for three quarters of the GEC-T target learning beneficiary population⁹⁴. Many smaller projects are not associated with any significant learning over and above their comparison groups.

How do girls' midline learning levels compare with international standards and benchmarks?

- **Most GEC-T girls still fall short of international standards of oral reading and comprehension:** at age 10, GEC-T beneficiary girls only read 45 words-per-minute on average across projects. This is at the lower end of a contextually relevant reading fluency benchmark of between 45 to 60 words-per-minute⁹⁵. Children in

⁹² The average time between the baseline and midline data collection is 18 months, which is the average time girls have had to improve their learning. A ten percentage point increase means that girls who scored 50% correct at baseline scored 60% correct at midline, or that girls who scored 10% correct at baseline scored 20% correct at midline.

⁹³ As a matter of comparison, and although they have been developed for the USA curriculum hence for a different context, the DIBELS benchmark expects an increase of about +45 words-per-minute during Grade 2, and of about +30 words-per-minute during Grade 3.

⁹⁴ These are Camfed International in Tanzania, Zambia and Zimbabwe, EDT Kenya and Discovery Learning Alliance (DLA) in Ghana, Kenya and Nigeria.

⁹⁵ See Developing Cross-Language Metrics for Reading Fluency Measurement (Abadzi, 2012).

the USA would be expected to read at over 100 words per minute by the end of grade three⁹⁶. At age 11, girls can read as many as 63 words-per-minute on average.

- **Only 8% of GEC-T girls aged 10 at midline can understand what they read.** Using the UNESCO-led Global Education Standard that “all children should read and understand what they read by age 10, or at the latest by the end of primary school”⁹⁷, this implies that the average GEC-T girl is still trapped in *learning poverty* at midline. The share of girls who can understand what they read reaches about 30% towards the end of primary school. This share increases steadily as girls grow older but remains capped at 40% of girls aged 18.

RQ2. Who has learned the most (and least), and how much did they learn?

- The increase in girls' learning scores has been accompanied by a substantial decrease of the share of girls scoring zero, which implies that **even the most educationally marginalised girls learned between baseline and midline**. This improvement can be observed across all literacy and numeracy subtasks⁹⁸.
- **Progress in learning over and above the comparison group is consistent across key subgroups** as defined by girls' enrolment status, school level, disability status, orphan status, family status (mother or father not living in the household⁹⁹), household poverty and education levels. In other words, the learning of girls belonging to these subgroups has improved slightly more than the learning of girls of the same subgroups who live in comparison areas.
- **However, focusing on the progress of GEC-T girls between baseline and midline, some subgroups appear to have learned at a significantly slower pace than others.** This is the case for girls who are overage for their grade, girls living in poor households and girls whose primary caregiver or head of household have no education (in literacy and numeracy) and girls whose father or mother does not live in the household (in numeracy only).

RQ3. Why did some girls learn more than others?

- Intermediate outcomes (i.e. life skills, quality of teaching, and parental attitudes) have improved for the average GEC-T girl more than the average comparison girl between baseline and midline. Although recognised as key enablers to improving learning and transition in projects theories of change, **improvements in intermediate outcomes have translated into small gains in learning**, as seen in RQ1.

RQ4. Did girls successfully transition between baseline and midline?

- **Transition on the GEC-T tends to suggest an overly optimistic picture as girls who are not re-contacted in the midline (from baseline) – 42% of the sample of girls at baseline – have an unknown and unaccounted transition status.** These girls, however, are more marginalised than those recontacted: they were older, overage for their grade, attending higher grades, and more likely to be married and be mothers.
- **Overall, most girls progress through school (more than 9 out of 10) and grade progression for re-contacted girls in schools in the treatment group is higher than in the comparison group. When accounting for attrition (and applying our assumptions of percentage of girls not re-contacted who had a successful and unsuccessful transitions based on projects' reports), we estimate that 63% of girls have a successful transition outcome.**
- There are **large variations in grade progression by projects** explained by the group of girls they target (i.e., those in secondary, girls with disabilities) and small differences in grade-progression, repetition and dropout across key subgroups.
- GEC-T out-of-school girls only represent 2% of the sample. Few of these girls transitioned into schooling, with many receiving other support to transition into vocational training or employment.

⁹⁶ See DIBELS benchmarks: http://oregonreadingfirst.uoregon.edu/downloads/assessment/dibels_benchmarks_3x.pdf

Note these benchmarks have been developed for the English language. Seven out of 27 GEC-T projects have assessed girls' learning in different languages of assessment than English.

⁹⁷ See the official page of the Global Coalition for Education Data by the World Bank and UNESCO: <https://gaml.uis.unesco.org/learning-poverty/>

⁹⁸ Except for oral reading fluency: only 2% of GEC-T girls could not read a single word at baseline.

⁹⁹ The number of married girls and mother girls was too small to include here.

Limitations of the GEC-T evaluation design and effects on study findings

- There are many **inconsistencies in the way the GEC-T projects' quantitative evaluations have been designed** (number and type of cohorts, tracking and recontacting strategies, presence/absence and characteristics of comparison groups, sample sizes) and in the way outcomes have been measured (different learning assessments or subtasks used between baseline and midline, across GEC-T projects, and across grade levels or enrolment status within the same project, different definitions of transition). These inconsistencies limit the ability to compare and aggregate findings across projects, at both the portfolio and subgroup levels. For this reason, the analysis has often had to be conducted on reduced sample sizes or includes explicit caveats that restrict its external validity and generalisability.

6. Recommendations

The GEC-Transitions (GEC-T) Window is approaching the end of its lifetime with only four GEC-T projects remaining by October 2022. These recommendations are aimed at the FCDO and wider policy-makers to inform future girls' education programming, policy and measurement approaches. As a quantitative assessment of the GEC-T's impacts and associations with different variables (rather than causal relationships), these recommendations focus on high-level policy and programming objectives and measurement issues.

Policy

There is a global learning crisis and a need for continued support for girls' learning. GEC support has not sufficiently improved girls' literacy at an early enough age to enable them to learn as they progress through school, a stronger focus is needed on girls' literacy in early primary school grades.

While girls' reading fluency improves as they get older, their literacy is very low in the foundational years of primary school, which means that their progress is too slow as they do not have the reading skills needed to learn as they transition through school. At age 10, most GEC-T girls can hardly read (averaging just 45 words-per-minute) and less than 8% understand what they read. As reading fluently provides a foundation for learning other skills, a stronger focus is needed on understanding and addressing the factors preventing girls from learning the basics, to prevent them from being trapped in learning poverty. This could include greater attention to school readiness and girls' literacy in early primary school grades.

More targeted interventions are required to support those subgroups of girl learners who have fallen behind other GEC girls. Regularly assessing girls' learning and their learning needs is essential to identifying who is falling behind and what targeted support is needed to help those girls catch up.

Some groups of girls – those who are overage for their grade; are living in poor households; whose primary caregiver or head of household has no education; whose father or mother does not live in the household – learned at a significantly slower pace than other girls between baseline and midline, in both the treatment and comparison areas. These subgroups require more targeted interventions to enable them to catch up and keep up with their cohort.

Interventions to improve transition through education need well-defined pathways – especially for girls that drop out of school and /or migrate outside of project areas. This requires tracking strategies that allow projects to follow girls' progress across multiple pathways.

GEC-T was successful in supporting the girls that projects *could track* in their transition through education, but projects experienced high rates of attrition from their samples. When girls were sampled in their homes as opposed to schools, attrition rates were lower, but they were still high. Many projects were not designed to track girls outside defined project areas, meaning that when girls moved, they were lost.

This means the analysis on transition is biased as the “lost” girls are more likely to not have transitioned successfully. This is not just a statistical issue, but an education issue, as projects are losing beneficiaries early with the outcomes for these girls largely unknown.

It is crucial to invest in better tracking to identify what happens to girls who leave project areas (or move to pathways not supported by projects) but also to keep supporting girls in their transition to adulthood. Simple (though costly and time-consuming) activities can reduce attrition, such as keeping in touch with girls between evaluation points, collecting additional community contacts for the girl outside the household, or visiting girls at schools they moved to.

Measurement

A more consistent measurement strategy is needed across the portfolio to achieve the right balance between project-level adaptability and portfolio-level consistency to enable aggregate impact assessment and to compare performance across the portfolio.

The GEC-T does not easily lend itself to aggregating project-level data to assess the overall impact of the portfolio and to compare performance across different girls, interventions and contexts – projects used different targeting and sampling strategies; undertook bespoke adaptations of learning assessments; and used different ways of tracking girls. The measurement requirements for GEC Phase II were less standardised than in Phase I to improve the

flexibility and relevance of projects' evaluation designs. However, this limits a portfolio-level evaluation. The FCDO and other policy-makers need to consider the extent to which portfolio level aggregate assessments of impact and performance across a portfolio are important and the trade-offs involved when designing future girls' education programmes.

Intermediate outcomes need to be measured consistently using suitable scales, and they need to be linked to learning outcomes in the data. The same girls who are being assessed for changes in their intermediate outcomes should also be assessed for changes in their learning outcomes to establish these links.

The observed improvements in (some) intermediate outcomes have translated into small gains in learning. However, analysis has been limited by the inability to systematically link learning gains (captured through school assessments) to changes in intermediate outcomes (captured by household and girl surveys). As a minimum requirement, projects should measure the explicit link between changes in intermediate outcomes and their association with changes in the same girls' learning outcomes. Assessing such links requires tailoring evaluation designs at the portfolio and project-level to ensure all outcomes are captured systematically, so that they can be consistently linked together.

Further research

Further research would be beneficial to illuminate the factors driving the positive results achieved by the three large projects in the GEC-T that delivered most of the learning gains.

Larger projects (in terms of the number of beneficiaries) were more successful in increasing girls' learning outcomes compared to the comparison group. More investigation is required into the reasons why different types of interventions improve learning for different types of girls and under what conditions, including what effect the quality of a project's design has compared to the quality of its implementation and the role scale has on its success.

7. Bibliography for Learning Standards and Benchmarks

This bibliography presents references for the learning standards and benchmarks reviewed as part of this study.

1. Abadzi (2012). Developing Cross-Language Metrics for Reading Fluency Measurement: Some Issues and Options. <https://documents1.worldbank.org/curated/ar/611321468338475370/pdf/797740WP0wpm0e0Box0379789B00PUBLIC0.pdf>
2. Abadzi (2013). Literacy for All in 100 Days? A research-based strategy for fast progress in low-income countries. <https://openknowledge.worldbank.org/bitstream/handle/10986/16248/797850WP0ENGLIOBox0379789B00PUBLIC0.pdf?sequence=1&isAllowed=y>
3. Cally A., Gabrielle W., et al. (2021). Benchmarking oral reading fluency in the early grades in Nguni languages. *International Journal of Educational Development*. Volume 84. <https://reader.elsevier.com/reader/sd/pii/S0738059321000869?token=2F1268170F961497A1205507177AC0C497687C8D536303AE43FE1C077FF2FB454B6F0BE193641B1E7DB9022865CBB891&originRegion=eu-west-1&originCreation=20220701154558>
4. DIBELS (-). DIBELS Benchmark Goals and Indicators of Risk. http://oregonreadingfirst.uoregon.edu/downloads/assessment/dibels_benchmarks_3x.pdf
5. Linda P., Leanne K., Yasmin S. (2016). Using an Assessment of Early Mathematical Knowledge and Skills to Inform Policy and Practice: Examples from the Early Grade Mathematics Assessment. <https://ijemst.org/index.php/ijemst/article/view/86>
6. RTI. (2017). Setting and Using Benchmarks for Reading Performance. https://ierc-publicfiles.s3.amazonaws.com/public/resources/10746_Benchmarks_brief_02-16-17_r6.pdf
7. Science of Teaching for Foundational Literacy and Numeracy. (-). Assessment-Informed Instruction: Systems Level. https://assets-global.website-files.com/61366d43ebd6df56d9b67a11/61748baba5e6696e8389b952_cBuATsusFbTaPCp9_Dcu8bhE1tIMN7Sci-Assessment%20at%20systems%20level.pdf
8. The Centre for Global Education Monitoring. (2020). The Early Grade Reading Assessment: Assessing children's acquisition of basic literacy skills in developing countries. Assessment GEMS series No.2. <https://research.acer.edu.au/cgi/viewcontent.cgi?article=1017&context=assessgems>
9. UNSTATS (2022). SDG indicator metadata. <https://unstats.un.org/sdgs/metadata/files/Metadata-04-01-01.pdf>
10. UNESCO. 'Learning Poverty'. Retrieved from <https://gaml.uis.unesco.org/learning-poverty/> (accessed April 2022)
11. UNESCO Institute for Statistics. (2018). Minimum proficiency levels (MPLs): outcomes of the consensus building meeting. https://gaml.uis.unesco.org/wp-content/uploads/sites/2/2018/10/GAML5_4.1.1_04-MLPs_Outcomes-of-the-Concensus-building-meeting.pdf
12. USAID. (2009). Early Grade Mathematics Assessment (EGMA): A Conceptual Framework Based on Mathematics Skills Development in Children. https://pdf.usaid.gov/pdf_docs/PNADS439.pdf
13. USAID (2016). Early Grade Reading Assessment (EGRA) Toolkit. https://pdf.usaid.gov/pdf_docs/PA00M4TN.pdf
14. USAID & RTI International (2017). All Children Reading – Asia (ACR – Asia) EGRA Benchmarks and Standards Research Report. https://ierc-publicfiles.s3.amazonaws.com/public/resources/Asia%20ACR%20Benchmarking%20Report_FINAL%20REVIS ED_21%20Dec%2017.pdf
15. World Bank, 'Learning Poverty'. Retrieved from <https://www.worldbank.org/en/topic/education/brief/learning-poverty> (accessed April 2022)