TETRA TECH
International Development

Independent Evaluation of the Girls' Education Challenge Phase II – Evaluation Study 4: Educating Girls with Disabilities in GEC II

Annexes

UKaid
from the British people

Client Name: Foreign Commonwealth & Development Office (FCDO)

Project Name: Independent Evaluation of the GEC Phase II

Contract Number: PO 10019

Partners:

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

Authors: Singal, Nidhi[1]; Niaz, Laraib[1]; Gupta, Romanshi[2]; Bligh, Libby[2]; Ogando-Portela, María José[3]; Aslam, Monazza[4]; Rawal, Shenila[4].

May 2023

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

Programme Director

Signature:

[1] REAL Centre at Cambridge
[2] Tetra Tech International Development
[3] Fab Inc
[4] Oxford Partnership for Education Research & Analysis

Table of Contents

Acknowledgements iii
Acronyms iv
Annex A: Terms of Reference 1
Annex B: Research Design and Methodology 15
Annex C: Quantitative Report 47
Annex D: Details of Case Study Projects 117

Acknowledgements

The study team is grateful to the various partners and individuals who have contributed to this report.

Tetra Tech International Development Europe: Simon Griffiths, Programme Director; Pauline Rose, Technical Director; Monazza Aslam, Advisor; Shenila Rawal, Advisor; Paul Atherton, Advisor; Louise Cathro, Programme Manager, and Amy Macmillan, Assistant Programme Manager. Primary data analysis was supported by: Zinnie Cowing, Hannah Allroggen, Josh Shelley, and Anna Reid (Associate). Quantitative data analysis was supported by: Youngjin Kim, Tomáš Koutecký (Fab Inc.), and Sirin Tangpornpaiboon (Fab Inc.).

Academic Partners and Experts: Institute for International Development Studies; Dr. Niraj Poudyal (Nepal); Centre for Social Research (Malawi); Jenipher Mbukwa (Malawi/ Uganda).

Fieldwork Team: Julia Midland, Fieldwork Research Manager, for training data collection partners, overseeing the fieldwork, and piloting and adapting the fieldwork as needed. Our data collection partners – Rooster Logic, Research Options and Research Plus – who undertook the fieldwork. Brijendra Joshi from Rooster Logic, Ernest Poya from Research Options and Dixie Avugwi from Research Plus for overseeing the fieldwork. Sidra Mughal-Butt for providing Quality Assurance support.

Evaluation Studies Working Group (ESWG), comprising members of the Independent Evaluation (IE) team, Fund Manager (FM) and the Foreign, Commonwealth & Development Office (FCDO), and an Independent Advisory Group (IAG) for providing their written and verbal feedback throughout the various stages of the report.

Implementing Partners: All 27 IPs who participated in the key informant interviews, and especially the three case study project IPs for all their support in facilitating our fieldwork.

Consultants: Saba Saeed (independent consultant), Stephanie Nowack (PhD Candidate, University of Cambridge) for supporting the analysis of the qualitative documentation. Thilanka Wijesinghe (PhD Candidate, University of Cambridge) for helping with the Political and Economic Analysis (PEA).

Acronyms

AGES	Adolescent Girls Education in Somalia
CAMFED	Campaign for Female Education
CBE	Community-Based Education
CFM	Child Functioning Module
CRPD	Convention on the Rights of Persons with Disabilities
CRW	Community Resource Workers
CSU	Cheshire Services Uganda
CtC	Child to Child
CWDs	Children with Disabilities
DTL	Distance Teaching Learning
EGDUE	Empowering Girls with Disabilities in Uganda through Education
EGMA	Early Grade Mathematics Assessment
EGRA	Early Grade Reading Assessment
EIES-GWDs	Expanding Inclusive Education Strategies for Girls with Disabilities
EMIS	Education Management Information System
ENGAGE	Empowering a New Generation of Adolescent Girls with Education
ESWG	Evaluation Studies Working Group
FCAS	Fragile and Conflict-Affected Settings
FCDO	Foreign, Commonwealth and Development Office
FGD	Focus Group Discussion
FM	Fund Manager
GATE	Girls' Access to Education
GEC	Girls' Education Challenge
GEC II	Girls' Education Challenge Phase II
GEC-T	GEC-Transition
GESI	Gender and Social Inclusion
GIEN	Girls Inclusive Education Network
GWDs	Girls with Disabilities

IAG	Independent Advisory Group
ID	Identifier
IDI	In-Depth Interview
IE	Independent Evaluation
IERCs	Inclusive Education Resource Centres
IP	Implementing Partner
JCE	Junior Certificate of Education
KII	Key Informant Interview
KISE	Kenya Institute for Special Education
LCD	Leonard Cheshire Disability
LMIC	Low- and Middle-Income Countries
LNGB	Leave No Girl Behind
MSCE	Malawi School Certificate of Education
NGO	Non-Governmental Organisation
NSGE	National Strategy for Girls' Education
OOS	Out-of-School
OPDs	Organisations of Persons with Disabilities
PCG	Parent/ Caregiver
PEA	Political Economy Analysis
PRT	Physical Rehabilitation Therapy
PSLCE	Primary School Leaving Certificate for Education
PVM	Participatory Visual Methodologies
RQ	Research Questions
SEGMA	Secondary Grade Mathematics Assessment
SEGRA	Secondary Grade Reading Assessment
SIP	School Improvement Plan
SOMGEP	Somali Girls Education Promotion Programme
SSHRCC	Social Sciences and Humanities Research Council
STAGES	Steps Towards Afghan Girls

TENI	Tackling Education Needs Inclusively
ToC	Theory of Change
ToR	Terms of Reference
UDL	Universal Design for Learning
UIS	UNESCO Institute of Statistics
UNCPRD	Convention on the Rights of Persons with Disabilities
VSO	Voluntary Service Overseas
WG	Washington Group
WG-SS	Washington Group Short Set
WWW	Wasichana Wetu Wafaulu

Project Acronyms

GEC-T 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
Cheshire Services Uganda	CSU	Uganda
ChildHope UK	ChildHope	Ethiopia
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
Plan International	Plan	Sierra Leone
Promoting Equality in African Schools	PEAS	Uganda

GEC-T Project name	Acronyms used in report	Project location(s)
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 University Service of Canada	WUSC	Kenya
World Vision	WV	Zimbabwe

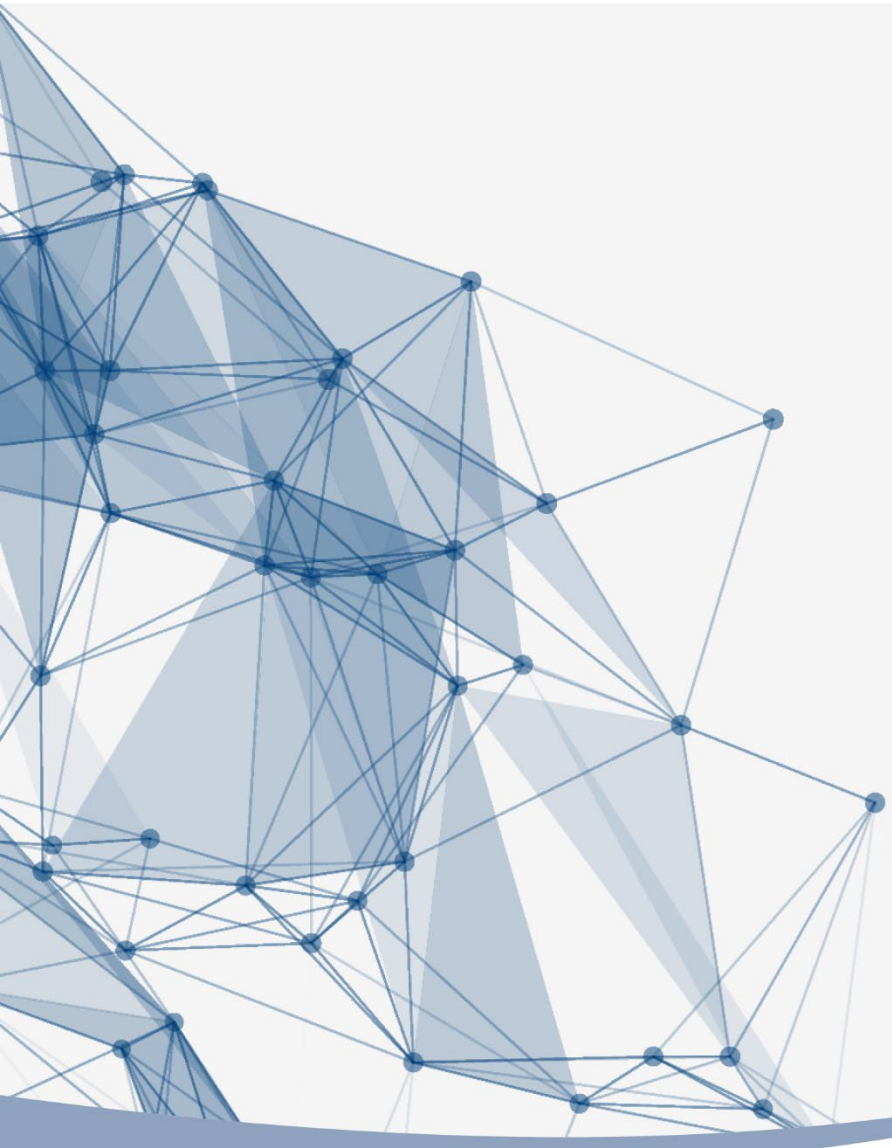
LNGB Project name	Acronyms used in report	Project location(s)
ACTED	ACTED	Pakistan
ActionAid International Kenya	ActionAid	Kenya
AKF	AKF	Afghanistan
CARE International UK	CARE	Somalia
International Rescue Committee [Pakistan]	IRC Pakistan	Pakistan
International Rescue Committee UK [Sierra Leone]	IRC Sierra Leone	Sierra Leone
Link Community Development International	Link	Malawi
People in Need (PIN), Ethiopia	Pin Ethiopia	Ethiopia
People in Need (PIN), Nepal	Pin Nepal	Nepal
Plan International UK	Plan	Zimbabwe
Population Council	Population Council	Ethiopia
Street Child	SC Nepal	Nepal
VSO	VSO	Nepal
World Education, Inc.	WEI	Ghana

Annex A: Terms of Reference

TETRA TECH
International Development

Independent Evaluation of the Girls' Education Challenge Phase II – Draft Terms of Reference for Evaluation Study 4 - Disability

March 2022



UKaid
from the British people

Independent Evaluation of the Girls' Education Challenge Phase II – Draft ToR for Evaluation Study 4

Client Name: FCDO

Project Name: Independent Evaluation of the GEC Phase II

Contract number: PO 10019

Version 3

31 March 2022

Tetra Tech International Development 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, Programme Director

Signature:

Table of Contents

Abbreviations and Acronyms iii
1. Background and Purpose 1
2. Scope of Work 2
3. Methodology 2
3.1. Approach and Key Phases 2
4. Research Ethics 6
5. Key Limitations, Risk Assessment and Mitigation Plan 7
6. Work Plan 8
6.1. Expected Deliverables 8
7. Team Composition 8
8. Stakeholder Engagement 9

List of Tables

Table 1: Proposed Methods and Criteria per Research Question 3
Table 2: Key Limitations, Risk Assessment and Mitigation Plan 7

List of Figures

Figure 1: Work Plan 8

Abbreviations and Acronyms

ESWG	Evaluation Studies Working group
FCAS	Fragile and Conflict-Affected Situations
FCDO	Foreign, Commonwealth and Development Office
GEC II	Girls' Education Challenge Phase II
GEC-T	Girls' Education Challenge – Transition
GWD	Girls with Disability
IAG	Independent Advisory Group
IP	Implementing Partners
LNGB	Leave No Girl Behind
OPD	Organisations of Persons with Disabilities
PEA	Political Economy Analysis
RQ	Research Question
ToR	Terms of Reference
WG	Washington Group

Terms of Reference for the Evaluation Study 4: Disability

1. Background and Purpose

1. The Girls' Education Challenge Phase II (GEC II) aims to reach the most marginalised girls and is committed to ensuring *all* girls receive a quality education. Disability has been identified by GEC II stakeholders as an important marker of marginalisation, with all GEC II projects expected to be inclusive of girls with disabilities (GWD).
2. This study forms part of the Independent Evaluation of the GEC II commissioned by the Foreign, Commonwealth and Development Office (FCDO) in 2020.
3. The **primary objective of this study (Study 4)** is to understand in what ways GEC II projects have engaged girls with disabilities (GWDs) through their interventions, and the perceived and observed effects of these approaches on their outcomes (including learning and socio-emotional wellbeing) and on the engagement of GWDs with their peers, teachers/ educators, and caregivers/families. The study timelines are set out in the "*Work Plan*" in section 6.
4. Early discussions with key GEC II stakeholders highlighted an important evidence gap in understanding 'what works' with respect to education programming among GWDs. Keeping this in mind, the following **research questions (RQs) have been preliminarily developed**:
    a) RQ1: To what extent have GEC II projects supported GWDs (with a particular focus on their learning engagement and outcomes) through their interventions, including approaches adopted during Covid-19, and what are the factors influencing these decisions?
    b) RQ2: What are the perceived and observed effects of the interventions adopted by selected GEC II projects on GWDs, their teachers/ educators, families/ caregivers, and communities?
    c) RQ3: What policy and programmatic implications can the FCDO and GEC Implementing Partners draw about inclusive approaches to education and what works well and less well for GWDs?
5. **Cross-Cutting Themes:**
    a) **Gender and social inclusion analysis:** To the extent possible, this analysis will explore differences for girls with different types of disabilities as well as intersections of disability, gender and other forms of exclusion (e.g., girls with children, out-of-school girls, etc.) with respect to the above research questions. Based on the availability of project data and the intended focus of this study, it is not considered feasible or appropriate to gather or analyse data from boys and /or compare these with GWDs.
    b) **Political Economy Analysis (PEA):** This analysis will explore the following questions – how and to what extent does the wider education system, socio-political environment and economy influence Research Question 2? How do these political, social, and economic factors interact with projects' strategies to support girls with disabilities? As the PEA methodology involves a review of country-specific frameworks, policies, and literature as well as interviews with a range of stakeholders (including government representatives, implementing partners community leaders, parents, girls), it is not possible to conduct PEA across the entire GEC II portfolio. As such, this will be conducted for Research Question 2 only.
6. The above research questions will be refined and detailed during the study design process, further described in the "*Methodology*" in section 3. The study design will additionally be informed by a rapid review of existing evidence as well as a review of the available secondary project documentation/ data. The IE will engage with representatives from the FM who have focused on disability in the GEC II (including about the use of the Washington Group questions). Additionally, initial consultations with project Implementing Partners (IPs) will begin at this stage to discuss IPs' capacities/ willingness to engage with the study.
7. This study is proposed to have an *evaluation for learning* rather than an *evaluation for accountability* objective. The evaluation for learning approach was taken for the first two evaluation studies to avoid

unfairly assessing projects during the Covid-19 period – the effects of which are still largely playing out at the time of designing Study 4. An evaluation for learning approach also lends itself to greater flexibility and opportunity to maximise learning potential from projects working on disability, even if these results were not originally envisaged and do not link directly to expected results.

8. This study offers an opportunity to gain a deeper understanding of which approaches work to engage GWDs by concentrating on the voices of GWDs and those around them on the reported effects of the interventions on their outcomes (such as learning) and their lives. This study can fill an important gap in external evidence by exploring GWDs' perspectives about what works to support their inclusion and is expected to inform future programming accordingly. Primary qualitative data will be gathered and triangulated, to the extent possible, with secondary data (including baseline/ midline/ endline data, project documentation, etc.) and wider literature.
9. The **primary stakeholder audiences** for this study are the FCDO (GEC II Programme Team, FCDO Education Advisors, Regional Education Advisors, Girls' Education Department), the Fund Manager (FM), and project Implementing Partners (IPs). The **secondary stakeholder audiences** for this study are other international donors, agencies, Organisations of Persons with Disabilities (OPDs) and stakeholders working in and investing in girls' education and disability inclusion more generally.

2. Scope of Work

10. As discussed above, this study is proposed as an evaluation for learning. An evaluation for accountability would require evidence-based judgements about the performance of the GEC II projects. A more flexible evaluation focus on learning is proposed at this point in time due to the disruption to project activities caused by Covid-19 The study will instead be focusing on the inclusion in and perceived and observed effects of the interventions on GWDs and their outcomes (such as learning, socio-emotional wellbeing, etc.).
11. This study is intended to be complementary to the ongoing and planned work conducted by the FM. Based on initial discussions with the FM, certain lines of inquiry have been dismissed to avoid duplication in efforts, the details of which will be included in the research design note (see "*Expected Deliverables*" in section 6.1 for more details).
12. Following discussions with key GEC II stakeholders, it is not considered essential to exclude closed projects from the study, as some projects may still be willing and able to engage in the study and take up the findings. Thus, during the research design phase, the IE team will contact all IPs involved with the GEC II to discuss the levels of engagement and ascertain IPs' capacities and willingness to engage/ participate in the study.
13. Available project documentation (e.g., baseline/ midline/ endline datasets, monitoring reports and external evaluation reports) will be used where possible to triangulate the primary data. The data sources will vary based on the proposed research question (see "*Methodology*" in section 3 below). Research Question 1 (RQ1) is intended to be answered through portfolio-level project documentation and learning data from GEC II projects as well as interviews with IPs. Research Question 2 (RQ2) will focus on collecting primary data from a select number of projects identified as case studies. The number of projects included for RQ2 is envisaged to be 2-3, based on the final selection criteria. This is in consideration of the requirements related with adapting research tools to different contexts (language, customs etc), types of disabilities, time and budgetary constraints, and associated fieldwork management processes. Research Question 3 (RQ3) is intended to synthesise and draw upon the findings from Research Questions 1 and 2. This will further be informed by a 'sensemaking' workshop with IPs to discuss and identify the broader learnings across contexts.

3. Methodology

3.1. Approach and Key Phases

14. As described above, this study proposes three research questions, with Research Question 3 (RQ3) intended to synthesise the findings from Research Question 1 (RQ1) and 2 (RQ2). This means that no additional data will be collected for RQ3 outside of a workshop with IPs to identify lessons.

15. The main methodology proposed for **RQ1** is key informant interviews with IPs. This will involve understanding which factors (such as those related to design, context and delivery) influenced IPs' targeting (or not) of GWDs, as well as any changes to their interventions during their project cycles. A preliminary mapping of project interventions indicates that 6 IPs did not focus on disability at inception (including those that may have integrated disability later in their project cycle), while 35 IPs did. Thus, we propose interviewing all IPs that did not focus on disability at inception, and due to time/ resource constraints, purposively select an equal number of IPs that did focus on disability at inception. The purposive selection from the 35 IPs that targeted GWDs will be based on criteria that will be developed further based on the IE's portfolio review, to enable diverse operational and contextual representation. To the extent possible, we will analyse learning data from both Girls' Education Challenge – Transition (GEC-T) projects (as has been conducted for Study 3) and Leave no Girl Behind (LNGB) projects for RQ1.
16. The main methodology proposed for **RQ2** is a case study of projects working with GWDs situated within their wider social milieu (e.g., parents, peers, and communities) using primary qualitative research methods. The nature of the case design will be finalised once projects have been selected. A key criterion for inclusion in RQ2 is that projects must have interventions with an explicit (though not necessarily an exclusive) focus on GWDs. This is intended to enable the study to draw on rich cases that offer the greatest potential for learning, rather than looking at projects which have not had an explicit focus on GWDs.
17. The purpose of primary data collection is to elevate the voices of the participants (particularly girls with disabilities) to understand why projects did or did not target GWDs, in what ways the interventions have engaged GWDs and the perceived and observed effects on their lives, including their inclusion in education.
18. Where feasible, the primary data will be triangulated with the available secondary data. The preliminary inclusion criteria for projects included to answer each RQ, as well as the proposed methods/ data sources has been outlined in *Table 1* below.

Table 1: Proposed Methods and Criteria per Research Question

Research Question	Projects Included	Proposed Methods/ Data Sources
RQ1: To what extent have GEC II projects supported GWDs (with a particular focus on their learning engagement and outcomes) through their interventions, including approaches adopted during Covid-19, and what are the factors influencing these decisions?	1. All projects across the GEC II portfolio 2. Availability of project documentation/ data	Key Informant Interviews with the 6 projects which do not have any focus on disability and an equal number of projects which do have a focus on disability, selected on a criterion which will be developed after an initial mapping of documents has been undertaken. Learning Data from GEC II projects (as far as possible) Project Documentation
RQ2: What are the perceived and observed effects of the interventions adopted by selected GEC II projects on GWDs, their teachers/ educators, families/ caregivers, and communities?	1. All projects across the GEC II portfolio willing and with capacity to engage with in-depth fieldwork	Key Informant Interviews In-Depth Interviews/ FGDs

Research Question	Projects Included	Proposed Methods/ Data Sources
	2. Strong focus on disability from inception 4. Other shortlisting criteria to be determined	Classroom Observations/ Other participatory qualitative methods In all cases accessibility and adaptability of methods to suit the diverse needs of GWDs will be central
RQ3: What policy and programmatic implications can the FCDO and GEC Implementing Partners draw about inclusive approaches to education and what works well and less well for GWDs?	1. All projects across the GEC II portfolio willing and with capacity to engage with IE team	Project Documentation Key Informant Interviews In-Depth Interviews/ FGDs Classroom Observations/ Other participatory qualitative methods Review of existing evidence Workshop with IPs

The **study approach and methodology** will be staged across five phases:

Phase 1: Scoping and Research Design. This phase of the study will be divided into two stages.

a) **Stage 1** involves reviewing existing data on the GEC II portfolio and initial consultations with GEC II stakeholders to frame the overall study and to inform the design of the primary qualitative tools. This will further be divided into four parts**:**

- A rapid review of external evidence on educational interventions to support GWDs to frame and contextualise the study.
- A review of the LNGB projects' evaluation data, including those on learning outcomes, to explore the possibility of analysing these data (GEC-T projects' learning data will be drawn from the IE's Study 3). Additionally, a rapid review of secondary project documentation (including Covid-19 Medium-Term Response Plans) will inform the design of the qualitative tools, with a focus on framing questions about project design, context and delivery.
- Initial consultations with representatives from the FM to understand disability in the GEC II more closely, particularly how learning has been understood and captured, as well as the use of the Washington Group questions.
- Initial consultations with IPs to understand their capacity and willingness to engage with the study for RQ1 and RQ2. Based on these, interviews to answer RQ1 will include IPs that did not target GWDs at inception, and a purposive selection of an equal number of IPs that did target GWDs from inception. RQ2 will include IPs who are able to engage with the study and have a strong focus on disability. Active engagement from IPs will support the IE team during vital aspects of the fieldwork, such as identifying GWDs, developing and adapting research tools, gaining greater knowledge of on-the-ground realities, as well as promoting uptake and relevance of study. Further shortlisting criteria will be developed to select projects in the instance there are more than 3 projects - this may include any one or more of the following: duplication of countries, size of the

potential sample of GWDs, and representation of geographical regions (e.g., including one project from each of the main GEC II regions – Sub Saharan Africa and South Asia, to the extent possible). In the event that there are fewer than three projects in the shortlist, the preliminary criteria will be revisited and relaxed to draw from a greater number of projects. In addition, we will consult closely with the FM and Regional Education Advisors (REAs) as part of the Evaluation Studies Working Group (ESWG), as well as with IPs themselves, to assess whether there are additional projects that may be relevant to the study, but which may otherwise have been precluded from inclusion.

b) **Stage 2** will involve designing the primary qualitative tools for RQs 1 and 2. For RQ1, this will relate to the IPs that demonstrate capacity and willingness to engage in key informant interviews. For RQ2, this will relate to the 2-3 shortlisted projects that have a strong focus on GWDs following further consultations with IPs, the ESWG and the Independent Advisory Group (IAG). The IE will also consult with OPDs to design the research tools and instruments. This stage will include further refining the research questions, sampling design, research tools and fieldwork plan for each case study. The methods of data collection will be designed in a way that is sensitive to the needs of GWDs and, is inclusive of different types of disabilities (to the extent that they emerge from the sample), which will be fostered through the engagement with OPDs.

c) Both stages will involve engagement with the IPs to help inform the study design. They will also involve consultations with data collection partners, Southern Academic Partners and the FCDO. At the end of this phase, a **Research Design Note** will be developed that will detail the overarching research questions, the proposed research tools and the shortlisted projects selected for the study.

Phase 2: Fieldwork.

The IE team proposes answering RQ1 through **key informant interviews** with IPs, including all IPs that did not focus on disability at inception and purposively selecting an equal number of IPs that did focus on disability at inception.

The IE team proposes answering RQ2 through **semi-structured interviews and /or focus group discussions, as well as non-participant observation of teaching and learning** (with the aim of using the narrative observation approach to understand how girls with disabilities interact with their teachers, peers, and other stakeholders in the larger educational arrangements they are a part of). The study will also draw from other more participatory sources of information such as visual data and audio diaries depending on the needs and abilities of girls with disabilities. The key stakeholders for data collection will include:

a) GWDs to gather their perceptions on how the interventions have affected their life trajectories, and on other multi-faceted outcomes such as learning and socio-emotional wellbeing. The research design will be adapted according to individual contexts keeping in mind the needs and rights of the GWDs. This will help in prioritising the voices of GWDs in a way that respects the principle of *"nothing about us without us".*

b) Educators/ teachers, to understand their perceptions of GWDs and any perceived and observed effects of the interventions on their practice. In addition to semi-structured interviews, we propose conducting non-participant observations to understand the effects of the interventions, e.g., pedagogical approaches adopted, changes in teachers' support.

c) Parents/ caregivers, to understand their perceptions of the benefits of schooling/ learning for GWDs.

d) School governance stakeholders (head teachers, school management committee leaders etc.), to understand their perceptions of how project activities have affected GWDs and barriers/ drivers of inclusion.

e) Key informant interviews will also be conducted with a wider range of stakeholders such as government representatives (national and local level), community members etc., to inform the political economy analysis.

All primary data collection will be guided by the safeguarding and ethical frameworks (see the GEC Independent Evaluation Ethical Research and Safeguarding Framework and see section 4 on

Research Ethics and Safeguarding). Primary data collection will be conducted with the support of our local data collection partners. The study will also engage persons with disabilities (for instance in the capacity of enumerators) to help with primary data collection in the field.

Phase 3: Analysis. This phase will include the analysis of findings from the qualitative transcripts of the primary data (interviews, focus groups discussions, etc) and triangulation with secondary data, where used. The analytical techniques used to answer each research question will be determined during Phase 1 of the study and will be completed based on the data collected during the fieldwork phase.

Phase 4: Validation of Research Findings. This stage will involve consultations with the key stakeholders such as selected IPs, research partners, Southern Academic Partners and the ESWG to validate the key research findings. In particular, the IE team proposes to hold a workshop with IPs to develop lessons learned and identify the transferability of findings to their programmatic contexts in a participatory approach.

Phase 5: Reporting. This phase will result in the development and dissemination of the key outputs of this study, including an emerging findings workshop, final report, and policy brief.

4. Research Ethics

1. All research and evaluation activities conducted as part of this study will adhere to the guidelines for ethical research as referenced in the Ethical Research and Safeguarding Framework.
2. The guidelines in the framework are developed to ensure that all primary research (involving individuals, stakeholders, or other programme stakeholders) is conducted ethically and in a safe manner. The Disability Study will be conducted in a way that gives precedence to the rights and dignities of its research participants and protects them from harm.
3. This ethical research framework is fully compliant with the guiding concepts and principles set out in FCDO's Evaluation Policy (2013) and FCDO's Research Ethics Guidance (2019)12; and the UK Data Protection Act (2018).
4. The Disability Study research design note will include an ethical research and safeguarding section pertaining specifically to this study. The ethical permissions will be applied for and adhere to the Cambridge Faculty of Education ethics process.
5. The process of obtaining all required government research permissions will be started as soon as the projects are shortlisted and countries are selected.

5. Key Limitations, Risk Assessment and Mitigation Plan

Table 2: Key Limitations, Risk Assessment and Mitigation Plan

Risk	Likelihood	Impact	Mitigating Actions	Impact following mitigation
Risks associated with fragile or conflict-affected areas make primary data collection unfeasible, and/or threaten the safety of GEC II cohorts and stakeholders.	Moderate	High	The GEC II portfolio and some IPs work with girls and schools threatened by FCAS, and this study aims (as a desirable inclusion criterion) to include one or more projects from FCAS. We will endeavour to monitor FCAS, through communication with the FCDO and FM, to inform decision making around primary data collection feasibility and safety.	Moderate
Low quality of project secondary data used to inform the research design.	Moderate	Moderate	The study team will liaise with the FM and IPs to obtain all the relevant data. After approval of the ToR, the first phase will be exploring the data to be able to frame what can and cannot be done as part of this study. The projects to be included in the final study will be chosen accordingly.	Low
Covid-19 results in school closures and hampers primary data collection.	High	High	We will factor the risk of school closures in our research design phase and include contingency plans (remote data collection by phone or online). We will liaise with IPs and country partners to make sure data collection is safe and make sure no data collection activities pose a risk to the health of any participants or team members.	Moderate
Delays in obtaining research permissions/ ethical approvals.	Moderate	High	We will use the primary data collection plan stage to develop positive relationships with stakeholders in proposed primary data collection countries and sites.	Moderate
Low sample size of girls with disabilities which may challenge identifying and building the sample, particularly if GWDs have emigrated due to Covid-19.	Moderate	High	We will focus on projects that directly target girls with disabilities and liaise with IPs to identify as many GWDs as we can. Additionally, as part of the fieldwork process, we will administer the WG questions to screen GWDs.	Moderate
Key stakeholders in the primary data collection countries may not be easily accessible during the data collection phase.	Moderate	Moderate	We will use the primary data collection plan stage to develop positive relationships with stakeholders in proposed primary data collection countries and sites and will factor in generous and flexible time allocation for primary data collection in case of delayed access to stakeholders.	Low

6. Work Plan

Figure 1: Work Plan

	2022												2023			
Activity	J	F	M	A	M	J	J	A	S	O	N	D	J	F	M	A
ES 4: Girls with Disabilities																
Developing Terms of Reference (TORs)																
Preliminary portfolio mapping																
FCDO workshop to review research questions																
ESWG meeting to review TOR																
Deliverable: TORs approved			◊													
Design Phase																
Data scoping																
Primary data tools development																
Research permissions and ethics approvals																
Deliverable: Design Note						◊										
Fieldwork Phase																
Primary data collection																
Data cleaning																
Deliverable: Fieldwork report												◊				
Analysis and Reporting Phase																
Data analysis																
Emerging findings workshop													◊			
Deliverable: Report draft													◊			
Deliverable: Final report															◊	
Deliverable: Policy brief																◊

6.1. Expected Deliverables

6. The **final ToR** for this study is due in March 2022.
7. Following approval of the ToR, a **Research Design Note** with details of the research design will be submitted for approval in June 2022.
8. A **Fieldwork Report** detailing the fieldwork process will be delivered to the FCDO in December 2022.
9. An **emerging findings workshop** will be conducted in January 2023.
10. A **draft report** presenting findings, conclusions, and recommendations for the FCDO, IPs and FM and wider stakeholder audiences as per the FM's GEC Learning Strategy will be delivered in January 2023.
11. **Final report** will be delivered in March 2023
12. **Final policy brief** will be delivered in April 2023.

7. Team Composition

13. This evaluation study will be led by a core study team under the guidance of the Principal Investigator and **Lead Author** (Nidhi Singal), **Research Associate** (Laraib Niaz) and **Lead Qualitative Analyst** (Romanshi Gupta). The study will be managed by **Programme Manager** (Louise Cathro) and supported by **Assistant Programme Manager** (Angela Nkonu). Additional support will be brought on as required to support with the data transcription, cleaning, coding, and analysis.
14. The Political Economy Analysis will be led by the **IE Team Leader** (Monazza Aslam) and **Deputy Team Leader** (Shenila Rawal).

15. Quality assurance processes will be overseen by the **Programme Director** (Simon Griffiths), **Technical Director** (Pauline Rose), **Team Leader** (Monazza Aslam) and **Deputy Team Leader** (Shenila Rawal).
16. Data collection, including enumerator training, fieldwork management and data quality assurance, will be managed by the **IE Fieldwork Manager** (Julia Midland). A wider team drawing on local data collection partners will also be assembled to support with in-country data collection. Southern academic partners will also be engaged throughout the study – including from outset during the research design phase – where they will provide analytical and advisory support to help inform and contextualise the research findings.

8. Stakeholder Engagement

17. The IE team will seek to interact with the following categories of external stakeholders in the course of the research (where relevant, some of these stakeholders will be consulted through the Evaluation Studies Working Group – ESWG):
    - FCDO UK;
    - FCDO Regional Education Advisors;
    - GEC II Fund Manager;
    - Girls' Education Department
    - Implementing Partners (IPs);
    - Beneficiaries of GEC II interventions; and
    - Other bilateral and multilateral agencies collaborating with GEC II or otherwise operating in the same sectors or thematic areas.
18. Engagement with IPs and receiving their inputs and feedback is a critical element of this study.
19. Engaging with OPDs to help with contextualisation of the research is an important part of this study.
20. Engaging with IPs of the selected projects will also ensure that we have a good understanding of their interventions, have identified relevant documentation and data for review and are able to triangulate findings for accuracy.
21. It is important that the IE team seek to engage with the beneficiaries who will be sampled for inclusion in the study. Engaging with project beneficiaries (such as GWDs and their teachers and parents/caregivers with whom GEC II projects have been working) will ensure they can meaningfully contribute to the individual studies to ensure that they can benefit from any evidence and learning generated – and the study team will seek to do this to the extent possible.
22. To ensure greater awareness about the research and continued engagement with key stakeholders, a communication strategy will be developed by the Fund Manager in collaboration with the IE team.

Annex B: Research Design and Methodology

1. Study design and methodology

1.1. Development of research design and questions

This section outlines the steps involved in the development of the research design which included an initial review of GEC II portfolio documentation, a rapid review of wider evidence and consultations with the FM/ FCDO.

1.1.1. Rapid review of wider evidence

A rapid literature review of key evidence and issues was initially undertaken to inform the framing and contextualisation of Study 4 in March 2022. This review was conducted using a purposive search strategy using the University of Cambridge's academic search engine (iDiscover) to identify recent research evidence around the education of GWDs. Evidence from the past 10 years has been examined using multiple sources ranging from academic literature (journal articles, book chapters etc.) to project evaluation reports published by Non-Governmental Organisations (NGOs) and international NGOs. This review of the literature helped to develop the research design of the study.

1.1.2. Initial review of GEC II portfolio documentation

An initial review of GEC II portfolio documentation (of all 41 projects) was conducted to understand how the projects have or have not supported GWDs through their interventions. This included firstly an examination of GEC II and FM documents to understand how GEC II projects support GWDs. In the context of GEC II, GWDs have the right to participate in all activities as active members of their communities. Individual projects (in both the LNGB and GEC-T windows) therefore adapt their interventions for the inclusion of GWDs. The implementing partners are also required to examine the barriers faced by GWDs in their individual contexts and undertake steps to mitigate these. The emphasis for inclusion in education is placed on reducing barriers and promoting opportunities for participation, rather than expecting girls with disabilities to 'fit in' as best they can (GEC, 2019: 3; GEC, 2020: 2).

A key activity in the desk-based document review was to undertake a detailed intervention and an internal activity mapping exercise. This mapping exercise resulted in a shortlist of projects Project design and implementation information was harvested from multiple project documents, including the theory of change, baseline, midline and endline evaluation reports. The mapping exercise resulted in a shortlist of 19 projects that focused on GWDs as shown in Figure 1.

Figure 1: Project Mapping Sheet according to interventions targeting GWDs

Note: This includes the projects that were reported to have some sort of disability focus. The interventions are only ticked when GWDs are specifically mentioned as either the direct or indirect recipient	Central Africa				East Africa											Southern Africa					West Africa				South Asia								
	Uganda	Uganda	Uganda	Uganda	Ethiopia	Ethiopia	Ethiopia	Ethiopia	Kenya	Kenya	Kenya	Kenya	Malawi	Somalia	Somalia	Mozambique	Zimbabwe	Zimbabwe	Zimbabwe Zambia Tanzania	Tanzania	Ghana	Ghana	Sierra Leone	Sierra Leone	Afghanistan	Afghanistan	Afghanistan	Nepal	Nepal	Nepal	Pakistan	
	Promoting Equality in African Schools (PEAS)	Viva	Cheshire Services Uganda	Opportunity International	Link Education International	Child4hope UK	People in Need	Population Council	Leonard Cheshire	World University Service of Canada (WUSC)	Education Development Trust (EDT)	ActionAid International	Link Education International	Relief International UK	Care International UK	Save the Children	World Vision UK	Plan International UK	Camfed International	Camfed	Plan International UK (formerly Varkey)	World Education Inc	Plan International UK	IRC	BRAC	Aga Khan Foundation	Aga Khan Foundation	VSO	Street Child	VSO	IRC	
GWDs - directly targeted through intervention		✓1	✓1				✓1	✓1	✓1			✓1	✓1	✓1	✓1			✓1			✓1	✓1	✓1	✓1	✓1	✓1	✓1		✓1	✓1		19
GWDs - indirectly targeted through intervention (e.g., project mentions marginalised girls such as...GWDs...)				✓1	✓1	✓1				✓1	✓1					✓1	✓1		✓1	✓1								✓1			✓1	11
Enabling access to school/learning for GWDs																																
Intervention: Transport provision		1	1						1													1										3
Intervention: Provision of inclusive learning resources		1	1						1													1		1					1			6
Intervention: Provision of assistive devices			1				1	1	1			1	1									1	1	1			1			1		11
Intervention: Adaptations to school/learning environment (e.g., accessible entrances/sanitary facilities)		1	1				maybe, to be checked	1	1												1		1	1						1		8
Intervention: Financial support/in-kind support		1	1		1				1					1	1			1					1		1	1	1			1	1	13
Intervention: alternative care support for resettled GWDs			1																													1
Inclusive practices (pedagogical/ accommodation) in classroom environment																																
Intervention: Modifications/ adaptations to curricula									1					1																		2
Intervention: Life skills curricula			1					1	1					1																1		5
Intervention: Psychosocial support			1						1																					1		3
Intervention: Peer activities/clubs						1			1					1																		3
Intervention: remedial/catch-up classes			1																			1								1		3
Intervention: Classroom set-up/seating arrangements																					1	1			1							3
Teacher guidance, training or recruitment																																
Intervention: Teacher Recruitment/PD for Inclusive Education (incl. a wide range of teachers; guidance to identify GWDs)		1	1	1	1		1	1	1	1		1		1	1	1	1	1			1		1	1		1	1		1	1	1	22
Intervention: Disability Training for Staff/Partners						1							1								1		1						1			5
Advocacy/outreach/awareness-raising																																
Intervention: capacity-building of stakeholders (school governance - SMCs/board of management)		1	1	1	1				1	1					1		1											1	1	1		10
Intervention: Engagement with Community Members/Families		1	1		1	1	1	1	1	1					1						1	1	1	1		1	1			1		16
Intervention: Engagement with other non-profit/disability organisations									1	1	1												1									4
Intervention: Advocacy/Outreach with Governments		1	1					1	1		1	1	1		1				1	1										1		11
Total by project	0	8	13	2	4	3	3	6	14	4	2	3	3	5	4	1	2	2	1	1	5	5	7	5	2	3	4	1	4	10	2	

1.1.3. Consultations with the FM/ FCDO

In addition to the initial project documentation and wider evidence review, FM and FCDO were consulted during the research design phase. This involved:

1. Consultations and feedback from the FCDO and Independent Advisory Group (IAG)) on the ToRs for Study 4 which outlined the scope, rollout, and proposed methodology. This included a workshop on the ToRs on 28 January 2022 and 10 February 2022 and multiple rounds of feedback from the FCDO and IAG to develop the ToRs.
2. Feedback from Roger Drew of the IAG and Ian Attfield of the FCDO on the project selection strategies for RQ1 and RQ2. Based on their feedback the data collection strategy for RQ1 was changed from a sample of 10 IPs (and associated KIIs) to a focus on all of the projects to generate more generalisable findings across the portfolio. The project selection strategy for RQ2 was approved.
3. Consultations between Nidhi Singal (Principal Investigator and Lead Author for Study 4) and colleagues at the FM. The FM provided feedback on project selection including which projects could be considered for inclusion/ exclusion from the shortlist for RQ2.

1.2. Research design and methods of data collection

1.2.1. Research strategy

Table 1 provides an overview of the types of data used to answer the RQs.

Table 1: Research strategy

Research Question	Projects included	Secondary Data Sources	Primary Data Methods/ Sources	Participants for primary data collection
RQ1: To what extent have GEC II projects supported GWDs through their interventions, including approaches adopted during COVID-19, and what are the factors influencing this?	All projects across the GEC II portfolio willing and with capacity to engage with the IE team	Project data and documentation	Key Informant Interviews	One person from each IP team (41)
RQ2: What are the perceived and observed effects of the interventions adopted by selected GEC II projects on GWDs (with a particular focus on their learning outcomes), their teachers/ educators, families/ caregivers, and communities?	Three projects identified based on sampling criteria given in	Project documentation	Key Informant Interviews In-Depth Interviews/ FGDs Classroom Observations/ other participatory qualitative methods	IPs, National and regional level stakeholders (government officials) GWDs, parents, caregivers, communities GWDs, Teachers

1.2.2. Research Question 1

This sub-section highlights the selection strategy and research methods and analysis for RQ1:

To what extent have GEC II projects supported GWDs (with a particular focus on their learning engagement and outcomes) through their interventions, including approaches adopted during Covid-19, and what are the factors influencing these decisions?

Project selection strategy

RQ1 was answered using secondary data analysis (both quantitative data and project documentation analysis) for all 41 projects on how the projects have supported GWDs through their interventions.

RQ1 was supplemented by KIIs with all IPs (one per project) depending on their capacity and willingness to engage with the study team. Of the 41 IPs contacted for this study, 27 IPs agreed to partake in an interview.

Quantitative Data Analysis

For the quantitative analysis, we looked at the projects' measures of intermediate outcomes, and the learning outcomes of girls with disabilities across the GEC II projects.

The datasets cover Phase II of the programme, which is operating between 2017 and 2025. Under this, 41 projects are receiving £500 million to support their activities across two windows: (1) the GEC-T window, which is supporting 27 successful GEC Phase I projects across 14 countries to transition to the next stage of their education; and (2) the LNGB window, which supports 14 projects in 10 countries to support up to 500,000 highly marginalised, adolescent girls who have never attended school or have already dropped out of school with literacy, numeracy and life skills. Of the 14 LNGB projects, we could use data from 12 projects.

The quantitative analysis is based on descriptive statistics, whereby we describe the data and how it compares across groups and across time. Where possible, we complement this with a statistical comparison of the averages, through t-tests of group differences. We also, where the data permits, look at a 'difference-in-difference' regression analysis, where we used ordinary least squares analysis on the changes in learning between data collection periods, compared to disabled girls who did not receive interventions (non-treated group).

The difference-in-difference estimates are used to try to estimate the (non-casual) impact of the programme on the learning outcomes of girls. This is estimated by first estimating the treatment group's midline (or endline) score minus the baseline score. This is the first difference. Due to the lack of a non-treated group in the LNGB window, this first difference is what we reported as an impact for LNGB. For GEC-T, we then calculate the difference for disabled girls who did not receive treatment. The difference between the two is the 'difference-in-difference'. 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 non-treated group.

We conducted this for both the cross-sectional sample (any girls with baseline or midline data) and panel sample (including the same girls recontacted in both periods) for GEC-T but only with a cross-sectional sample for LNGB (due to data quality issues discussed next). The difference-in-difference is conducted for all girls, and separately for disabled and non-disabled girls for comparison.

We cannot include any covariates in the difference-in-difference analysis as sample sizes drop significantly when covariates are introduced into the model. Hence, the difference-in-difference estimates show progress made by the treatment girls over the non-treated group, without regarding the differences they may have in their characteristics – which may bias the estimates upwards or downwards depending on how these characteristics intersect. As most of the excluded characteristics are working against disabled girls, this means that our coefficient is likely to be downward biased (as disabled girls are more likely to be poorer, and poorer girls are more likely to have lower learning outcomes).

The statistical significance of the difference-in-difference coefficients is reported for P-values of below 0.05 and 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.

Beyond a traditional 'difference-in-difference' regression, we compare the changes in learning between disabled and non-disabled treatment girls (rather than compared with a non-treated group), to investigate the extent to which disabled girls have made progress in learning compared to non-disabled girls. Therefore, this difference-in-difference estimate shows the progress disabled girls have made over and above non-disabled girls.

Further, we analyse changes in intermediate outcomes (such as life skills and teaching quality) by running a 'difference-in-difference' on these variables, including covariates of age and being overage (relative to grade

level). These regressions are run for all samples, and separately for disabled and non-disabled girls in GEC-T. The analysis was not done for LNGB as only a few projects have data on intermediate outcomes.

Table 2 below shows the different comparison groups utilised in the analysis. We compare disabled girls who received interventions with different groups, depending on the type of analysis.

Table 2: Comparison group for quantitative analysis

Area of focus	Analysis	Comparison group
Profiling	Prevalence of disability	No comparison group
	Baseline characteristics of girls with disabilities	All girls who received interventions (regardless of disability status)
	Reasons girls did not attend school	Non-disabled girls who received interventions
Examining changes in outcomes for girls with disabilities	Change in learning outcomes	For GEC-T, comparisons were made with: (i) Disabled girls who did not receive interventions (ii) Non-disabled girls who received interventions For LNGB, the following group was compared: (iii) Non-disabled girls who received interventions

Prior to discussing the findings, we highlight the sample composition and the main sampling decisions that are made due to restrictions in the underlying data.

Secondary qualitative data collection and analysis methods

In addition to quantitative data analysis, disability-specific project documentation analysis for all 41 projects was conducted to answer RQ1. The review of project documentation was conducted over the period of October and November 2022 by two consultants- Stephanie Nowack (for GEC-T projects) and Saba Saeed (for LNGB projects). This included examining the baseline, midline and endline evaluation reports and monitoring reports to collate the responses of GWDs, their teachers, caregivers, and the IPs to see the extent to which the projects supported GWDs (particularly their learning engagement) in their interventions. An important aspect of the secondary documentary analysis included an examination of the COVID-19 Medium-Term Response Plans to see how individual projects responded to COVID-19 and how or how not they adapted their interventions to suit the needs of GWDs given the impact of the pandemic. To analyse the data the following coding framework was developed (see Table 3).

Table 3: Coding framework for qualitative documentation (RQ1)

Theme 1: Conceptualisations of disability
Sub-theme
Washington Group
Ultra-marginalised girls
Physical impediment
Recognition of social barriers
Confusion about disability conceptualisation resulting in 'false' high rates of disability in sample
Theme 2: Are there any disability-specific interventions?
Sub-theme
Disability sensitive, but not disability transformative

Teacher-focused interventions
Supply of assistive devices and model schools
Services of a sign language research assistant
Sole focus on GWD: Teaching quality, attendance, self-esteem & life skills, community and parental attitudes, economic empowerment, policy, and governance
Theme 3: Political, social, and economic factors affecting the inclusion of GWDs in the education system
Sub-theme
Transport
Food insecurities
After-school clubs
Bullying
Community perceptions
Classroom inclusion
Lack of resources and infrastructure at school
Home chores
Theme 4: Methodological challenges to include GWD in projects
Sub-theme
"Falling through the cracks"
Small sample sizes (tokenistic inclusion or methodological challenges?)
Challenges in data collection methods
Data disaggregated by sex and age, but not disability
Lack of human, financial, and physical resources to include GWD in projects
Clear attempts to include GWD in projects - Future recommendations to include GWD in projects
Theme 5: Covid-19 response plans (low for GWDs) **'A new or revised project-specific Theory of Change was not developed in light of Covid-19. As such, the endline evaluation explores how the original Theory of Change was maintained despite the school closures.' [Avanti Communications Group]**
Sub-theme
Cash transfers
Narrowing technological gap
Distribution of learning packs

Primary data collection methods

Quantitative and qualitative secondary data for all 41 projects was supplemented by primary data to answer RQ1. The primary data collection method for RQ1 was KIIs with project IPs. These KIIs were held to explain and understand which factors influenced the decisions of the IPs to focus on disability and focused on the following thematic areas (this list is not exhaustive and further thematic areas were developed):

- Rationale and motivation of project IPs for focusing on GWDs;
- To what extent have IPs integrated GWDs in their interventions; and

- Reasons for a shift in the focus on GWDs (from baseline to midline).

All KIIs lasted up to an hour and were conducted in English. The interviews were audio recorded (after due consent was taken) and transcribed in English.

The key informant interview transcripts were analysed using a thematic approach. A coding framework was developed after the interviews were ready for analysis. Given the substantial amount of data, which was generated with the interviews, we streamlined and created a thematic coding framework based on the interview questions instead of using open coding. This involved the Senior Research Lead and Qualitative Analyst of the Study 4 team, Laraib Niaz and Romanshi Gupta, reading the interview transcripts and making sense of them. As a second step, they went through the data line by line to code and then categorise these codes to understand the themes emerging from the data.

1.2.3. Research Question 2

This sub-section highlights the selection strategy and research methods and analysis for RQ2:

What are the perceived and observed effects of the interventions adopted by selected GEC II projects on GWDs, their teachers/ educators, families/ caregivers, and communities?

Project selection strategy

RQ2 was explored through in-depth qualitative case studies where the IE team collected primary data. To shortlist projects for these case studies, we developed selection criteria as outlined in Table 4 below.

Table 4: Selection criteria for case studies with IPs (RQ2)

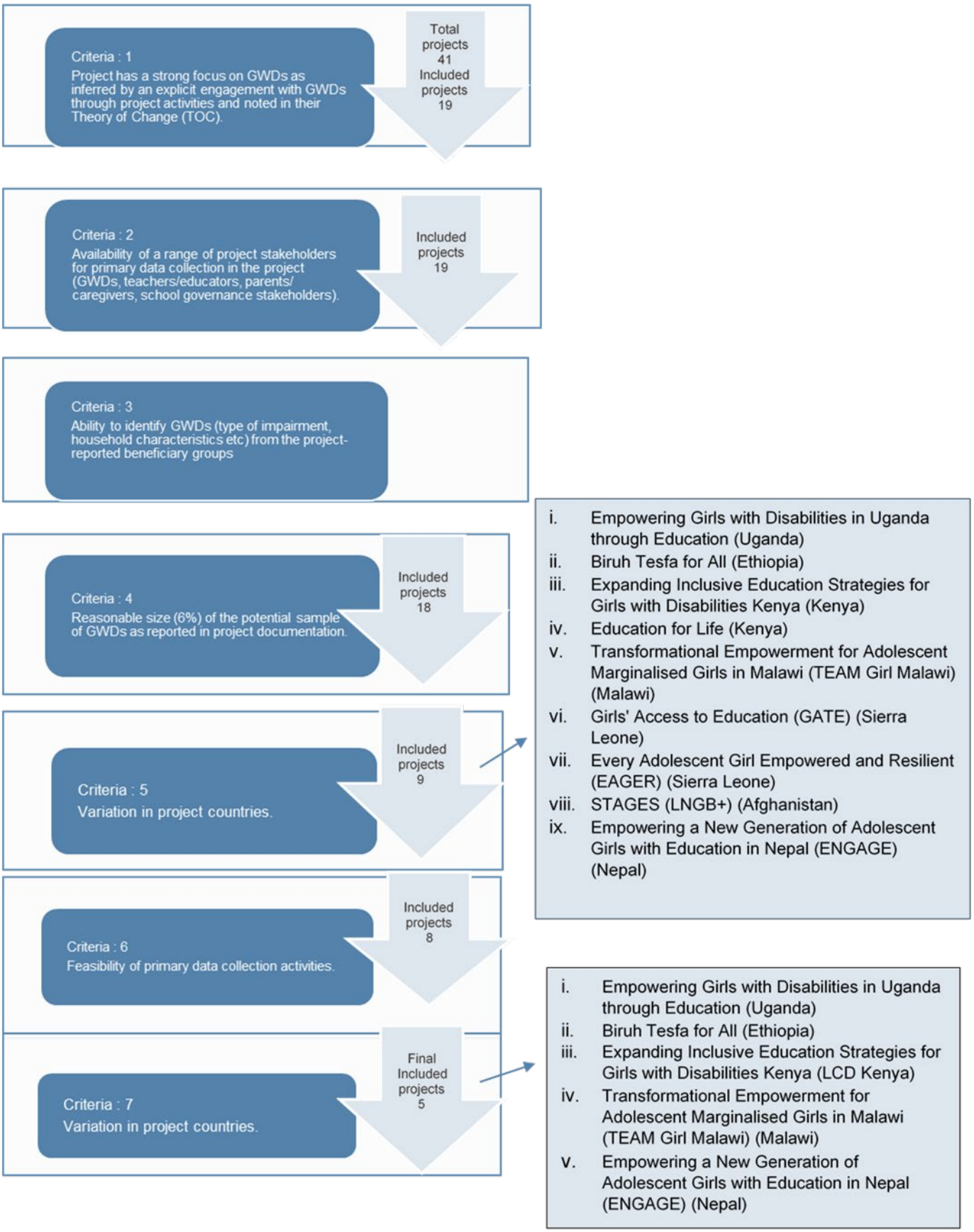
Criteria	Essential	Desirable
Project has a strong focus on disability as inferred by an explicit engagement with GWDs through project activities and noted in their Theory of Change (ToC).	✓	
Availability of a range of project stakeholders for primary data collection in the project (GWDs, teachers/ educators, parents/ caregivers, school governance stakeholders).	✓	
Ability to identify GWDs (type of impairment, household characteristics etc.) from the project-reported beneficiary groups.	✓	
Reasonable size (6%) of the potential sample of GWDs as reported in project documentation.	✓	
Feasibility of primary data collection activities	✓	
Variation in project countries.	✓	
Capacity and willingness of the selected IP to engage with the IE team during the timeframes of the study.	✓	
Avoiding duplication with other IE studies	✓	
Prevalence of different types of disabilities among the GWDs.		✓
Capacity and willingness to engage with the IE team.		✓
Diversity in the proportion of GWDs in the target beneficiaries.		✓

Based on the above selection criteria, projects were included and excluded from the final shortlist. The shortlisting process is explained below.

1. **Strong focus on disability:** The review of project documentation and ToCs showed that 19 out of 41 projects had a strong focus on disability at baseline. These 19 projects were therefore included in the shortlist. These projects offer an opportunity for in-depth learning about interventions supporting GWDs.

2. **Availability of stakeholders and ability to identify GWDs:** 19 projects fulfilled this criterion and hence all were included. This is essential to ascertain which projects have reported the prevalence of GWDs within their beneficiary groups. Further details (types of impairment, household characteristics) will support the sampling.
3. **Reasonable size of the potential sample of GWDs:** For the 19 projects, we calculated the median value (6%) as a proxy for reasonable size. We did not calculate the mean as there were a few outliers in the data (0.17%, 100%) that would have skewed the mean value. Using this metric, nine projects out of 19 were selected. This has enabled us to identify a sample of GWDs from within individual projects, particularly in consideration of risks of attrition/ relocation.
4. **Variation in project countries:** Two projects in Kenya were identified based on our initial criteria. However, *Expanding Inclusive Education Strategies for Girls with Disabilities* is considered a better fit given that it covers all five stakeholder groups (GWDs; parents/ caregivers; community members; other NGOs, OPDs; government representatives), and has a larger sample of GWDs as a proportion of total project beneficiaries (100%). In contrast, *Education for Life* only covered two stakeholder groups and has a smaller sample of GWDs (30%). In Sierra Leone, two projects were identified – Plan International UK GATE and IRC EAGER – both of which were selected for data collection in the IE's Study on *Teachers and Teaching* (Study 2). Since IRC EAGER was previously included for primary data collection in Study 2, we chose to exclude this project in an effort to **avoid duplication of projects** across the studies. While primary data collection eventually did not occur with Plan International UK GATE for Study 2, the project closed in July 2021, which reduced the likelihood that we would be able to identify and collect data from beneficiaries – particularly given the risks of attrition and relocation due to COVID-19. Thus, we chose to exclude both projects from the list.
5. **Feasibility of primary data collection activities:** Given the ongoing disruptions in Afghanistan, it was deemed too high risk to consider conducting fieldwork in Afghanistan. As such, it is not possible to include STAGES (LNGB+) as a case study.
6. **Capacity and willingness to engage with the IE team:** from the resulting shortlist, Population Council's Biruh Tesfa for All in Ethiopia was excluded due to a lack of capacity to engage with the IE team.
7. **Diversity in the proportion of GWDs in the target beneficiaries:** Since LCD Kenya and Cheshire Services Uganda (CSU) Uganda both have 100% and 99% samples of GWDs, CSU Uganda was excluded, and LCD Kenya was included to allow for diversity in GWD sample sizes. This would allow the final three projects to have varied samples of GWDs (translating to 2260, 1000, and 538 GWDs in Kenya, Malawi, and Nepal respectively). LCD Kenya was chosen because of its bigger sample overall and range of interventions being adopted, and stakeholders being targeted.

Figure 2: RQ2 Project selection process



After the shortlisted projects were approached, two of them agreed to participate in the study while the third project, LCD Kenya declined. We then decided to consult with the FM and FCDO to decide which projects to approach next. We presented two options to them, which included Plan International UK's project Girls' Access to Education (GATE) in Sierra Leone and Cheshire Services Uganda (CSU) Uganda's project Empowering Girls with Disabilities in Uganda through Education.

We were advised to approach Plan International UK's project, but we did not receive any response from them. We then approached the only other GEC-T project with a substantial sample of GWDs (170) and a range of interventions targeting GWDs (learning centres, special/ resource schools, teacher training, community awareness campaigns) - VIVA Uganda's Building Girls to Live, Learn, Laugh and 'SCHIP' in Strong, Creative, Holistic, Inclusive, Protective, Quality Education which agreed to participate (see Table 5).

Table 5: Project shortlist for RQ2

Country	Implementing Partner	Project	GEC-T /LNGB
Malawi	Link Education International	Transformational Empowerment for Adolescent Marginalised Girls in Malawi (TEAM Girl Malawi)	LNGB
Nepal	VSO	Empowering a New Generation of Adolescent Girls with Education in Nepal (ENGAGE)	LNGB
Uganda	VIVA	Building Girls to Live, Learn, Laugh and 'SCHIP' in Strong, Creative, Holistic, Inclusive, Protective, Quality Education	GEC-T

Primary data collection and analysis methods

RQ2 was answered using a primary data collection approach, across different stakeholder groups: GWDs, their parents/ caregivers, teachers, community groups and government or national agency representatives. This multi stakeholder approach ensured multiple, diverse perspectives from relevant groups and individuals on the perceived and observed effects of project interventions on GWDs and those around them. A range of primary qualitative methods (semi-structured interviews, participatory methods, focus group discussions and observations of educational arrangements) were used to collect data for RQ2.

Focus Group Discussions

FGDs were conducted with two groups of stakeholders - girls and community members. A focus group interview is described as "a carefully planned discussion designed to obtain perceptions on a defined area of interest in a permissive, non-threatening environment" (Krueger and Casey, 2009: 2). In this sense, FGDs can be used to provide a safe peer environment, especially for children.

FGDs with girls were conducted with both girls with and without disabilities together. The focus group discussions were conducted with groups of eight to 10 girls. The IE team endeavoured to identify and mobilise an equal mix of GWDs and girls without disabilities, where possible. These FGDs were conducted in a manner that was inclusive of the disabilities that the girls faced, with exercises being suitable and adaptable for various types of disabilities.

FGDs with community members were conducted to understand their perceptions of disability and changes in attitudes, if any, after the interventions. These FGDs were used to elicit the community members' perspectives on how well the interventions work for GWDs, their families and their communities.

Semi-structured Interviews

An important aspect of this study was one-on-one **SSIs with GWDs, their parents/ caregivers, teachers/ education providers and government officials**. Interviews allowed the researcher to collect open-ended data, explore participant thoughts, beliefs, and feelings about the interventions and delve deeply into personal and sometimes sensitive issues by using prompts and then follow up questions and probes. The Principal Investigator of Study 4, Professor Nidhi Singal, has substantial experience researching Children with Disabilities, which was crucial in the design and analysis of the study. SSIs, therefore, were useful for eliciting the thoughts and perceptions of a range of stakeholders for RQ2.

We developed an SSI schedule with probes and prompts, and each interaction lasted between 45 minutes and an hour. These were audio recorded (after due consent was taken). They were all conducted in local languages (Nepali in Nepal, Chichewa in Malawi, Luganda in Uganda) and then translated into English. Please see section 2.1.1 for target and achieved samples for SSIs.

Table 6 includes a list of thematic areas under which questions were asked from the different stakeholder groups (this is not an exhaustive list, but an illustrative one).

Table 6: Thematic breakdown for semi-structured interviews

Stakeholder Group	Themes
GWDs	Nature of the engagement with the intervention. How has the intervention helped GWDs access education (and overcome barriers). Perceived effect of the intervention on learning. Perceived effect of the intervention on confidence and other socio-emotional skills. How the project changed their aspiration for the future.
Parents/ caregivers	Awareness of the impact of the intervention on the lives of GWDs. Nature of the engagement with the intervention. How has the intervention helped GWDs access education (and overcome barriers). Perceived effect of the intervention on GWDs' outcomes (learning, socio-emotional skills). Context specific questions pertaining to PEA. How the project changed GWDs' aspirations for the future.
Teachers	Type and level of engagement with the intervention. How has the intervention helped GWDs access education (and overcome barriers). Effect of the intervention on teaching processes. Effect of the intervention on teachers' attitudes and confidence in teaching GWDs. Effect of the intervention on GWDs learning. Effect of the intervention on GWDs' confidence and other social skills. Context specific questions pertaining to PEA.
Government officials	Context specific details to inform PEA, such as: • Political and economic factors which influence the inclusion of GWDs • Social and cultural norms which influence the inclusion of GWDs • Stakeholders related to disability inclusion and their incentives or disincentives • Additional barriers to/ drivers of change regarding disability inclusion

Participatory methods

Other than FGDs and SSIs, the **participatory methods of photovoice and audio diaries** were used to stay true to the objective of the study - to listen to the voices of GWDs. These methods needed adaptation according to the sample of GWDs (and the types of impairments faced by the GWDs).

Photovoice is a participatory technique that allows participants to record and present photographs where the photograph represents information, feelings, or memories (Shaw, 2020). In this study, GWDs were given prompts where they were asked to present photographs based on those prompts such as "What do I like about my school/ learning centre". In this sense, the photographs provided a tool with which to conduct an interview. The analysis did not centre on the photographs but rather on the girls' voices around what the

photographs represented. Similarly solicited audio diaries, especially with girls who are unable or do not wish to use a camera, allowed a closer look into the lives and perceptions of GWDs. Participants were given the same prompts as those for photovoice and were free to structure their entries in whichever format they wanted, for example as structured information or more abstract reflections (for up to three minutes). Audio-recorded diaries have the advantage of being more amiable in contexts where there is a prioritisation of the oral culture.

Audio diaries (for girls with visual impairments) and photographs (for girls with hearing impairments) were uploaded on a secure platform and were anonymised. We provided the GWDs with disposable cameras to take pictures in the school setting. Similarly, for audio diaries, we provided them with recording devices.

Observations in educational arrangements

Observations were undertaken depending on the educational arrangements covered in the targeted programmes. These observations were semi-narrative; clear guidance was provided to the researchers on the areas to focus on and make notes about. In one case in Nepal, where there were no formal classes (it was a day-care centre), we required the enumerators to visit the intervention sites to make field notes. Enumerators were given guidance on the focus of these observations, which was shaped by Spradley's (1980) guidance on observation. Spradley (1980) provides a systematic approach (consisting of nine steps) through which observations are conducted as an "ethnographic research cycle" rather than a linear sequence. Step one for instance consists of locating a social situation (in this case an educational arrangement), step two consists of participant or non-participant observation (in this case non-participant observation of GWDs in the educational arrangement), step three being the recording observations in fieldnotes (key to writing these fieldnotes are three principles - the language identification principle, the verbatim principle, and the concreteness principle).

The sequencing of these methods – holding focus group discussions prior to individual interviews and participatory methods and holding observations last – as well as the duration/ timing of these methods was carefully considered before fieldwork commenced and depended on the individual project contexts. An overview of the respondents and methods for the evaluation study is illustrated in Table 7.

Table 7: Summary of methods and respondents

Respondents	Primary qualitative methods			
	FGDs	Individual/ KI interviews	Participatory Methods	Classroom/ learning centre observation
Girls	X	X	X	X
Teachers		X		X
Parents/ caregivers		X		
Community groups	X			
National/ Regional government officials		X		

The study team worked collaboratively with IPs, OPDs, SAPs and data collection firms in developing the design of the research tools and methods to ensure IPs' priorities and experience were accommodated within the design. Consideration was made of context, and the tools were adapted according to the type of disability.

Data analysis

This study involves multiple approaches to qualitative analysis, including content and thematic analysis to bring rich contextual detail to findings drawn from the individual case studies and to bring multiple voices to the research questions and overarching area of inquiry.

The focus of analysis and reporting for RQ2 is on the three GEC II projects using primary qualitative methods. The analysis focuses on developing contextually nuanced insights into shortlisted projects. While data is

presented thematically, and around key research questions and findings, the analysis is conducted primarily at the level of the project.

Coding was undertaken by the Study 4 team and ensured that large amounts of qualitative data became manageable units for analysis. Coding was conducted in three stages:

1. Open coding (where initial codes were assigned to the data collected);
2. Axial Coding (where initial codes were reviewed to see which ones were being frequently used); and
3. Thematic coding (where axial codes were reviewed to generate themes from the data).

Coding was an iterative process, and the framework for coding was refined as the team got a deeper understanding of the information emerging from the primary data. Once coding had been completed, thematic analysis was undertaken across the databases. Furthermore, a deeper case study analysis was undertaken to understand the effects of the interventions for GWDs and those around them. Thematic analysis was conducted to facilitate comparison between projects while also highlighting rich details on individual project contexts, barriers, and interventions.

1.2.4. Cross-cutting themes: Political Economy Analysis

A political economy lens was applied to inform the primary objective of this study which was to understand in what ways GEC II projects have engaged GWDs through their interventions, and the perceived and observed effects of these approaches on their outcomes and on the engagement of GWDs with their peers, teachers/ educators, and caregivers/ families. Specifically, the PEA aimed to further inform the findings of this study (in relation to the two research questions specified) and in terms of the following aspects:

- **Political, economic, socio-cultural environment and other wider influences within the contexts in which these programmes were operating:** an explanation of the political, social, and economic climate that affected the inclusion of people with disabilities in the country in which the project was operating as well as any regional and /or local contextual issues that may positively influence or hinder project strategies and outcomes related to the inclusion of people with disabilities.
- **Stakeholders, bargaining mechanisms and power dynamics, and incentives/ disincentives:** identifying who the key stakeholders regarding inclusion and disability are, how they are organised and how they relate to each other, bargaining processes amongst different stakeholders and finally, motivating factors that influence these stakeholders and the incentives/disincentives they faced to include people with disabilities.

This analysis aimed to explore how each of the above elements can be a driver of or barrier to change in terms of how they influence whether and to what extent projects are able to support GWDs.

The PEA leads (Dr Monazza Aslam and Dr Shenila Rawal) worked with the core research team to integrate questions pertaining to these aspects into their primary data collection and analysis efforts. PEA-related questions were incorporated into KIIs with IPs for RQ 1. PEA-related questions were also included in the qualitative data collection instruments (KIIs, FGDs etc. with IPs, national level stakeholders e.g., government officials at the national and district levels, GWDs, parents, caregivers, communities, and teachers etc.) for RQ2. This included adding questions related to the support provided by the government and communities as well as other organisations such as OPDs.

The PEA for the three short-listed country contexts was also informed by secondary data analysis based on evidence gained from a literature review, country and regional contextual research, and documentary analysis (programme reports, documentation, national reports, etc.) which was conducted by a consultant (Thilanka Wijesinghe). All PEA related information (data from the transcripts), review of secondary documentation and review of policy frameworks were handed over by the consultant to the PEA leads who then developed a PEA report. Segments from the PEA report were then used in the Study 4 Final Report.

2. Fieldwork management and primary data analysis

2.1. Sampling process

The sampling approach for primary data collection was purposive and iterative and developed by the Study 4 team in consultation with IPs.

We attempted to follow the methodology detailed below to select specific locations for research, as well as select specific GWDs.

1. Calculate the number of GWD in each district using the information of beneficiaries given to us by the IPs[1].
2. Select two districts with the largest number of GWDs.
3. Select sampling areas (municipalities/ schools/ learning centres) with the highest number of GEC-supported GWDs.
4. Random sampling was employed to choose the sample of six GWDs (according to the proportional representation of disability in girls at the chosen school/ learning centre where possible)[2]. The sampling frame consisted of schools/ learning centres with the highest number of GEC- supported GWDs. Random sampling was conducted by using the randomisation function on Microsoft Excel.
5. Two of the selected girls in each location were randomly selected for the case study and would participate in all four research tools – participatory methods, SSIs, FGDs, and classroom observations while the rest of the four girls would participate in SSIs and FGDs.

After selecting the GWDs in each location, the primary caregiver/ parent of each selected girl was automatically selected for their respective interview. GEC-supported teachers/ educators were randomly selected from lists provided by the IPs. Classroom observations were completed in spaces that included at least one case study GWD and at least one of the selected teachers/ educators.

We did not draw specific samples for either the community leader FGDs or the national/ district-level officials. Instead, we worked closely with the IPs to purposively select relevant individuals for this research.

Specific sampling information for each country is detailed below.

Nepal

In Nepal, VSO implemented their ENGAGE project in three districts, Parsa, Banke and Sarlahi. Following the methodology outlined above, Parsa and Banke districts were selected for research because they had the highest number of GWDs benefitting from GEC programming.

Within each district, we calculated the number of GEC supported GWDs in each participating school throughout the district and selected the two schools with the highest sample of girls. The sample of girls in each of these schools was very small (with most schools having between one and seven GWDs each). In Parsa, we selected Sri Ne Ra Ma Bi Bhikhampur Pipra in Jagarnathpur and Sri Trijudha Mahabir Pd Ragubir Ram Secondary School in Birgunj Sub Metro. In Banke, we selected Adarash Basic School Khajura in the municipality of Khajura and Samjhana Learning Center Baijanath 5 in the municipality of Baijanath.

To select specific girls for research in Nepal, we then listed all the GEC-supported girls for each school and selected randomly. Since the sample of GWDs in each school was low, there was no need to select a sample of girls from within the school for all schools as all girls were automatically selected. We aimed for a sample of six girls from each GEC-supported school. The time available for fieldwork in Nepal was limited due to the national holidays in October. Additionally, it was difficult to collect data from girls with intellectual and severe disabilities since they became exhausted easily. Therefore, there were some sample replacements for the

[1] Important to note that there are limitations to these data as the IPs will only tend to identify girls that are known to schools/ administrators. Other girls who might be a part of the projects may be excluded.
[2] In practice this was hard to do in all locations other than Malawi where the sample of GWDs with functional impairments in each learning centre was high |(between 10 and 30). In Nepal the sample was too small in each school as was the case in most schools in Uganda.

girls. We replaced a girl from within the shortlisted school where possible. This did not particularly impact our research design since our sampling criteria was school-specific i.e., we were shortlisting girls from the schools/ learning centres with the highest amount of GWDs.

Malawi

In Malawi, Link Education International's project Team Girl Malawi operates in three districts, Dedza, Lilongwe and Mchinji. Following the methodology outlined above, Dedza and Mchinji were selected for research because they had the highest number of GWDs benefitting from GEC programming. Within each of these districts, we selected the municipalities with the highest number of GEC-supported GWDs. In Dedza, we selected two municipalities, Chimkombero 3 and Mapuyu, which together included 68 GEC-supported GWDs.

Following the above protocol, in Mchinji the two municipalities with the highest number of GWDs, Mdzomba 3 and Mwase 3, were selected. However, the Community-Based Education (CBE) meeting point in Mdzomba 3 only included girls with visual impairments. To obtain a more diverse sample of GWDs, we selected the municipality of Mponda instead. From within the chosen CBE meeting points, we selected the main respondents (GWDs) on the basis of proportional representation with respect to the type of disability they are facing. This in practice was difficult to achieve since the sample size of GWDs across impairment types and severity across other contexts. Malawi was the only context where the sample of GWDs was large enough for us to choose GWDs according to proportional representation of the type of disability.

For each municipality, we needed to select six GWDs. We first determined the number of girls with each different type of disability (hearing, seeing, cognition, or walking[3]) and then calculated the number of girls we would need to select in order to achieve proportional representation by disability. To select specific girls for research, we then listed all GEC supported girls and selected them randomly, in line with disability proportions. Two of the selected GWDs were then randomly selected for the case study.

Uganda

Viva CRANE's GEC II project Building Girls to Live, Learn, Laugh and SCHIP (Strong, Creative, Holistic, Inclusive, Protective, Quality Education) operates in eight districts, Arua, Buikwe, Gulu, Kampala, Luweero, Mukono, Nakaseke, and Wakiso. Kampala and Wakiso districts were selected for research using the methodology detailed above. In Kampala, we then selected Kawempe and Rubaga municipalities. In Wakiso, we selected Masuliita and Kakiri.

In Kampala, there is one operational school each in the chosen municipalities, Pearls of Africa Special Needs Centre in Kawempe and Mukisa in Rubaga. Similarly, in Wakiso there is one operational school each in the chosen municipalities, New Hope in Kakiri and New Hope School in Masuliita. These four schools were purposively sampled for research.

The selection of the six GWDs for research was a mix of purposive and random sampling. Since the sample of GWDs in each school was small, we could not purposively select girls according to a representation of the type of disability. We instead chose three girls who were a part of the formal schooling pathway and three who were a part of the vocational pathway from each school. The three individual girls in each stream were listed and selected using random sampling. Random sampling was conducted by using the randomisation function on Microsoft Excel.

Note that Ebola and Ebola-related school closures significantly impacted our sample. First and foremost, our timeframe for completing the fieldwork was cut short due to school closures mandated by the Government of Uganda to control the spread of Ebola. These closures meant that schools had to move exam schedules forward, limiting our ability to work inside of the selected schools. As a result, we worked closely with the IP to identify schools that would allow us to conduct research with girls and teachers, and the IP worked closely with school administrators to bring girls and their caregivers at specified times outside of school hours so that research could be completed quickly and efficiently. Furthermore, these exams and school closures prevented us from completing the target sample of classroom observations. The completed sample, therefore, is significantly different to the original sample. The details above therefore describe the initial, intended, sampling process. We were however able to conduct research in the selected schools that we shortlisted

[3] The Washington Group questions incorporate six domains **seeing, hearing, walking, cognition, self-care, and communication**. We only focused on the first four visible ones as the remaining two were less reported by the IPs.

through our sampling design which was the same in all three country contexts. Hence even though there were replacements of girls within the schools in the end and differences in the final sample, the selection strategies for the schools/ learning centres were the same in all three countries.

2.1.1. Target and achieved sample

All fieldwork was completed face-to-face using local, trained interviewers (see Section 2.2.1) who were familiar with the operating context. Target and achieved sample for each type of research, by country, is displayed in Table 8.

Table 8: Target and achieved sample, by country

Research Type	Target Sample	Achieved Sample		
		Nepal	Malawi	Uganda
SSI Girls	**24**	24	24	23
SSI Parents	**24**	28	24	23
SSI Teachers	**16 - 24**	23	24	16
FGD Girls	**8**	8	8	8
FGD Community Members	**8**	8	8	8
Photo/Audio Voice	**8**	7	8	7
Classroom Observation	**12**	14	12	6
Local Government KII	**2**	3	2	2
Central Government KII	**2**	2	2	3

Table 9 below displays the total number of GWDs interviewed per country, disaggregated by disability type as defined by the IPs.

Table 9: Types of disabilities included, as identified by the IPs by country

Type of Disability	**Number of Interviewed Girls**			
	Nepal	**Malawi**	**Uganda**	**Total**
Hearing impairment	10	4	5	**19**
Marginalised	3	0	0	**3**
Physical impairment	4	4	0	**8**
Visual impairment	7	8	0	**15**
Multiple impairments	2	2	0	**4**
Severe/Profound	4	0	0	**4**

Type of Disability	Number of Interviewed Girls			
	Nepal	Malawi	Uganda	Total
Intellectual disability	1	0	9	**10**
Learning disability	0	6	7	**13**
Speech Impairment	0	0	1	**1**
ADHD	0	0	1	**1**
Emotional disorder	0	0	1	**1**
Total sample	**31**	**24**	**24**	**79**

2.2. Designing research instruments

The primary data collection tools were developed iteratively in response to the overarching questions described above.

The key informant interview guides for RQ1 were designed based on the review of project documentation to explore *how* and *why* IPs supported girls with disabilities as part of their beneficiary groups.

The tools for RQ2 were preliminarily developed in accordance with the literature on inclusive research methods and child-focused research methods and were centre-safe, ethical, and participatory research. Consultations with the shortlisted IPs, as well as with the FM/ FCDO and SAPs enabled the tools to reflect stakeholders' learning priorities and this was conducted over a period of two months between July and August. Specifically, the IPs and SAPs helped in making the tools context-specific.

All the primary data collection tools were tailored to each project/ national context and research question as appropriate, to ensure they were culturally appropriate and responding to the study's research questions. In addition, data collection tools were tailored to projects' activities and interventions to ensure that they respond to the research questions of this study. While the tools were context-specific, we ensured that as far as possible, they were standardised to facilitate comparisons at the analysis stage. All tools were translated from English into the local languages (Nepali in Nepal, Chichewa in Malawi, and Luganda in Uganda) to facilitate data collection.

We also included the following types of adaptations:

- For girls with visual disabilities, we adapted the tools to Braille and/ or had qualitative researchers/ facilitators read out the questions and record oral responses.
- For girls with hearing disabilities, we used sign language interpreters.
- For girls with speech disabilities, we tried to obtain written responses.
- For girls with physical disabilities, we closely liaised with the Fieldwork Manager and the data collection partners to ensure that accommodations are made for any FGDs/ interviews.

Finally, tools were tested through a pilot study in each of the three countries (see Section 2.3).

2.2.1. Local data collection partners

In each shortlisted country, the IE team identified and contracted a local partner. For this study, the IE team liaised with local partners to ensure that the researchers had the specialist expertise in the required data collection methods (Braille interpretation, sign language interpretation, experience collecting data from people with disabilities). These local partners were responsible for recruiting qualitative researchers and facilitators

with the relevant skills (including language skills, Braille/ sign language interpretation skills, as well as professional competencies; and locally recruited female enumerators) and at least a few years of experience required for the primary data collection.

The local partners were:

- Nepal: Rooster Logic;
- Malawi: Research Options; and
- Uganda: Research Plus.

2.2.2. Engagement with SAPs

Throughout the duration of this study, we engaged with the Southern Academic Partners (SAPs), who were experts in the field of disability, on an ongoing basis. The SAPs were:

- Niraj Poudyal and Institute for Integrated Development Studies (IIDS) in Nepal;
- Jenipher Mbukwa and Centre for Social Research (CSR) for Malawi; and
- Jenipher Mbukwa for Uganda.

The SAPs were asked for feedback during the preparation of the research design note, on the fieldwork tools and the Study 4 Draft Report.

2.2.3. Ethical research and safeguarding

The GEC Independent Evaluation Ethical Research and Safeguarding Framework[4] forms the overarching ethical framework for all research and data collection protocols for the GEC II IE. These guidelines relate to the design, implementation and reporting of all activities conducted as part of the IE. The Ethical Research and Safeguarding Framework is compliant with the guiding concepts and principles set out in the FCDO's Evaluation Policy (2013) and the FCDO's Research Ethics Guidance (2011); the DFID Ethical Guidance for Research, Evaluation and Monitoring Activities (2019); and the UK Data Protection Act (2018).

All necessary research permissions were obtained from National Council for Research in the Sciences and Humanities (NCRSH) in Malawi and Mildmay Uganda Research and Ethics Committee (MUREC) in Uganda prior to data collection taking place. There were no national level requirements for research permissions in Nepal. These were managed through our local data collection partners in each of the sampled countries, who submitted the research application and managed all processes associated with gaining the approvals needed. In addition to this, we also received ethical approval through our consortium, partner, the REAL Centre at the University of Cambridge. Where necessary, we requested FCDO support in the provision of a letter, or other documentation, to support our applications.

All research and evaluation activities conducted as part of this study adhered to the guidelines for ethical research as referenced in the GEC II IE Ethical Research and Safeguarding Framework. These guidelines were developed to ensure that all primary research (involving individuals, stakeholders, or other programme stakeholders) is conducted ethically and safely.

Study 4 was conducted in a way that gave precedence to the rights and dignities of research participants and protected them from harm through:

- Developing ethics forms (including consent/ assent forms) and protocols with our local data collection partners and consulting with IPs participating in the research.
- Training enumerators in the use of these forms and protocols and piloting them at the same time as piloting the research tools.
- The inclusion of specialised training for working with marginalised populations and sensitive subjects for all enumerators, supervisors, and DCP staff. This training offered specific considerations and protocols for working with minors, for working with women and girls, and for working with people with disabilities.

In cases where an enumerator was led to believe that a child was at risk of serious harm, action was taken to report this concern to the relevant IP and to the study team who in turn submitted a safeguarding report to FCDO and the FM. The interviewer reported the concern to her team supervisor, disclosing only the nature of

[4] A copy of the GEC IE Ethical Research and Safeguarding Framework is available upon request.

the concern while maintaining the anonymity of the girl. The supervisor then reported the concerns to both the IP and the DCP management staff, who filed a written safeguarding report, which was submitted to Tetra Tech. Following the procedures outlined in the Ethical Research and Safeguarding Framework, the Tetra Tech Safeguarding Lead submitted a report on the concern raised to the FCDO and the FM. Throughout the course of fieldwork in all three countries, a total of four reports were filed, two for Malawi and two for Nepal. In response to the reports, IPs were informed, and necessary action was taken in the field, for instance informing the GWD's parents and family members if there was a risk of harm.

2.2.4. Data management

Original copies of primary data were stored and organised to facilitate retrieval and analysis at the analysis stage, with data protection and privacy security checks (such as password protected access and encryption where necessary). The IE catalogued the data including details such as time, date, and location of data collection, language of data collection, duration of interview/ FGD, critical identifiers for all respondents, and other relevant pieces of information. Data translation, transcription, and cleaning was conducted by the local data collection partners. Consideration was made of how and when tools were translated from English into local languages, how primary data was recorded – e.g., hand-written notes, typed notes, audio recording, visual recording – and how data was translated back into English, where necessary. All primary data transcripts were anonymised, transcribed (e.g., transcribed from written text to computer/ digital copy) and translated into English (where necessary) as soon as feasible after collection. Primary data was cleaned, including checking for anonymity and missing data that may have occurred throughout processes associated with writing, transcribing (from audio to written transcript), translation (into and from English into the local language), storage, transmission (sharing from the primary data collectors to the IE team), or uploading/ digitisation of any data.

2.2.5. Quality Assurance Protocols

Our quality assurance protocols for data collection included the following:

All local partners ensured rigorous standards during fieldwork to ensure quality control. These standards included:

- All moderators, observers, quality control officers, and management staff recruited to work on this study signed non-disclosure and confidentiality agreements before they were engaged for fieldwork.
- Local partner staff ensured that all data collection was completed in line with standard research practice and complied with ethical standards of consent. All staff were transparent with respondents regarding the aim and objectives of the project and fully explained the process prior to commencing interviews.
- Throughout fieldwork, data collection control sheets to record the types and numbers of interviews conducted were completed by team leaders at the end of every interview day. Team leaders and research assistants went through every filled-out research tool cover sheet after the research assistants had checked the questionnaires and were satisfied that they were filled correctly.
- Research team supervisors accompanied researchers throughout fieldwork and offered feedback on facilitation/ moderation.
- Debrief meetings were held at the end of a fieldwork day.
- All local partners oversaw a systematic and transparent approach to data transcription.
- Daily calls and check-ins between the Fieldwork Manager and the local research teams allowed for the resolution of issues during fieldwork, as needed. Further, the Fieldwork Manager worked closely with the Tetra Tech team to manage the data collection process through updates as needed and weekly calls.

Our quality assurance protocols for this study required that each deliverable (including drafts) was reviewed prior to submission to the FCDO for consistency by the PI and Lead Author, Deputy Team Leader, Team Leader, Technical Director and Programme Director.

2.3. Research tools, training, translation, and piloting

All research tools were developed to respond to the research and evaluation questions. The Tetra Tech team held initial discussions with shortlisted IPs about the research questions and the focus of the research study.

Further consultations were held to ensure primary data collection tools reflect project-level priorities within the scope of the research questions and included project-level interests and learning priorities.

Any tools for use with children, such as FGDs with girls, were developed with reference to the literature on child-focused research methods, and particularly on safe and ethical research with girls and marginalised children.

All tools were translated into relevant local languages: Nepali in Nepal, Chichewa in Malawi, and Luganda in Uganda. Professional translators were used for all local language translations. Initial translations were reviewed by independent translators and our Southern Academic Partners (SAPs) for accuracy. Back translation was also undertaken to verify translation accuracy. All translations were further verified during training when local data collection staff were given the opportunity to discuss the nuance of specific words used and offer feedback on the translations. Piloting the tools at the end of training further tested the adequacy of these translations. Suggested revisions to the tools were made after piloting.

For all three countries, eight guides were developed as follows:

1. KII with national and district level officials and stakeholders;
2. SSI with GWDs;
3. SSI with parents/ caregivers;
4. SSI with teachers/ educators;
5. FGD with girls, including GWDs;
6. FGD with community leaders;
7. Participatory methods – photo and audio voice – with GWDs; and
8. Classroom observations.

Clear instructions were developed for facilitators responsible for implementing the data collection tools. These instructions were included in remote facilitator training and provided along with all data collection tools.

Training for interviewers, moderators, and supervisors was managed by our Fieldwork Manager. The study team offered feedback and provided input into all training materials and also provided training to the fieldwork manager. The training of the fieldwork manager was conducted by the Principal Investigator, Prof Nidhi Singal.

Prior to commencing in-country training with local field teams, the Fieldwork Manager was trained on the specificities of the project and all the research tools, with a special focus on the purpose and intent of each, by the study team. In each country, training took place in person over six days, with all supervisors and enumerators/ moderators. In Nepal, training took place in Kathmandu between 27 August and 3 September 2022. In Malawi, training took place in Lilongwe from 12 to 19 October 2022. In Uganda, training was held in Kampala between 31 October and 5 November 2022.

The training aimed to ensure that enumerators were able to efficiently and effectively conduct the research required; this included ensuring that enumerators are able to follow and strictly adhere to the programme's safeguarding practices and ethical protocols.

The first day of training in each country included a briefing from the local IPs to offer insight into their intervention. Additional topics covered included:

- Project overview, objectives, and purpose;
- Research methodology, sampling, and quality control;
- Safeguarding, research ethics, consent, and interview techniques;
- Data protection;
- Research tool review;
- Interview and moderation techniques;
- Classroom observation protocols; and
- Role playing and dummy interviews.

Specialised supervisor training took place at the end of enumerator training and included sample management, data verification, and team logistics and management.

Two days of piloting was carried out in each country after training was completed. The purpose of the pilot was to test for research tool sensitivities and comprehension and to allow research staff the opportunity to practice prior to commencing fieldwork. In all three countries, consent for research was obtained from each school in advance of the pilot.

In each country, team supervisors met with school administrators prior to piloting, often the day before. to discuss the purpose of the pilot and to make logistical arrangements. The team identified the girls needed for piloting and spoke with their caregivers to arrange times to complete these interviews. Consent and assent forms were completed for all caregivers and girls, as well as with the schoolteachers identified for interviews and classroom observation.

Across all three countries, the average time for interviews with teachers and caregivers was around one hour. More time was needed to complete all activities with the GWDs (interviews and FGD) to help create a space of ease and safety for the girl and allow for patience despite communication challenges. In all three countries, these FGDs took at least two hours, with some taking up to three hours, with breaks taken as needed to support girls who were struggling or needed to rest.

All three data collection partners submitted transcripts and observation narratives from the pilot for review by the Tetra Tech team.

Once the transcript review was complete, an additional day of training was offered in each country to review the pilot and offer feedback on any issues that arose with the transcripts. Post-pilot, research teams required some clarification on question intent and wording but in general, few issues were raised.

An overview of pilot research conducted, by country, can be found in Table 10 below.

Table 10: Pilot research conducted by country

Country	Nepal	Malawi	Uganda
Location	Shree Nepal Rastriya Secondary School, Banke	Yepa and Mbeta Learning Centres, Lilongwe	Creative Learning Centre (CLC) located inside the Kampala Baptist Church
Dates	6 to 7 September	20 to 21 October	7 to 8 November
SSI Girls	3	5	2
SSI Parents	5	6	4
SSI Teachers	2	4	3
FGD Girls	1	3	1
FGD Community Members	0	3	1
Photo/Audio Voice	1	4	1
Classroom Observation	3	5	1
Government KII	0	0	0

2.4. Analysis of Primary Data

To answer the research questions, a set of 337 primary qualitative transcripts received in English were gathered, cleaned, and analysed. This section describes how these data were analysed.

2.4.1. Coding processes and framework

Analysis of the primary qualitative data (transcripts) was conducted through the computer-assisted qualitative analysis software, NVivo. NVivo is a qualitative data analysis software which allows researchers to import, organise, code, analyse and collaborate on data. A team of six coders, including Dr Laraib Niaz and Romanshi Gupta from the study team, coordinated through Cloud Collaboration. This enabled multiple coders to simultaneously code all 337 transcripts gathered for the study.

All 337 transcripts were distributed among members of the coding team. To improve intra-coder reliability and efficiency, coders were allocated the entire set of one type of transcript (arising from a particular tool), for instance, one coder was responsible for coding the semi-structured interview transcripts with GWDs while another focused on semi-structured interview transcripts with parent/ caregivers.

Coding frameworks were provided to the team and codes were generated keeping this framework in mind also inductively, where new codes based on patterns or key themes emerging from the data would be developed by the coders.

Separate coding frameworks were developed for each of the tools. Once data was coded, the study team analysed it by reading all the data carefully and finding connections and themes within the data. The main messages from the data were derived and added to the report with quotes and figures added to provide evidence.

Example coding frameworks for semi-structured Interviews with GWDs and their parents/ caregivers are given below in Table 11.

Table 11: Qualitative coding frame for semi-structured interviews with GWDs

Main Code	Sub Code	Description
Characteristics of the GWD		Describes the respondent. Use this code if the sub codes do not apply
	1.1 Age	
	1.2 Siblings	Use this code for any information related to the siblings (number of siblings, their gender, age, level of education).
	1.3 Education	Describes the level of education of the girl. Use this code if the sub codes do not apply.
	1.3.1 Grade	This is the current grade the girl is enrolled in. Use this even if the girl is now enrolled in livelihood classes.
	1.3.2 Schooling	Use this code for previous schooling experience. For instance, if the girl was out of school before, number of years she has studied before.
	1.4 Involvement in household chores	Describes the girl's daily involvement in household chores and routine at home.
	1.5 Type of disability	Describes the type of disability the girl has (for instance physical, visual impairment, hearing impairment, cognitive).

Main Code	Sub Code	Description
	1.6 GWDs Relationships	Use this code for any description of GWDs' engagement and relationships with people. Use this code if the sub code does not apply.
	1.6.1 Relationship with parents	
	1.6.2 Relationship with siblings	
	1.6.3 Engagement with community members	
	1.6.4 Engagement with neighbours	
	1.6.5 Relationship with friends	
Barriers to inclusion of GWDs		This includes any challenges faced by the respondent in accessing education. Use this code if the sub code does not apply.
	2.1.1 Access to schools	This includes difficulties in access to schools (for instance transportation issues, lack of nearby resource schools).
	2.1.2 Challenges faced inside schools	Any challenges faced by the girls in the school such as lack of attention by teachers, getting teased by other students etc.
GWDs' perceptions on schooling		The respondent's perceptions and attitudes towards going to school. Use this code if the sub codes do not apply
	3.1 Reasons for dropping out of school	Use this code only for respondents who had previously dropped out of school.
	3.2 Motivation for going to school	
	3.3 Perception on the importance of education	Respondents' views on whether and why education is important for them.
Nature of Engagement with the intervention		Describes the respondent's participation in the intervention. Use this code if the sub codes do not apply Use this code if the sub codes do not apply
	4.1 Awareness of the Intervention	Use this if the respondent knows about the project interventions.
	4.2 In kind support for the girl	This includes resources in goods and services provided by the project such as sewing machines, bicycles, soap etc.

Main Code	Sub Code	Description
	4.3 Learning support for the girl	Use this code for any support provided to the respondent to enhance their learning such as the provision of assistive devices, extra attention given etc.
	4.4 Monetary support for the girl	Use this code for any financial support provided by the project such as yearly or monthly cash payments.
	4.5 Support during COVID-19	Use this code for any support provided to the respondents by the project during COVID-19.
	4.6 Engagement and training with parents	This includes any resources provided to the parents and/or any training and awareness sessions given to them.
	4.7 Additional support provided to family	Use this code for in kind and monetary support provided to the family. Note this is different from 4.2 and 4.4 above.
	4.8 Project-specific support	Use this code for any interventions/ support specific to individual projects.
	4.8.1 Big Sister Nepal	Use this for any information on how the Big Sisters supported/ interacted with the GWDs.
	4.8.2 Bridge class Nepal	Use this code if the respondent had access to/ attended the bridge class.
Perceived effects of the intervention		This includes any effects of the intervention that the respondent mentions. Use this code if the sub codes do not apply
	5.1.1 Learning Outcomes	Use this code if the sub codes do not apply
	5.1.1 Changes in learning outcomes	Use this code for any perceptions around changes in the learning outcomes of the respondent.
	5.1.2 Confidence in learning environment	Use this code for any changes in the confidence level and participation of the girl in the school.
	5.1.3 Vocational skills acquired	Use this code for acquisition of any vocational skills following the intervention such as sewing, farming etc.
	5.2 Changes in parents' attitudes	Use this code for any changes in the parent's attitudes and behaviour towards the respondent following the training.
	5.3 Changes in community members' perceptions	Use this code for any changes that the respondent has witnessed around community members' perceptions and attitudes towards the GWD following the intervention.
	5.3 Usefulness of COVID-19 resources	Use this code for how the respondents used the COVID-19 resources provided to them.
	5.4 Usefulness of project resources	Use this code for respondents' use of any other resources provided by the intervention.

Main Code	Sub Code	Description
	5.5 Effect on income	Use this code for any effect on income generated due to participation in the livelihood classes (through selling items that the respondent learnt to make in the classes).
	5.6 Project-specific effects	Use this code for effects of any project-specific support. Use this code if the sub codes do not apply.
	5.6.1 Role of Big Sister Nepal	Use this code for the respondent's views on how helpful/ or not the big Sister has been.
External Support		Any support received by the girl other than from the project. Use this code if the sub codes do not apply
	6.1 Support from other organisations	Any other organisations that have provided support to the Girl such as community organisations, other NGOs.
	6.2 Support from government	Any support from the local Ward, Municipality or the district and national level government offices.
	6.3 Support of family members	Support of extended family members.
Aspirations for the future		This includes any goals for after the project ends such as further education, working, marriage etc.
Type of additional support needed		Any additional support needed such as more financial help, attention by teachers, accessible spaces etc.

Table 12: Qualitative coding frame for semi structured interviews with parents/ caregivers

Main Code	Sub Code	Description
1. Characteristics of the GWD		Describes the girl with a disability. Use this code if the sub codes do not apply
	1.1 Age	
	1.2 Siblings	Use this code for any information related to the siblings (number of siblings, their gender, age, level of education).
	1.3 Education	Describes the level of education of the girl. Use this code if the sub codes do not apply.
	1.3.1 Grade	This is the current grade the girl is enrolled in. Use this even if the girl is now enrolled in livelihood classes.
	1.3.2 Schooling	Use this code for previous schooling experience. For instance, if the girl was out of school before, number of years she has studied before.
	1.4 Girl's involvement in household chores	Describes the girl's daily involvement in household chores and routine at home.
	1.5 Type of disability	Describes the type of disability the girl has (for instance physical, visual impairment, hearing impairment, cognitive).

Main Code	Sub Code	Description
	1.6 GWDs Relationships	Use this code for any description of GWDs engagement and relationships with people. Use this code if the sub code does not apply.
	1.6.1 Relationship with parents	
	1.6.2 Relationship with siblings	
	1.6.3 Engagement with community members	
	1.6.4 Engagement with neighbours	
	1.6.5 Relationship with friends	
1 Barriers to inclusion of GWDs		This includes any challenges faced by the parents or the child in accessing education. Use this code if the sub code does not apply.
	2.1.1 Access to schools	This includes difficulties in access to schools (for instance transportation issues, lack of nearby resource schools).
	2.1.2 Challenges faced inside schools	Any challenges faced by the girls in the school such as lack of attention by teachers, getting teased by other students etc.
2 Nature of Engagement with the Intervention		Describes the respondent and their child's participation in the intervention. Use this code if the sub codes do not apply
	3.1 Awareness of the project	Use this if the respondent knows about the project interventions.
	3.2 Provision of resources for the family	This includes resources in goods and services provided by the project such as sewing machine, bicycle, soap etc.
	3.3 Project Meetings with parents	This includes any meetings hosted by the project around the support for GWDs that the parents attended.
	3.4 Monetary support for the family	Use this code for any financial support provided by the project such as yearly or monthly cash payments.
	3.5 Project team's visits	Use this code for any visits the project team conducted at the respondents' houses.
	3.6 Support during Covid	Use this code for any support provided to the respondents by the project during COVID-19. Use this code if the sub codes do not apply.
	3.6.1 Provision of resources	This includes resources such as masks, sanitisers, soaps etc. given by the project.
	3.6.2 Vaccine	Use this for any help in accessing the vaccine.
	3.7 Training for parents	This includes any sensitization and awareness training provided by the project to the parents.
	3.8 Project-specific support	
	3.8.1 Big Sister Nepal	Use this for any information on the Big Sister's interactions with the girl.
4. Perceived Effects of the intervention		This includes any effects of the intervention that the respondent mentions. Use this code if

Main Code	Sub Code	Description
		the sub codes do not apply
	4.1 Behavioural changes in Girl	This includes any changes in the behaviour/ attitudes of the GWD that the parent has witnessed. Use this code if the sub code does not apply.
	4.1.1 Changes in confidence	Use this code for any changes in the confidence level of the girl in school or outside.
	4.1.2 Changes in interactions with friends	Use this code for any changes the parents have witnessed in the girl's interactions with their friends.
	4.1.3 Changes in engagement with family members	Use this code for any changes the parents have witnessed in the girl's interactions with their family members (parents, siblings, grandparents).
	4.1.4 Changes in motivation for going to school	Use this code for any changes witnessed in the motivation for going to school in the GWD.
	4.1.5 Change in engagement with community	Use this code for any changes the parents have witnessed in the community members' engagement with the GWD.
	4.2 Changes in learning	Use this code for any perceptions around changes in the learning outcomes of the GWD.
	4.3 Effects of training for parents	Use this code for any changes in the respondent's attitudes and behaviour following the training.
	4.4 Changes in community members' perceptions	Use this code for any changes that the parents have witnessed around community members perceptions and attitudes to the GWD following the intervention.
	4.5 Usefulness of COVID-19 resources	Use this code for any information on how the respondents used the COVID-19 resources provided to them.
	4.6 Usefulness of Government funds	Use this code for how the respondents used the government funds provided to them.
	4.7 Usefulness of intervention resources	Use this code for respondents use of any other resources provided by the intervention.
	4.8 Project-specific effects	Use this code for effects of any project specific support. Use this code if the sub codes do not apply.
	4.8.1 Role of Big Sister Nepal	Use this code for the respondent's views on how helpful/ or not the big Sister has been.
3 Aspirations for the Girls Future		Parents' hopes for the child's future. Use this code if the sub codes do not apply
	3.1 Further education	Parent wants the girl to pursue further education. Use this code also for any apprehensions around getting further education.
	3.2 Marriage	Parent wants the girl to get married in the future. Use this code also for any worries around the prospect of marriage.
	3.3 Job	Parent wants the girl to be working in the future. Use this code also for any apprehensions around the prospect of working.

Main Code	Sub Code	Description
4 **External Support**		Any support received by the girl other than from the project. Use this code if the sub codes do not apply
	4.1 Support from other organisations	Any other organisations that have provided support to the family such as community organisations, other NGOs.
	4.2 Support from the government	Any support from the local Ward, Municipality or the district and national level government offices.
	4.3 Support from the community	Any support provided by the community members.
5 Additional types of support needed		Any additional support needed such as more financial help, attention by teachers, accessible spaces etc.

2.4.2. Triangulation

Where one GWD participated in multiple methods (for instance in photovoice, classroom observations and interviews, and focus group discussions), multiple transcripts relating to one respondent were generated. Transcripts were linked via unique identifiers (IDs) through a classification sheet that was created through NVivo. For instance, a case study GWD who participated in photovoice, interview, FGD and classroom observation would have four transcripts coded to her unique ID. These transcripts were then compared to see what the case study girl is saying in the FGD and interviews as well as linking to her observed behaviour in the classroom.

These subsets of transcripts were then analysed to compare for instance what a teacher says about their teaching practices against what practices they were observed implementing in the classroom.

GWDs were also linked to their parents/ caregivers and teachers/ educators to allow for triangulation between the various participants. The study team also focused on triangulating responses to similar questions by different participants. For instance, the coding frameworks for GWDs, parents/ caregivers, teachers/educators as well as community members all have codes for challenges faced by GWDs and these responses were collated and analysed.

2.5. Limitations and mitigation

1) Limited availability of project data/ quality concerns regarding quantitative, longitudinal data impacted the type of analyses possible to assess changes over time. We used the quantitative data to the extent possible and communicated the rationale for including/ excluding certain datasets and the final methods to the FCDO as required, for their feedback. This study also included a substantial amount of primary qualitative data collection and analysis allowing us to fully respond to the research questions.
2) While we endeavoured to include girls with all types of disabilities in our qualitative sample, it was not always possible to do so due to limited sample sizes. We liaised with IPs to identify the beneficiaries with different types of disabilities and incorporated this into our sampling approach as a criterion.
3) Key stakeholders were not easily accessible or refuse to participate in the data collection in some cases. We liaised with IPs and our Fieldwork Manager to identify a larger sample of stakeholders in case of refusals with replacements for refusals.
4) Delays in obtaining research permissions/ ethical approvals. We allowed sufficient time in the work plan to obtain all relevant permissions and scheduled this task as early as possible prior to the start of fieldwork.
5) Disruptions in fieldwork time (September to November) due to public holidays (for instance Diwali, Dussehra in October in Nepal). We planned fieldwork in the three countries according to their respective

school calendars from September to December (for instance conducting fieldwork in Nepal in September due to the holiday season in October).

6) For RQ1, it was only possible to interview 27 out of the 41 IPs due to non-response (7 projects), refusal to participate (2 projects) and scheduling challenges (5 projects) in the timeframe of completing interviews. The IPs who did not respond were primarily those whose projects ended by the time we contacted them for interviews. One IP declined to participate as their project did not include GWDs, and another IP declined to participate as the project had ended and most staff had moved onto other projects/ organisations.
7) For RQ2, the purposive selection strategy for projects means that selected projects are not intended to be representative of the wider GEC II portfolio – rather the aim is to understand what works and what can be learnt. The selection focuses on projects that included specific interventions related to girls with disabilities.
8) While we endeavoured to include girls with all types of disabilities in our qualitative sample, it was not possible to do so in all three countries. In Nepal, the sample of GWDs in each school/ learning centre was too small (between 1 to 7) to include proportional representation of disability. In Uganda and Malawi, some schools with a big sample of GWDs did not have any girls with visual impairments.
9) The selection of tools (WG-SS or CFM set) for collecting data on disability and the criteria (the levels of difficulty carrying out an action) for determining a girl's disability, as well as the respondent (girl self-reporting or PCG reporting on the girl's disability)[5] varied among projects. This prompted decisions to maximise sample size and ensure consistency in the analysis of quantitative data. To this end, it was decided to use girls' self-reported data on disabilities and (wherever possible)[6] fix the status at baseline (to ensure the same girls are compared over time). This means that relevant project data on disability are excluded (for example, where projects collect data from PCGs but not directly from girls), thus allowing for the profiling of disabled girls from 23 GEC-T and 10 LNGB projects. Additionally, fixing status at baseline can introduce bias when the status of disability changed significantly from baseline to midline (of the 2,564 girls reported disabled at baseline 431 reported disabled at midline in GEC-T. In LNGB, the figure is 22 out of 226 disabled girls in baseline).
10) In GEC-T, we evaluate learning progress over and above the non-treated group, allowing us to attribute gains in girls' outcomes relative to the intervention. However, the majority of the LNGB projects did not include in their design non-treated groups (girls who did not receive intervention). Therefore, for the LNGB window, the focus of the analysis is on the treatment girls – how they learn over time and comparing them to non-disabled girls. While we acknowledge that girls with disabilities may have different challenges compared to girls with no disabilities, the comparison yields useful information in that finding no differences between them would constitute a positive outcome. For consistency, we also provide this comparison for the GEC-T window (see full quantitative background paper in Annex C).
11) There are differences across projects in terms of the tools (WG-SS or CFM set) used to collect data on disability and the level of difficulty carrying out an action considered by projects, which prompted decisions to ensure consistency in this study. This was accomplished by using girl-reported data on disabilities, following WG guidelines concerning the level of functional difficulty considered as a disability ("a lot of difficulty" or "cannot carry out the action at all"; and "daily" for depression and anxiety). This means that relevant data on disability are excluded (for example, where projects collect data from PCGs but not from girls).
12) GEC-T projects primarily implemented the WG-SS set, while most LNGB projects used the CFM scale instead. We use the tools that are frequently used by each window (WG-SS for GEC-T and CFM for LNGB). This resulted in a higher prevalence of disability being observed in the LNGB window due to the increased number of dimensions captured in the CFM set compared to the WG-SS set.
13) 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). The profiling of girls maximises the data available; as such, each variable used for profiling yields a different sample size. No

[5] Within the GEC-T and LNGB, there was a disparity between the proportion of girls who self-reported disability and the proportion of PCGs who reported on girls' disability, with 73% and 80% of total samples respectively self-reporting disability in comparison to 48% and 51% of PCGs respectively.
[6] Cross-sectional analysis of learning data requires disability status to be determined at each data collection point.

attempt has been made to keep the sample size consistent as that would lead to a race to the bottom and a sample not representative of the portfolio of projects. Similarly, for examining changes over time, we are not able to include any covariates in the analysis as our sample size would drop significantly.

14) The LNGB data include a specific set of questions related to the barriers to schooling faced by girls. This data is not collected for the GEC-T which mostly focuses on in-school girls. Due to a lack of barrier data in the follow-up round, we did not investigate change over time in barriers to schooling for LNGB girls.
15) Difference-in-difference strategies are employed to assess learning changes on: (1) the cross-sectional (i.e., sample of baseline and midline/endline girls); and (2) panel sample (i.e., sample of girls followed over time).[7] Due to the lack of common girls' unique IDs between the rounds across most LNGB projects (at the time of writing this report), only GEC-T projects were able to be included in the panel analysis.[8] For LNGB projects, we look at change over time in the cross-sectional sample, which included 493 disabled girls. Results from this analysis should be read with caution as they are based on a much smaller sample size for disabled girls compared to GEC-T.
16) The WG tools condense the complexities of conceptualising disability into an easy to implement tool. However, changes observed over time in reporting present a challenge for analysis (e.g., a girl may report having a disability at baseline, but then not report it at the midline or endline, or vice versa). To avoid this, whenever possible, we use baseline data to assign disability status. The cross-sectional learning analysis is an exception, meaning that a girl can have a different disability status at baseline and midline/endline. Results from this analysis should be interpreted with caution as it is not possible to measure the potential bias of such transitions between states.
17) Data is gathered for the GEC-T external evaluations from a sample of 'treated' girls (girl beneficiaries) and a sample of girls with comparable characteristics who do not receive any intervention (non-treated). As such, the effects of GEC II interventions relative to the non-treated group can be, and is, estimated. The majority of LNGB window projects do not include non-treated groups.[9] Therefore, it is not possible to attribute gains in girls' outcomes relative to a non-treated group for the LNGB. For the LNGB, the alternative is to focus the analysis on treatment girls – first how they learn over time, and also comparing this to non-disabled girls. While we acknowledge that girls with disabilities may have different challenges compared to girls with no disabilities, the comparison yields useful information in that finding no differences between them would constitute a positive outcome.
18) Intermediate outcomes are assessed through individual questions asked to girls or PCGs which serve as proxy indicators for attendance, economic empowerment, life skills, quality of teaching, gender-based violence, and parental attitudes. These questions are adapted from various scales and provide an indication of understanding change in these outcome areas, even though they are not measured reliably. Changes in the intermediate outcomes are measured as shifts in answers based on Likert scales (ranging from Totally Agree to Totally Disagree).
19) There may be evidence of attrition bias, which affects analysis involving change over time in learning outcomes. In the treatment group, disabled recontacted girls scored higher in baseline numeracy than those lost to attrition. In the non-treated group, however, disabled recontacted girls scored worse in baseline literacy than those lost to attrition. This implies that those recontacted in treatment and non-treated group have different baseline literacy and numeracy than those lost to attrition. Learning gains of treatment group are analysed over non-treated group with lower literacy level than the overall baseline non-treated sample. Vice versa, recontacted treatment group have higher level of numeracy than the overall baseline treatment sample.
20) Given the timelines of the study, there was limited time for coding which meant that coders could not cross-code across different tools to compare the coding outcomes from different coders. For this reason, each coder was assigned a specific tool and coding framework with which to work.

    Other challenges in the coding process included:

[7] Projects with only baseline data available are excluded from learning analysis.
[8] While the LNGB projects could in theory be included in panel analysis, the follow-up sample constitutes 493 disabled girls with any learning data, of which only 231 were recontacted (see Table 10). Given such a small sample size and coverage of only three projects at follow-up, we do not conduct a panel analysis for LNGB.
[9] Because of the ethical implications of targeting highly marginalised girls, the majority of programmes do not have non-treated groups; instead, they offer the intervention to all girls identified as meeting the programme criteria. One project, Pin Nepal, did employ a non-treated group with girls in that group intended to receive the intervention in enrolling cohorts.

Coders were coding qualitative data that they had not collected leading to potential bias. To redress this:

a) The study 4 team undertook a full day of training that was led by the Senior Research Lead, Dr Laraib Niaz, who spoke at length about the study, describing the research design, tools, coding, and expectations from the analysis.
b) Team check-ins with the coders, where the Senior Research Lead met with the coding team to discuss challenges, emerging codes, and suggested priority areas for analysis.
c) The Qualitative Analyst, Romanshi Gupta, created a coding log through which coders reported on the coding process which was regularly reviewed by the Senior Research Lead.
d) Note that Ebola and Ebola-related school closures significantly impacted our sample. First and foremost, our timeframe for completing the fieldwork was cut short due to school closures mandated by the Government of Uganda to control the spread of Ebola. These closures meant that schools had to move exam schedules forward, limiting our ability to work inside of the selected schools.

2.6. Dissemination

The findings, conclusions and recommendations of the study are being shared through the study report as well as a knowledge brief that will summarise the main points from the report but directed towards a wider audience (students, teachers, IPs, OPDs, government stakeholders, donors). Additionally, the study findings will also be disseminated through presentations at academic and other conferences and seminars and through the development of academic articles and blog posts.

Annex C: Quantitative Report

Evaluation Study 4 – Educating Girls with Disabilities

Quantitative Findings

Acronyms

ALP	Accelerated Learning Programme
BL	Baseline
CBE	Community Based Education
CFM	Child Functioning Module
EGMA	Early Grade Mathematics Assessment
EGRA	Early Grade Reading Assessment
GEC	Girls' Education Challenge
GEC-T	Girls' Education Challenge – Transition
GS	Girls Survey
GWD	Girls with Disabilities
HHS	Household Survey
IO	Intermediate Outcomes
IS	In-School
ISG	In-School Girls
JSS	Junior Secondary Schools
LNGB	Leave No Girl Behind
ML	Midline
OLS	Ordinary Least Squares
OOS	Out-of-School
OOSG	Out-of-School Girls
PCG	Primary Caregiver
SeGMA	Secondary Grade Mathematics Assessment
SeGRA	Secondary Grade Reading Assessment
WG	Washington Group
WG-SS	Washington Group Short Set on Functioning

Project Acronyms

GEC-T 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
Cheshire Services Uganda	CSU	Uganda
ChildHope UK	ChildHope	Ethiopia
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
Plan International	Plan	Sierra Leone
Promoting Equality in African Schools	PEAS	Uganda
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 University Service of Canada	WUSC	Kenya
World Vision	WV	Zimbabwe

LNGB Project name	Acronyms used in report	Project location(s)
ACTED	ACTED	Pakistan
ActionAid International Kenya	ActionAid	Kenya
AKF	AKF	Afghanistan
CARE International UK	CARE	Somalia
International Rescue Committee [Pakistan]	IRC Pakistan	Pakistan
International Rescue Committee UK [Sierra Leone]	IRC Sierra Leone	Sierra Leone
Link Community Development International	Link	Malawi
People in Need (PIN), Ethiopia	Pin Ethiopia	Ethiopia
People in Need (PIN), Nepal	Pin Nepal	Nepal
Plan International UK	Plan	Zimbabwe
Population Council	Population Council	Ethiopia
Street Child	SC Nepal	Nepal
VSO	VSO	Nepal
World Education, Inc.	WEI	Ghana

1. Summary

This paper provides an overview of girls with disabilities supported by GEC II project interventions and examines changes in their learning and life skill outcomes over time, based on 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 follow-up (midline or endline).

Phase II of the GEC programme, which is running from 2017 to 2025, has allocated £500 million to 41 projects across two windows: the GEC Transitions (GEC-T) window and the Leave No Girl Behind (LNGB) window. The GEC-T window is providing support to 27 successful GEC Phase I projects in 15 countries for 1.3 million girls to transition to the next stage of their education, while the LNGB window is helping 14 projects in 10 countries to support up to 500,000 highly marginalised, adolescent girls who have never attended or have already dropped out of school, to enrol back into school or pursue an alternative pathway.

This paper contributes to cross-sectional and longitudinal research evidence on disability, using the Washington Group questions, a tool designed to identify people with a disability. It seeks to make visible the unique and often marginalised experiences of girls and young women with disabilities who are supported by the GEC II projects. The aim of this study is twofold: the first one is *profiling beneficiary girls* in terms of their functioning disabilities. The second one is *examining changes in girls' learning and intermediate outcomes*, conditioned by disability status.

As this study is based on projects' external evaluation data, we are limited by the data that was collected, which impacts the analytical methods we are able to employ to this dataset. We set out these limitations and key methodological decisions at the start of this paper, such as our focus on girl-reported disability. Overall, out of the 27 GEC-T projects, we were able to utilise data from 23 projects, involving 34,030 beneficiary girls; of the 14 LNGB projects, we used data from 10 projects, involving 13,296 beneficiary girls.1

The number, and share, of girls reporting to have a disability varies substantially across projects and GEC II windows. As expected, two GEC-T projects that focused on girls with disabilities have higher shares. The share of girls reporting depression and anxiety at any time is high, with more than half having experienced depression and 1 in 25 having this daily.

The data suggests that girls with disability are also more marginalised than non-disabled girls in terms of having other characteristics of marginalisation - being poor or orphaned. In LNGB projects, disabled girls are also more likely to be married and mothers than non-disabled girls. Though caregivers of LNGB girls with and without disability have identified financial reasons as the main reasons for being out of school, this reason becomes more prominent for disabled girls. Caregivers of disabled girls are more likely to report reasons related to 'potential' functional difficulty than caregivers of non-disabled girls, such as "girl having a health condition that prevents her from going to school".

Girls with disabilities have lower baseline learning levels compared to non-disabled girls – but encouragingly, in GEC-T, disabled girls have shown similar levels of improvement as non-disabled girls. Disabled girls who received treatment have also shown improvements beyond disabled girls who did not receive treatment. For LNGB, despite starting at a lower baseline learning level, disabled girls have shown improvement above and beyond non-disabled girls in both literacy and numeracy (though results are only statistically significant in numeracy).

In terms of intermediate outcomes (including attendance, economic empowerment, life skills, teaching quality, school related gender-based violence, and parental attitudes), GEC-T girls with disability felt that aspects of economic empowerment (e.g., reduction in difficulty affording schooling), life skills (e.g., reduction in feeling nervous doing Maths in front of others or not focused on goals), and teaching quality (e.g., less GWDs felt that their teacher treated boys and girls differently) had improved over the course of the intervention, including their perception of increased safety in travelling to/ from school. However, negative changes were observed in terms of caregiver-reported school attendance and primary caregivers felt there was less safety for girls travelling to schools. No significant changes in other intermediate outcomes were observed. Positive changes in intermediate outcomes, found in Independent Evaluation of the Girls' Education Challenge Phase II – Evaluation Study 3: Aggregate impact of GEC-T projects between baseline and midline (Study 3), are primarily driven by non-disabled girls.

[1] For profiling of disabled girls. The sample sizes reduce when investigating barriers and changes in learning over time. Caveats of this are discussed in the Methodology section.

2. Methodology

Box 1: Key methodological decisions made in the analysis

Summary of key methodology decisions

1. We use the tools that are frequently used by each window (WG-SS for GEC-T and CFM for LNGB). We define a functional difficulty level as recommended by the Washington Group, which categorises disability as either "a lot of difficulty" or "cannot carry out the action at all".
2. We focus on girl-reported disability in this study to capture girls' experiences of functional difficulty in their environments; therefore, projects without girl-reported disability information are excluded. Using girl-reported disability also maximises the sample size which otherwise would be too small to conduct reliable analysis
3. We use baseline data to define disability status, while acknowledging that the prevalence of functional difficulties can fluctuate over time.
4. The profiling of beneficiary girls maximises the data available; as such, each variable used for profiling yields a different sample size.
5. Analysis is conducted primarily on girls who received treatment and focuses on two areas: (1) profiling beneficiary girls in terms of their functioning disabilities; and (2) examining changes in girls' learning and intermediate outcomes.
6. Difference-in-difference strategies are employed to assess learning changes on: (1) the cross-sectional (i.e., sample of baseline and midline/endline girls); and (2) panel sample (i.e., sample of girls followed over time). However, due to the unavailability of girls' unique IDs across rounds for most LNGB projects (at the time of writing this report), panel analysis was only conducted for GEC-T projects.
7. In GEC-T, difference-in-difference analysis was conducted to compare learning changes against disabled girls who did not receive intervention (non-treated group). For the majority of LNGB projects, however, non-treated group did not exist. In these cases, learning changes between disabled and non-disabled treatment girls were compared instead.

2.1. Measurement of disability

In the paper, disability is measured using the Washington Group method, which condenses the complexities of conceptualising disability into a tool that produces cross-nationally comparable population-based measures of functioning disability. Because the tool is easy to implement in a variety of cultures, with varying levels of economic resources, it has been widely used internationally in surveys and national censuses.

The basic WG model (the WG Short Set on Functioning (WG-SS)) is based on six questions that represent six actions most often found to limit an individual and result in participation restrictions. While arguably limited, the results from this set of questions guarantee that (a) the majority, not all, of the people with limitation in basic actions will be represented, (b) the most commonly occurring limitations in basic actions will be represented, and (c) people with similar limitations across countries/ contexts will be captured2. In addition to the WG-SS, the WG has developed five other question sets. The other one used in the paper is the WG Child Functioning Module (CFM), specifically the version for children aged 5-17 years old. The CFM is designed to be administered to primary caregivers (PCGs) and consists of the six core WG-SS questions plus 18 new questions on seven developmental disabilities domains including learning, concentrating, accepting change, controlling behaviour, making friends, anxiety, and depression.3

Both the GEC-T and LNGB windows utilised the WG-SS and the CFM set, asking questions about the girls' disabilities, both directly to the girls themselves (through girls' surveys) and to their PCGs (through household questionnaires).4 At baseline, the GEC-T window primarily administered the WG-SS to girls. In addition, a sample of GEC-T projects administered CFM for primary caregivers, while others used the WG-SS. In contrast, the

2 Read more in https://www.washingtongroup-disability.com/question-sets/wg-short-set-on-functioning-wg-ss/ and see Questionnaire in Appendix 1.
3 Read more in https://www.washingtongroup-disability.com/question-sets/wg-short-set-on-functioning-wg-ss/ and see Questionnaire in Appendix 1.
4 While information is reported by both the girls (through girls' surveys) and their primary care givers (through household questionnaires), all of the information reported is about the girls themselves. In this report we use the terms girl-reported to denote the girls' own responses and PCG-reported to denote information reported by primary caregivers

LNGB window administered the CFM to girls and their PCGs, with the exception of one project that administered the WG-SS to PCGs instead.5

As there are differences across projects in terms of design choices – e.g., who was administered the household questionnaire – and tools used, the sample sizes for analysis differ depending on the availability of the relevant variables. While ideally, we would maintain a balance, the differences across projects lead to a race to the bottom in which relevant data get excluded.

The sample varies with the choice of tool. Out of the 27 GEC-T projects, 19 projects6 and CAMFED Tanzania and Zimbabwe have both girl-reported data and PCG-reported data (collected using either WG-SS or CFM). Four projects (Avanti, STC DRC, Varkey, and Viva) and CAMFED Zambia do not have girl-reported disability data and four projects (Childhope, MC Nigeria, and PEAS) do not have data on disability as reported by the primary caregiver (as measured by the WG-SS or CFM). One project (Varkey) did not have disability data at all.

In LNGB, the availability of disability information, both girl-reported and PCG-reported, varies across projects. At the time of writing the report, data was available for 30 LNGB cohorts (within 12 projects). Thirteen LNGB cohorts from six projects simultaneously collected girl- and PCG-reported disability data. In addition, 10 cohorts from 7 projects (ActionAid C3, IRC Pakistan Distant, IRC SL C1, Pin Ethiopia C3, Pin Nepal C1 & C3, Plan GS, Street Child (C1, C2 & C3) girl-reported disability while five cohorts from three projects (Plan HH, VSO C1 and WEI C1 Formal, C1 Nonformal and C2 Nonformal) have primary caregiver data. Pin Nepal C3 only has disability data reported by the primary caregiver using the WG-SS. One cohort (Acted C2 L&N) does not have baseline data. Appendix 2 describes this in greater detail.

The lack of data is explained by choices made by projects at the design stage. **In the GEC-T window, the sample is considerably reduced when using PCG-reported disability, since many projects only administered the household questionnaire for a subset of the girls. In the LNGB window, the sample is reduced because projects administered the WG questions to either the girl or primary caregiver.**

Overall, in GEC-T, 73% of the total sample of girls receiving interventions has girl-reported disability data while only 48% has PCG-reported disability data. In the LNGB, 80% of the overall sample receiving interventions has girl-reported disability data, while only 51% has PCG-reported disability data.

Table 1: Number of baseline GEC-T and LNGB observations with disability data

	GEC-T		LNGB	
	Number of observations	% of total observations	Number of observations	% of total observations
Total treatment girls	46,704		16,719	
Has girl-reported disability data	34,030	73%	13,296	80%
Has PCG-reported disability data	22,233	48%	8,545	51%

In this paper, we use girl-reported measures of disability for the analysis. We also used the sets that are frequently used by each window (WG-SS for GEC-T and CFM for LNGB). Maximising the sample size is one important reason for focusing the analysis on girl-reported disability. However, the choice between girl- and PCG-reported data goes beyond this. While the Washington Group questions are asked to parents or caregivers for young children, their views might not always accurately represent the experience of the child as children and parents have different perceptions of reality. Previous research conducted by Outhred et al. (2020) has shown that girls receiving GEC-T interventions often reported being disabled at a greater proportion than their caregivers, with little agreement when the same set of questions was asked to both girls and caregivers. We found similar discrepancies in the data collected in the GEC-T and the LNGB windows (i.e., higher prevalence of disability when the tool is administered to the girl) when examining the answers of a subsample that has both girl- and PCG-reported disability (see Appendix 3).

Moreover, research has shown that the prevalence of functional difficulties among adults fluctuates over time, with some transitioning in and out of severe disability over time (Mitra, 2017, p. 73). Reasons for these transitions can be due to aging, but also to changes in reporting behaviours. These fluctuations are also observed in our data (when we

5 For further information about the administration of these scales across the portfolio of projects, please refer to Appendix 2.
6 AFK, BRAC, CAMFED (ex-BRAC), Care, CSU, DLA, EDT, HAP, ICL, LCD, Link, MC Nepal, Opportunity, Plan, Relief, STC Moz, VSO, World Vision and WUSC.

compare the same girls over time). For instance, among a panel sample of GEC-T girls with disability and learning data at both baseline and midline, 2,564 (of which 1,493 are treatment girls) were categorised as being disabled at baseline and only 431 (246 in treatment) of these were categorised as disabled at midline. Similarly, among a sample of LNGB girls with follow-up data, of 226 categorised as being disabled at baseline, only 22 girls were categorised as disabled at midline. These differences could be explained by changes in interpretation of the disability questions, as well as changes in girls' lived experience of functioning difficulties (e.g., where assistive devices have been provided). Due to these differences and fluctuations, and in order to facilitate comparison, **this paper fixes the disability status at baseline for all the analysis with the exception of cross-sectional analysis.**

As the differences in reporting are large, we ideally would conduct robustness checks to see whether our results are driven by our choice of focusing on girls-reported disability and fixing it at baseline. However, an imminent challenge with this is the resulting sample sizes: 100-162 observations when disability reported by PCG coincides with the girl-report of disability and 151-386 observations when girls are categorised as disabled at both at baseline and midline (regardless of whether the type or number of disabilities match). Despite this challenge, we reanalysed the data and the results are presented in Appendix 4. However, the results need to be interpreted with caution.

The WG-SS and the CFM define the level of difficulty carrying out an action on the following continuum: no difficulty; some difficulty; a lot of difficulty; or cannot do at all. Certain domains in the CFM, such as anxiety and depression in this paper, have different response options. The WG recommends the adoption of the strict cut-off, where **a functional difficulty level is assessed as either "a lot of difficulty" or "cannot carry out the action at all", for estimating disability prevalence rates. This same cut-off was used in the majority of external evaluations of GEC-T (22 projects) and LNGB projects (25 cohorts) and has been used for this analysis.** 7 Three GEC-T projects and four LNGB cohorts (within CARE project) reported both cut-offs separately.

Table 2: WG cut-offs

Cut-offs	GEC-T	LNGB
Applied strict cut-off only	19 projects	21 cohorts (within 11 projects)
Applied both loose cut-off *and* strict cut-off	3 projects	4 cohorts (within 1 project)
Unknown	4 projects	3 cohorts (within 3 projects)

Note: This table includes 26 GEC-T projects; it excludes Varkey which does not have disability data. For the LNGB it includes 28 cohorts; it excludes Acted C2 L&N which does not have baseline data and Pin Nepal C3 which only collected WG data from primary caregiver. Loose cut-off includes girls reporting "some difficulty" as disabled.

Robustness check: comparison against external evaluation

As a robustness check, we looked for differences between the data, definitions of disability, and the disability status stated in each external evaluation's baseline report. Overall, in all 17 GEC-T projects with girl-reported disability data, the baseline report presented prevalence based on girl-reported data and the estimates in the report mostly correspond to our data. Discrepancies were detected in Childhope and CSU projects.8 For LNGB, 14 cohorts presented a disability status by using CFM data collected from girls' surveys in the baseline report. All corresponds to our data with the exceptions of two projects: Acted C1 ALP and Pin Ethiopia C1. This shows that there are not major discrepancies between the data used in this report and the data used in the evaluations. The full tables can be seen in Appendix 5.

2.2. Method of analysis

This paper has two main focal areas: the first one is *profiling beneficiary girls* in terms of their functioning disabilities. Based on a cross-sectional examination of all girls in the GEC-T and LNGB portfolios, the profiling first shows the extent and types of disabilities faced by beneficiary girls and then describes the main characteristics of girls with disabilities in terms of age, education (enrolment, repetition, overage for grade), marital status, pregnancy and childbearing, orphanhood, poverty, and main barriers to education (see Section 3). To put the profiling in context, disabled girls' characteristics are compared to the characteristics of all girls in the portfolio.

The second focal area is examining changes in girls learning and intermediate outcomes, conditioned by disability status (see Section 4). This follows a quantitative analysis based on descriptive statistics of the relevant data,

7 For anxiety and depression, when a girl reports experiencing these conditions "daily", we categorise the girl as disabled.
8 For Childhope, the discrepancy resulted from the use of the loose cut-off definition of disability, including girls who stated, "Yes, Some Difficulty".

showing how it compares across groups and across time. We complement this with a statistical comparison of the averages, through t-tests of group differences between disabled girls in treatment and non-treated group, and between girls with disabilities and those without.

Where the data permits, we also look at a 'difference-in-difference' regression analysis, where we used ordinary least squares analysis (OLS) of the changes in learning between data collection periods for disabled girls who received the intervention, compared to disabled girls who did not receive intervention. The difference-in-difference estimates are used to estimate the (non-casual) impact of the programme on learning outcomes of girls. This is estimated by firstly calculating the difference between treatment group's learning score in midline (or endline) and baseline. This is the first difference. Due to the lack of a non-treated group for LNGB, the result of the first difference is what we report for this group. For GEC-T, we are able to calculate the difference between the treatment and the non-treated group. The 'difference-in-difference' estimate is then derived by the combination of difference in time and the treatment status. 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 non-treated group.

We conduct this for both the cross-sectional sample (any girls with baseline or midline/endline data)9 and panel sample (including the same girls recontacted in both periods) for GEC-T and only with cross-sectional sample for LNGB (due to data quality issues discussed in the limitations section). The difference-in-difference is conducted for all girls, and separately for disabled and non-disabled for comparison. We are not able to include any covariates in the difference-in-difference analysis because our sample sizes would drop significantly. Hence, the difference-in-difference estimates show progress made by the treatment girls over the non-treated group, disregarding any differences they may have in their characteristics – which may bias the estimates upwards or downwards depending on the intersection of these characteristics. As most of the excluded characteristics are working against disabled girls, this means that our coefficients are likely to be downward biased (as disabled girls are more likely to be poorer, and poorer girls are more likely to have lower learning outcomes).

The statistical significance of the difference-in-difference coefficients are reported for P-values below 0.05 and 0.10. The statistical analysis was conducted 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.

Beyond a traditional 'difference-in-difference' regression, we also compare changes in learning between disabled and non-disabled treatment girls (rather than a comparison of treatment and non-treated group), to investigate the extent to which disabled girls have made progress in learning relative to non-disabled girls. This difference-in-difference estimate shows the progress disabled girls have made over and above non-disabled girls.

Further, we analyse changes in intermediate outcomes (such as life skills and teaching quality) by running a 'difference-in-difference' for these variables. These regressions are run for the panel sample of girls, and separately for disabled and non-disabled girls in GEC-T. Girls are categorised as having or not having intermediate outcomes depending on the variables used to estimate intermediate outcomes. Some of the variables have answers of yes and no.10 In which case, girls are considered having the outcomes when their answers correspond to the intermediate outcomes.11 Other questions are based on Likert-type scales with answers ranging from 'Strongly Agree' to 'Strongly Disagree',12 'Often' to 'Never',13 and 'Never' to 'Everyday'.14 For these questions, girls are categorised to have the intermediate outcome when they answered, 'Strongly Agree', 'Agree, or 'Neither Agree nor Disagree',15 'Often' or 'Sometimes', and 'Once or Twice' or 'Everyday'. Improvements in intermediate outcomes were measured as a change over time (from baseline to midline) in the proportion of treated girls having the intermediate outcome over and above non-treated girls. As the intermediate outcomes were worded negatively, negative coefficients (reduction in the outcome) indicate positive changes.

The analysis was not done for the LNGB portfolio due to limited data availability. Very few LNGB projects have available data on intermediate outcomes, especially in follow-up round and for disabled girls which does not allow for the difference-in-difference estimation.

9 We limit the analysis to projects with available midline or endline data to ensure the same sets of projects are being compared over time.
10 Including attendance, economic empowerment, teacher does not suggest ways to continue study after school, and aspects of feeling unsafe.
11 All intermediate outcomes were worded negatively for ease of interpretation. However, some of the actual questions asked are worded positively (such as 'Since the start of the most recent school year, has girl attended her (main) school on most days that the school was open?'). In this case, caregivers answering No are categorised as having the intermediate outcome of 'Did not attend school on most days since the start of the most recent school year'.
12 All life skills questions.
13 Teacher does not use a different language to help understand and Teacher does not encourage students to participate.
14 Teacher used physical punishment on other students in the past week and Teacher used physical punishment on girl in the past week.
15 Similar to the yes/no questions, all life skills categories in the intermediate outcomes section were worded negatively. In cases where the actual questions are worded positively (such as 'I can stay focused on a goal despite things getting in the way'), girls answering 'Strongly Disagree', 'Disagree', and 'Neither Agree nor Disagree' are categorised as having the intermediate outcome of 'Not focused on a goal'.

Comparison against Study 3

In Study 3 (Evaluation Study 3: Aggregate impact of GEC-T projects between baseline and midline), to analyse the change in learning and intermediate outcomes, the authors used difference-in-difference regressions for the panel sample (recontacted) girls and using beneficiary-population weights. Projects included in the analysis are those with both baseline and midline data.16 In the analysis of intermediate outcomes, the authors also control for age and overage for grade status.

In this study, there are some deviations on how the calculation is done compared to Study 3. Notably, project-equal weighting is used instead of beneficiary-population weights. This is because we lack information on the number of beneficiaries with a disability. Beneficiary-population weights would put more weight on projects with a greater number of beneficiaries, but not necessarily projects with more disabled girls. The analysis of changes in intermediate outcome analysis is carried out without the use of weights. Additionally, Study 3 excludes from subgroup analysis projects where the number of observations is less than 100. As this study focuses on the subgroup of girls with disabilities, for which the prevalence among projects is often low, we chose not to exclude any project based on sample size. We also do not control for age and overage status in the analysis of learning and intermediate outcomes to maximise sample and ensure consistency across windows, with the variable of overage not being applicable in the LNGB window since the girls are out-of-school. As a result, results and sample sizes deviate from those reported in Study 3.

2.3. Projects included in the analysis and sample sizes

This section describes the availability, and validity, of GEC II projects' external evaluation datasets for profiling and for examining changes in learning levels and intermediate outcomes among girls with disabilities. For the profiling of girls, we include all girls with girl-reported disabilities at baseline; and to examine changes in learning and intermediate outcomes analysis, we include girls in projects with both baseline and follow-up (midline or endline) data, girl-reported disabilities and valid learning data.17

GEC-T

The GEC-T supports girls and young women in 15 countries through 27 projects.18 For the purposes of this study, four projects19 and CAMFED Zambia were excluded from the study because they lacked information on girl-reported disabilities. This left 23 of the 27 GEC-T projects for analysis.

Profiling of beneficiary girls

These 23 GEC-T projects are included in profiling baseline descriptives of girls supported by GEC-T projects. **Profiling of baseline characteristics includes 34,030 girls receiving GEC-T interventions for whom disability data is available.**

Many useful intersectional factors for profiling – such as if girls are married, are a mother, or an orphan, and economic variables are usually asked in caregiver/household surveys. In GEC-T, when restricting to those with data available for both disability and other key variables, the number of girls observed dropped significantly (Table 3). The main reason for this is that such key variables are collected from PCGs, and this is only available for a subset of girls. For instance, when intersecting disability with economic barrier data, only 41% of the full sample is retained.

16 Childhope, Plan, and Opportunity are excluded from panel analysis as they have poor identifiers.
17 Disaggregated learning data with at least one available learning subtask.
18 CAMFED International and DLA multi-country projects are counted as a single project.
19 Avanti, STC DRC, Varkey, and Viva

Table 3: Number of GEC-T baseline observations with available data on disability and key variables

	Number of observations	% of total observation
Total treatment girl	46,704	
Has girl-reported disability, AND	34,030	73%
Has learning data	27,763	59%
Has at least one married, mother, or orphan	20,796	45%
Has at least one economic variable	19,291	41%
Has at least one intermediate outcome	33,208	71%

Examining changes in girls' learning and intermediate outcomes

Learning analysis is conducted for the cross-sectional sample (full sample of baseline and full sample of follow-up) and the panel sample (girls followed over time). The starting sample is the sample of 23 projects included in the profiling of beneficiary girls.

For the cross-sectional analysis examining changes over time, we further exclude two projects20 as they do not have midline data and three projects21 because their learning data presents data quality issues. This leaves 18 projects in our cross-sectional sample analysis of learning. For our panel analysis, we are left with 17 projects. VSO is excluded from panel analysis (for learning and intermediate outcomes) due to a lack of common girls' unique IDs between the rounds (See Appendix 6 for project details).

Table 4 summarises the number of projects included in profiling and in examining change over time.

Table 4: Available projects for analysis (GEC-T)

	Number of projects	Excluded projects
Total with baseline data	27	-
Sample for disability descriptive analysis (projects with baseline and disability data)	23	Avanti, CAMFED Zambia, STC DRC, Varkey, and Viva
Sample for cross-sectional learning analysis (projects with baseline, midline, and disability data with no data quality issues)	18	Link, STC MOZ, Plan, Childhope, and Opportunity
Sample for panel learning analysis and intermediate outcomes (previous sample minus projects without unique ID)	17	VSO

*Note: Only the projects that are suitable for cross-sectional and panel learning analysis are included here (17 projects with valid midline, valid learning, and disability data).

Attrition in GEC-T projects

We examine attrition rates for girls in GEC-T projects to check for potential biases in the longitudinal analysis (i.e., changes in learning and intermediate outcomes). Attrition is defined as the percentage of girls whose data were collected at baseline but who were not re-contacted at midline, considering projects with disability data at baseline and learning data at baseline and midline.22 We matched girls' unique identifiers between baseline and midline to determine whether a girl has been re-contacted. Projects where a girl's re-contact status cannot be assessed by matching their unique identifiers are excluded from the attrition calculation.

For the GEC-T window, the base to estimate attrition includes 44,918 girls from 17 projects with disability data at baseline and learning data at baseline and midline (Table 5), for both treatment and non-treated groups. The figures

20 Link and STC MOZ
21 Plan, Childhope, and Opportunity
22 This is consistent with the definition used in Study 3.

for the sample at midline include girls from baseline – those who were recontacted – plus 'top up' girls (i.e. new girls who replaced those who were not found at midline) and who have learning and disability data.

At a portfolio level, overall attrition is high – 36.4% of girls in baseline disability and learning data were not recontacted at midline. Attrition does not vary significantly between the treatment (36.1%) and the non-treated group (36.7%), but it does vary slightly between disabled (37.8%) and non-disabled girls (36.1%) – mainly driven by differences in the non-treated group, where a higher share of disabled girls (39.5%) were lost to attrition.

Approximately 10-11% of the girls with available learning data in the baseline sample in the GEC-T are disabled; this same prevalence remains for the sample recontacted at midline.

Table 5: Number of GEC-T observations with available learning data and girl-reported disability

Sample	Subsample	Sample at baseline	Sample at midline	Girls recontacted at midline		
		N	N	N	% recontacted	% attrition
All	All	44,918	47,634	28,589	63.65	36.35
	Disabled	4,594	3,816	2,857	62.19	37.81
	Non-disabled	40,324	43,818	25,732	63.81	36.19
Treatment	All	26,760	29,973	17,098	63.89	36.11
	Disabled	2,579	2,241	1,637	63.47	36.53
	Non-disabled	24,181	27,732	15,461	63.94	36.06
Non-treated	All	18,158	17,661	11,491	63.28	36.72
	Disabled	2,015	1,575	1,220	60.55	39.45
	Non-disabled	16,143	16,086	10,271	63.63	36.37

Attrition varies considerably across projects, as shown in Table 6, with three projects showing attrition levels of less than 10% and, on the other extreme, eight projects with attrition of 41-60%, including one with an attrition level of 61-100%.

Table 6: Attrition level by project in GEC-T

Level of attrition	Number of projects	Projects
Very low (less than 10%)	3	CAMFED (ex-BRAC), DLA, ICL
Low (11-20%)	1	MC Nepal.
Medium (21-40%)	5	AKF, BRAC, HPA, Relief, World Vision.
High (41-60%)	7	CAMFED International (Tanzania and Zimbabwe), CSU, EDT, LC, MC Nigeria, PEAS, WUSC.
Very high (61-100%)	1	CARE International.

Within projects, attrition rates are not very different for disabled and non-disabled girls: in 11 out of 17 GEC-T projects, the differences in attrition between disabled and non-disabled girls are within +/- 10 percentage points (p.p.), as shown in Table 7. At the extremes we have two groups of three projects where either girls with disabilities have attrition rates of more than 10% difference than girls with no disabilities, or girls with no disabilities have a higher attrition rate, of more than 10%, than disabled girls.

Table 7: Attrition level differences by disability status in GEC-T

Differences in attrition between disabled and non-disabled girls	Number of projects	Projects
Attrition is much higher for disabled girls (more than 10 percentage points difference)	3	HPA, MC Nigeria, PEAS
Attrition of disabled and non-disabled girls is within +/- 10 percentage points	11	AKF, CAMFED International (Tanzania and Zimbabwe), CAMFED (ex-BRAC), CARE International, CSU, DLA, EDT, ICL, Relief, World Vision, WUSC
Attrition is much higher for non-disabled girls (more than 10 percentage points difference)	3	BRAC, LC, MC Nepal[23]

Systematic attrition can bias our estimates of changes in learning outcomes over time. To test for attrition bias, we compare baseline literacy and numeracy levels of disabled and non-disabled girls who were lost to attrition with those who were recontacted at midline (see Appendix 7), within the treatment and non-treated group, to see whether baseline learning levels of girls lost to follow-up differ between groups.

When comparing disabled girls between treatment and non-treated groups, the data reveals some level of attrition bias. Those in the treatment group who were recontacted had higher baseline numeracy scores (4.75 p.p.) than those who were not recontacted, whereas disabled girls in the non-treated group who were recontacted had lower baseline literacy scores (3.62 p.p.) than those not re-contacted. However, no statistical difference was found for literacy in the treatment group, nor numeracy in the non-treated group.

When comparing disabled and non-disabled girls, there is less concern for attrition bias due to the fact that the pattern of score difference between non-disabled lost and recontacted is similar to that of disabled girls. In the treatment group, non-disabled recontacted girls also scored higher in baseline numeracy (8.03 p.p.) than non-disabled lost girls. In the non-treated group, those recontacted scored lower in literacy (2.29 p.p.) compared to those not re-contacted. Nonetheless, in the non-treated group, non-disabled recontacted girls scored higher (7.22 p.p.) in numeracy than non-disabled girls not re-contacted. This pattern was not seen in disabled girls in the non-treated group.

Additionally, when comparing between disabled and non-disabled girls, disabled girls had lower baseline numeracy and literacy scores than their non-disabled counterparts in both the recontacted and lost sample, as well as across treatment and non-treated groups.

LNGB

The LNGB supports girls and young women in 10 countries through 14 projects. Out of 14 LNGB projects, AKF (Afghanistan) and Population Council (Ethiopia) were excluded from the study as their data was not available at the time of writing this report. This left 12 projects (30 cohorts24) with quantitative data for analysis.

Profiling

A requisite for being included in the profiling analysis is the availability of disability data. VSO Nepal, WEI Ghana, and a cohort from Pin Nepal and Plan were further excluded from the study for not having girl-reported disability data. This left 10 projects (23 cohorts) for profiling of girls' baseline characteristics.

Profiling of baseline characteristics includes 13,296 girls receiving LNGB interventions for whom disability data is available. Similar to GEC-T data, when restricting the samples to have both disability information and other key variables, the sample size is reduced (Table 8).

23 Attrition is 0% for disabled girls in MC Nepal as the project has only one disabled girl at baseline who was recontacted at midline.
24 Each project consisted of multiple cohorts with differing targeting and programme pathways. Plan's submitted separate girl and household survey data which could not be linked due to a lack of common identifiers. We separated these out as two cohorts for ease of analysis.

Table 8: Number of LNGB baseline observations with available data on disability and key variables

	Number of observations	% of total observation
Total treatment girls	16,719	
Has girl-reported disability data, AND	13,296	80%
Has learning data	11,630	70%
Has at least one married, mother, or orphan	12,578	75%
Has at least one economic variable	11,575	69%
Has at least one variable for reason girl is not in school[25]	7,391	44%

Examining changes in girls' learning

To examine changes over time, it is imperative to have two data points; in this case, baseline and follow-up. The follow-up samples include girls recontacted from baseline and 'top up' girls. Due to different programme designs, some LNGB projects collect data on midline (6 cohorts) and others on endline (10 cohorts) from nine projects (See Appendix 6 for project details). This study utilises either midline or endline data as data from follow-up period. For one project, IRC Sierra Leone, both midline and endline data are available. In this case, we use midline as follow-up data given the endline lacked learning variables.26

The starting sample is 10 projects and 23 cohorts with baseline and disability data. Then, 10 cohorts27 were excluded because they do not have follow-up data. Next, Link C1 is excluded as there is no disability data available at follow-up and Street Child (all three cohorts) is excluded as it does not have disaggregated learning data. Lastly, ACTED C2 L&N (excluded for baseline descriptive analysis) is included for learning as it serves as endline to ACTED C1 L&N. In summary, data from 5 projects and 10 cohorts are available for cross-sectional analysis at follow-up.

Changes in learning are estimated using the cross-sectional sample for the LNGB portfolio, as most projects' data did not have common identifiers between baseline and midline or endline. Moreover, only three projects have data for any recontacted girls of which only a few are disabled girls. This is because two of the three projects have a very low share of disabled girls (CARE and Pin Nepal C2).

Table 9: Available projects for analysis (LNGB)

	Number of projects	Number of cohorts	Excluded projects
Total projects with data	12	30	
Total with baseline data	12	29	ACTED C2 L&N
Sample for disability descriptive analysis (projects with baseline and disability data)	10	23	Pin Nepal (C3), Plan HH, VSO C1, WEI (C1 Formal, C1 and C2 Nonformal)
Sample for cross-sectional learning analysis: follow-up (projects with baseline, follow-up, disability, and learning data)	5	10[28]	ACTED (C1 ALP), ActionAid C3, CARE C1 NFE, IRC Pakistan (Learn, Earn, and Distant), Link (C1 and C3), Pin Ethiopia (C1 and C3), Plan GS, SC Nepal (C1, C2, and C3)

Table 10 summarises the learning sample sizes of treatment girls at baseline and follow-up, and the percentage of recontacted girls at the portfolio level. The table is divided in two parts: one for projects with unique identifiers used to give an indication of recontact rates, and one for projects used for cross-sectional analysis of learning. For LNGB any available follow-up evaluation round is considered (midline or endline). This varies by project, but IRC Sierra Leone is

25 This variable is only asked to out-of-school girls. In GEC-T, as most girls are in school, the sample size with available data for this variable is very small for meaningful portfolio-level analysis.
26 This is because learning data is only available in midline and not endline.
27 The number of cohorts does not add up to the number of midline and endline cohorts as presented above because some cohorts without follow-up had already been excluded from the starting number of 23 cohorts on the basis of having no disability data (such as WEI).
28 ACTED C2 L&N is added here as it has endline data.

the only project for which all three evaluation rounds are available. Data for non-treated group is not presented here as only one project has non-treated groups.

For LNGB, for projects with common identifiers, the share of girls reporting a disability at baseline is about 9.1%. The same prevalence is found for the sub-sample of girls recontacted (Table 10). For the sample used in cross-sectional analysis, prevalence of disability is higher at baseline compared to follow-up (11.5% versus 9.1%, respectively).

Table 10: Number of LNGB observations with available learning data and girl-reported disability

Sample	Subsample	Sample at baseline	Sample at follow-up	Girls recontacted at follow-up	
		N	N	N	% recontacted
Treatment (projects with common unique identifiers: 3 projects and 5 cohorts)	All	3,764	5,251	2,580	68.54
	Disabled	344	493	231	67.15
	Non-disabled	3,420	4,758	2,349	68.68
Treatment (projects in cross-sectional learning analysis: 5 projects and 10 cohorts)	All	5,423	5,251	N/A	N/A
	Disabled	624	493	N/A	N/A
	Non-disabled	4,799	4,758	N/A	N/A

*Note: Only the projects that are suitable for cross-sectional and panel learning analysis are included here (with valid follow-up, valid learning and disability data).

We do not explore attrition for the LNGB window projects, because we do not examine changes over time using panel data for this group.

2.4. Limitations

1) There are differences across projects in terms of the tools (WG-SS or CFM set) used to collect data on disability and the level of difficulty carrying out an action considered by projects, which prompted decisions to ensure consistency in this study. This was accomplished by using girl-reported data on disabilities, following WG guidelines concerning the level of functional difficulty considered as a disability ("a lot of difficulty" or "cannot carry out the action at all"; and "daily" for depression and anxiety). This means that relevant data on disability are excluded (for example, where projects collect data from PCGs but not from girls).
2) GEC-T projects primarily implemented the WG-SS set, while most LNGB projects used the CFM scale instead. We use the tools that are frequently used by each window (WG-SS for GEC-T and CFM for LNGB). This resulted in a higher prevalence of disability being observed in the LNGB window due to the increased number of dimensions captured in the CFM set compared to the WG-SS set.
3) 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). The profiling of girls maximises the data available; as such, each variable used for profiling yields a different sample size. No attempt has been made to keep the sample size consistent as that would lead to a race to the bottom and a sample not representative of the portfolio of projects. Similarly, for examining changes over time, we are not able to include any covariates in the analysis as our sample size would drop significantly.
4) The LNGB data include a specific set of questions related to the barriers to schooling faced by girls. This data is not collected for the GEC-T which mostly focuses on in-school girls. Due to a lack of barrier data in the follow-up round, we did not investigate change over time in barriers to schooling for LNGB girls.
5) Difference-in-difference strategies are employed to assess learning changes on: (1) the cross-sectional (i.e., sample of baseline and midline/endline girls); and (2) panel sample (i.e., sample of girls followed over time).[29] Due to the lack of common girls' unique IDs between the rounds across most LNGB projects (at the time of writing this report), only GEC-T projects were able to be included in the panel analysis.[30] For LNGB projects, we look at

29 Projects with only baseline data available are excluded from learning analysis.
30 While the LNGB projects could in theory be included in panel analysis, the follow-up sample constitutes 493 disabled girls with any learning data, of which only 231 were recontacted (see Table 10). Given such a small sample size and coverage of only three projects at follow-up, we do not conduct a panel analysis for LNGB.

change over time in the cross-sectional sample, which included 493 disabled girls. Results from this analysis should be read with caution as they are based on a much smaller sample size for disabled girls compared to GEC-T.

6) The WG tools condense the complexities of conceptualising disability into an easy to implement tool. However, changes observed over time in reporting present a challenge for analysis (e.g., a girl may report having a disability at baseline, but then not report it at the midline or endline, or vice versa). To avoid this, whenever possible, we use baseline data to assign disability status. The cross-sectional learning analysis is an exception, meaning that a girl can have a different disability status at baseline and midline/endline. Results from this analysis should be interpreted with caution as it is not possible to measure the potential bias of such transitions between states.
7) Data is gathered for the GEC-T external evaluations from a sample of 'treated' girls (girl beneficiaries) and a sample of girls with comparable characteristics who do not receive any intervention (non-treated). As such, the effects of GEC II interventions relative to the non-treated group can be, and is, estimated. The majority of LNGB window projects do not include non-treated groups.31 Therefore, it is not possible to attribute gains in girls' outcomes relative to a non-treated group for the LNGB. For the LNGB, the alternative is to focus the analysis on treatment girls – first how they learn over time, and also comparing this to non-disabled girls. While we acknowledge that girls with disabilities may have different challenges compared to girls with no disabilities, the comparison yields useful information in that finding no differences between them would constitute a positive outcome.
8) Intermediate outcomes are assessed through individual questions asked to girls or PCGs which serve as proxy indicators for attendance, economic empowerment, life skills, quality of teaching, gender-based violence, and parental attitudes. These questions are adapted from various scales and provide an indication of understanding change in these outcome areas, even though they are not measured reliably. Changes in the intermediate outcomes are measured as shifts in answers based on Likert scales (ranging from Totally Agree to Totally Disagree).
9) There may be evidence of attrition bias, which affects analysis involving change over time in learning outcomes. In the treatment group, disabled recontacted girls scored higher in baseline numeracy than those lost to attrition. In the non-treated group, however, disabled recontacted girls scored worse in baseline literacy than those lost to attrition. This implies that those recontacted in treatment and non-treated group have different baseline literacy and numeracy than those lost to attrition. Learning gains of treatment group are analysed over non-treated group with lower literacy level than the overall baseline non-treated sample. Vice versa, recontacted treatment group have higher level of numeracy than the overall baseline treatment sample.

3. Profiling

3.1. Prevalence of disability

At portfolio level

In this section, we examine the prevalence of functioning disabilities among girl beneficiaries32. To do this, we are reliant on the numbers in the sample rather than the overall beneficiaries. As projects do not report the total number of girls with disabilities, just overall beneficiaries, we can only conduct the profiling on the number of girls who have been sampled, and the share of each project evaluation sample who have girl-reported disability.

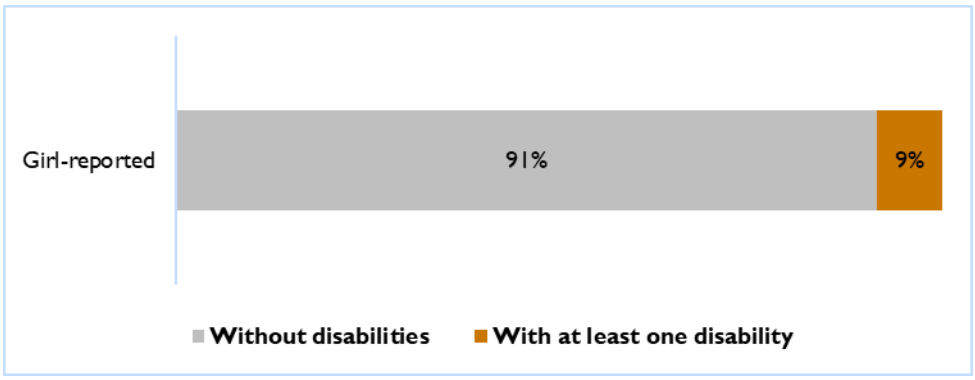
We report here the full set of girls with a disability across both GEC-T and LNGB portfolios, regardless of if they have learning data nor midline/endline data.

Overall, at baseline, 8.6% of the GEC-T sampled girls report having at least one disability. Unsurprisingly the disability focused projects CSU (43%) and LC (36%) have the highest share – but it is noticeable this is below half (and nearly 1/3rd) even for those projects. Others, such as LINK (25%) have a high proportion of girls with disabilities in their

31 Because of the ethical implications of targeting highly marginalised girls, the majority of programmes do not have non-treated groups; instead, they offer the intervention to all girls identified as meeting the programme criteria. One project, Pin Nepal, did employ a non-treated group with girls in that group intended to receive the intervention in enrolling cohorts.
32 The report presents many figures with low percentages (less than 10%). When these low percentages were presented in the Figures, they were rounded to whole numbers. The number is rounded down if the value of the first decimal falls between 1 and 4 and rounded up if the value falls between 5-9. For precision, when the low percentage figures were described in the text, they are reported with one decimal place.

target groups. Some of the projects with a large number of beneficiaries have a relatively greater proportion of girls with disabilities, especially CAMFED ex-BRAC (17%), CAMFED Tanzania (17%), CAMFED Zimbabwe (16%) and DLA Ghana (10%). Data disaggregated by projects can be found in Appendix 8.

Figure 1: Disability status of GEC-T baseline treatment girls



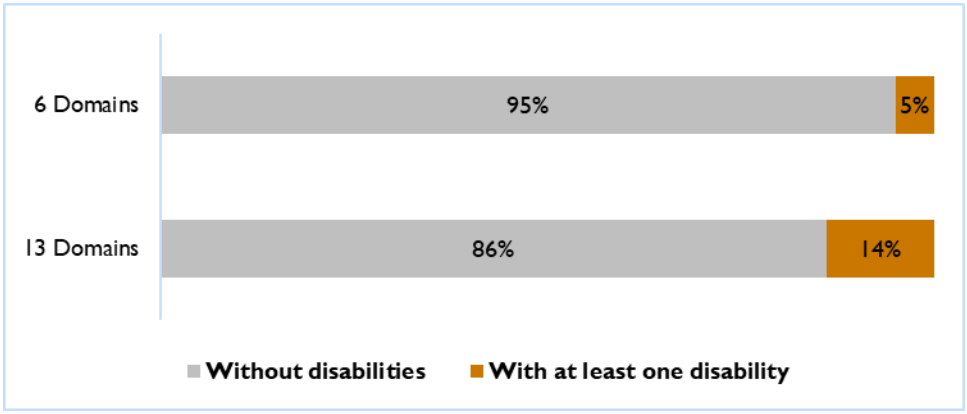
Note: N = 34,030

For LNGB girls, girl-reported data shows that 4.9% of girls have disabilities based on the six core domains (seeing, hearing, walking, remembering, self-care and communicating). This percentage is lower than the 8.6% prevalence of disability found in the GEC-T sample for the same six core domains.

LNGB girls were asked the Child Functioning Module (CFM), which gathers data on seven additional domains: difficulties learning, concentrating, accepting change, controlling behaviour, making friends, anxiety and depression33. Using the CFM set, 14% of girls report at least one functional disability.

Based on girl-reported data, Cohort 3 of Link (24%) has the greatest proportion of girls with disabilities, followed by Plan GS cohort (18%) and Link Cohort 1 (17%). Using all CFM domains, most LNGB projects and cohorts (20 out of 23 with girl-reported disability data) demonstrate an increase in the proportion of girls with disabilities compared to the common six domains. Cohort 3 of Link (41%) has the largest proportion of girls with disabilities followed by Link Cohort 1 (34%) and ActionAid Cohort 1 (33%). The estimates of LNGB girls with disabilities by project can be seen in Appendix 8.

Figure 2: High level disability status of LNGB baseline treatment girls



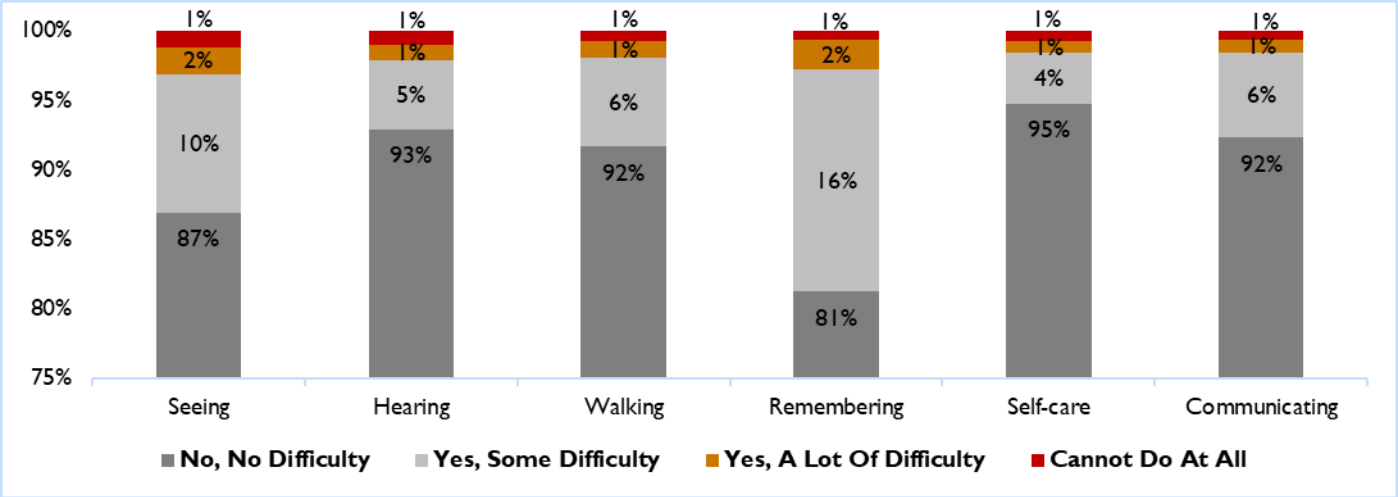
Note: N 6 domains = 13,295; N 13 domains = 13,296. The sample differs by one, since only one girl has data for the seven CFM dimensions and not the core six domains.

The figures below show the severity level of disability for each category based on girl-reported data from girls. The most common functioning difficulty among GEC-T girls are difficulties remembering, with 19% reporting this. Looking just at those with "a lot of difficulty" or who "cannot do at all", the shares fall noticeably and are relatively even across the groups - though visual disabilities are slightly more prevalent: 2.0% of the GEC-T girls reported they have a lot of difficulty in seeing and 1.2% said they cannot see at all. The least common disability among GEC-T girls is self-care

33 The extended CFM measures functional difficulties, which include domains beyond traditional measures of physical difficulty.

and communication disability: only 0.9% of the GEC-T girls reported they have a lot of difficulty in communicating while 0.7% said they cannot communicate at all.

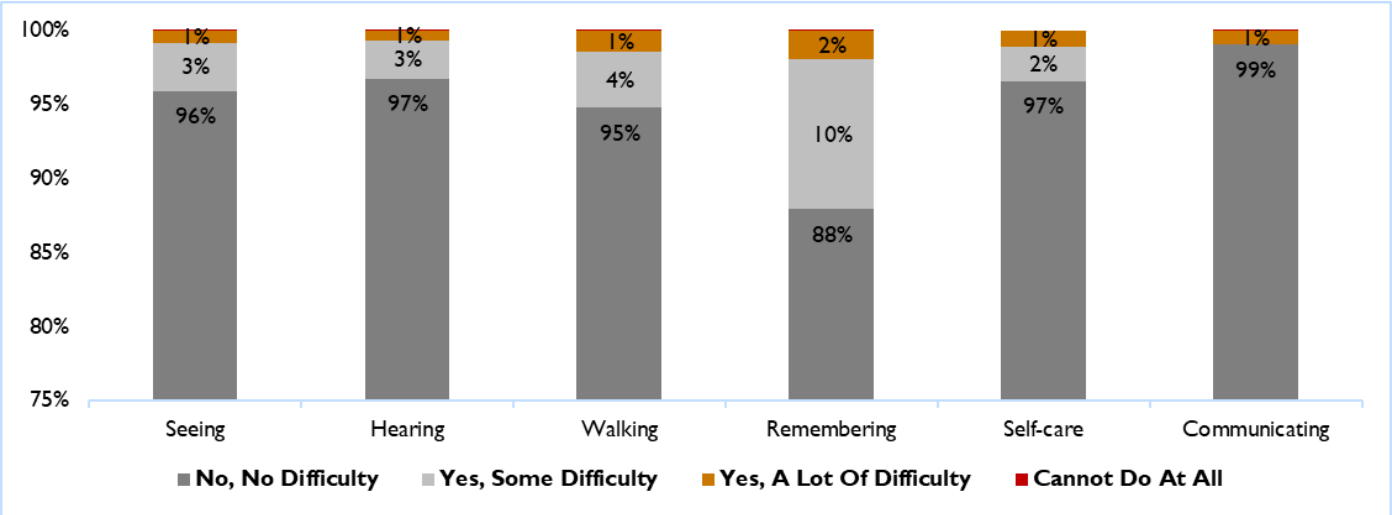
Figure 3: Disability status of GEC-T baseline treatment girls (based on girl-reported disability)



Note: For visual assistance, y-axis starts at 75%. N seeing = 31,103; N hearing = 31,126; N walking = 33,795; N remembering = 33,854; N communicating = 33,841

Similarly, LNGB girls also report difficulties remembering, though at a lower share at 12%. Again, once we use the stricter definition the shares fall: 1.9% of the LNGB girls reported they have a lot of difficulty or answered "cannot do at all" in remembering. The least common disability among LNGB girls is hearing disability as 0.7% of the LNGB girls reported they have a lot of difficulty or answered they cannot hear at all.

Figure 4: Disability status of LNGB baseline treatment girls – Six common domains (based on girl-reported disability)



Note: For visual assistance, y-axis starts at 75%. *N seeing = 12,833; N hearing = 12,830; N walking = 10,165; N remembering = 12,846; N communicating = 12,946

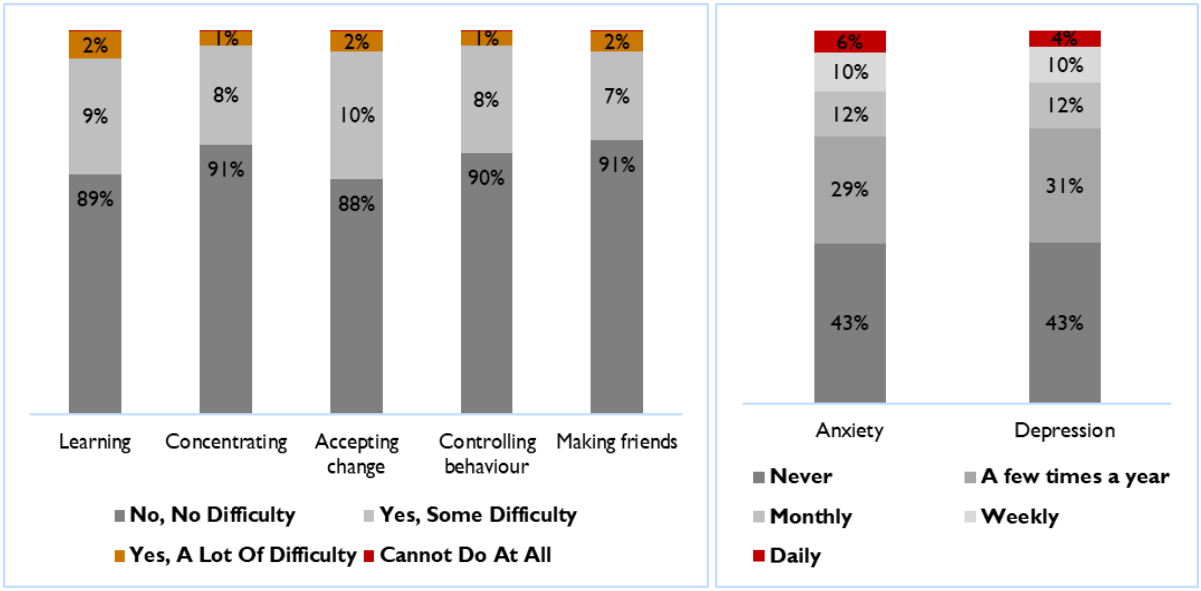
For LNGB, we can look wider across the other domains. The other seven domains consist of two categories: five domains with four scales (from 'no difficulty' to 'cannot do at all') same as the common domains and two other domains (anxiety and depression) with different scales ranging from "never" to "daily". It is important to note that using the recommended cut-off of CFM, the disability indicator includes "daily" for the questions on anxiety and depression and "a lot of difficulty" as well as "cannot do at all" for all the other categories.

When looking at daily occurrences of the other seven domains, the mental health related figures are striking. The most commonly reported domain is anxiety (5.9%) followed by depression (4.3%) while the least common disability is concentrating (1.1%). The majority of LNGB girls who have disabilities reported they have 'a lot of difficulties' rather than 'cannot do at all' for the learning, concentrating, accepting change, controlling behaviour and making friends.

The share of girls reporting depression and anxiety at any time is noticeably high – with only 43% reporting they never experience these. Having more than one in two girls experiencing depression in the sample is worth further investigation and support. Moreover, 5.9% and 4.3% reported having difficulties with anxiety and depression on a daily basis. This is an important finding, especially considering that UNESCAP (n.d.)'s tested the WG tool in six low- and middle-income countries and found that 76% and 84% of youth and children age under 18 reported never experiencing anxiety and depression (a much higher prevalence compared to LNGB girls) and 3.0% and 1.3% reported experiencing anxiety and depression daily (a much lower prevalence than for LNGB girls. These findings further highlight the need to assess and incorporate measures related to mental health in future programme design.

Some projects were seen with a higher proportion of disabilities in anxiety and depression: 17% of ACTED Cohort 1 L&N reported they feel anxious, nervous or worried every day, followed by Link Cohort 1 (13%) and ActionAid cohort 1 (10%). Also, 12% of Link Cohort 1 girls reported they feel sad or depressed each day, followed by ACTED Cohort 1 L&N (9.7%) and ActionAid Cohort 1 (9.0%).

Figure 5: Disability status of LNGB baseline treatment girls– the other seven domains (based on girl-reported disability)



Note: N learning = 13,216; N concentrating = 13,184; N accepting change = 13,152; N controlling behaviour = 13,164; N making friends = 13,217; N Anxiety = 13,207; N Depression = 13,199

Prevalence by projects

The GEC-T window has 27 projects, of which two are multi-country. For simplicity and given the differences in prevalence of disabled girls within those two multi-country projects, Table 11 treats each project-country as a separate project. Out of 31 GEC-T country-projects34, five do not have girl-reported disability data. Out of the 26 projects that do have disability data, 11 have between 2-6% of girls with disabilities in their treatment sample, two projects have between 6-10% of girls with disability, seven projects have over 10% of girls with disabilities, while six projects have less than 2% of girls with disabilities in the treatment sample.

34 CAMFED International and DLA (multi-country projects) are counted as three different projects per country

Table 11: Number of projects by share of girls with disabilities in their intervention at BL – GEC-T (girl-reported disability)

Windows	Share of girls with disabilities	Number of projects	Projects
GEC-T	< 2%	6	AKF, Care, Childhope, MC Nepal, Relief, STC MOZ
	2-6%	11	BRAC, DLA Nigeria and Kenya, HPA, ICL, MC Nigeria, Opportunity, Plan, PEAS, VSO, WUSC
	6-10%	2	EDT, World Vision
	>10%	7	DLA Ghana, CAMFED Tanzania, CAMFED Zimbabwe, CAMFED ex-BRAC, LINK Projects focusing on girls with disabilities: CSU and LC

Note: Five GEC-T projects (Avanti, CAMFED Zambia, STC DRC, Varkey and Viva) do not have girl-reported disability data.

The LNGB Window has data on 12 projects with 30 different cohorts; however, seven of these cohorts do not have girl-reported disability data.35 Of the remaining 23 cohorts, these, nine cohorts have less than 2% of girls with disabilities in the treatment sample when using the core six disability domains. Five cohorts have more than 10% of girls with disabilities in their intervention. Looking at the full CFM set – 13 domains – increases the number of cohorts in the group of more than 10% of girls with disabilities to 12.

Table 12: Number of projects by share of girls with disabilities in their intervention at BL – LNGB (girl-reported disability)

Windows	Share of girls with disabilities	Number of cohorts	Projects (Cohort)
LNGB (6 common domains)	<2%	9	ACTED (C4 L&N), Care (C1 ABE), Care (C1 NFE), Care (C1 Formal), IRC Pakistan (Learn), Pin Nepal (C2), Street Child (C1), Street Child (C2), Street Child (C3)
	2-6%	9	ACTED (C1 L&N), ACTED (C1 ALP), ActionAid (C3), Care (C4 NFE), IRC Pakistan (Earn), IRC Pakistan (Distant), IRC SL (C1), Pin Ethiopia (C1), Pin Nepal (C1)
	6-10%	0	N/A
	>10%	5	ActionAid (C1), Link (C1), Link (C3), Pin Ethiopia (C3), Plan (GS)
LNGB (13 domains)	<2%	4	Pin Nepal (C2), Street Child (C1), Street Child (C2), Street Child (C3)
	2-6%	3	Care (C1 ABE), Care (C1 NFE), Care (C1 Formal)
	6-10%	4	ACTED (C4 L&N), IRC Pakistan (Learn), IRC Pakistan (Earn), Pin Nepal (C1)
	>10%	12	ACTED (C1 L&N), ACTED (C1 ALP), ActionAid (C1), ActionAid (C3), Care (C4 NFE), IRC Pakistan (Distant), IRC SL (C1), Link (C1), Link (C3), Pin Ethiopia (C1), Pin Ethiopia (C3), Plan (GS)

Note: For the LNGB project, ACTED (C2 L&N), Pin Nepal (C3), Plan (HH), VSO (C1), WEI (C1 Formal), WEI (C1 Nonformal), WEI (C2 Nonformal) do not have girl-reported disability data.

3.2. Baseline characteristics of girls with disabilities

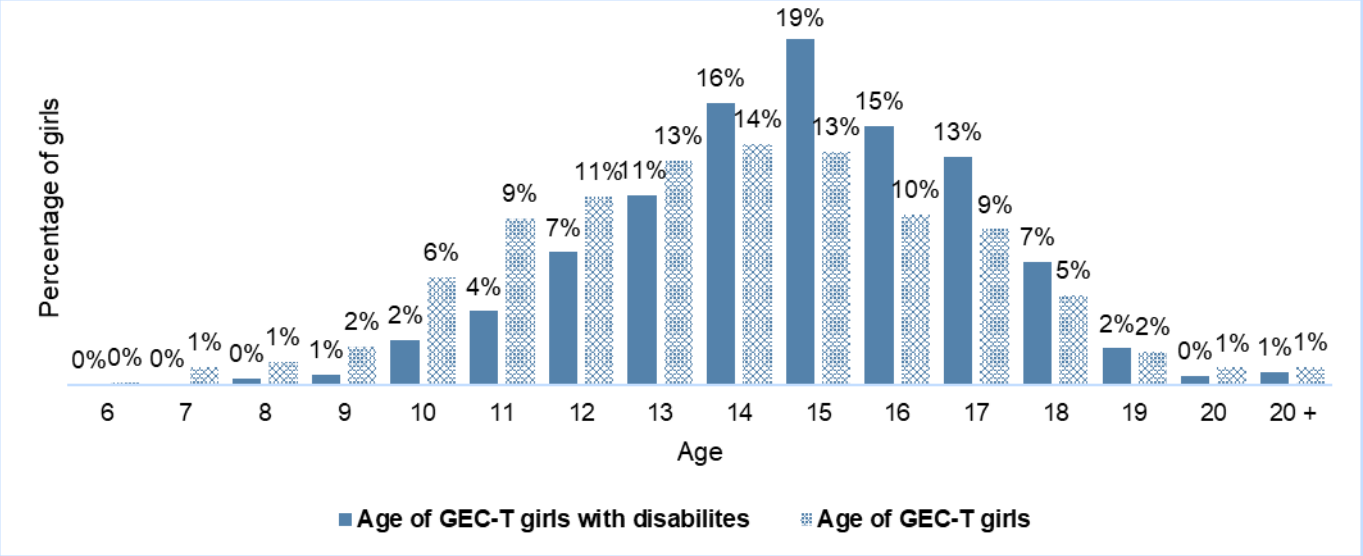
GEC-T

The age distribution for all GEC-T girls and GEC-T girls with disabilities is shown in the graph below. The average age of girls with disabilities in the sample is 15 years old. The youngest girls with disabilities are 7 years old, making up 0.1% of girls with disabilities. The oldest girls with disabilities are older than 20 years old36, representing 1.3%. The

35 Different cohorts within the same project were treated as different projects.
36 No specific age was provided

data indicates that GEC-T girls with disabilities had a less evenly distributed age range than GEC-T girls as a whole. The highest proportion of age of girls with disabilities is the 15-year-old group which represents 19%.

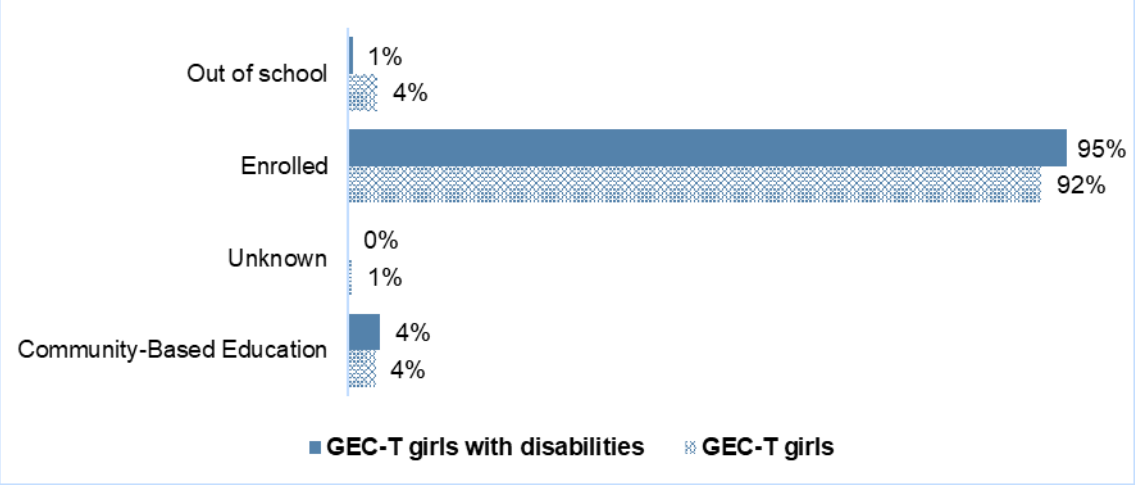
Figure 6: Age distribution of GEC-T girls with disabilities and GEC-T girls



Note: This figure differs slightly from the figure presented in the main report as it compares GWDs to all girls in the portfolio rather than GWDs and girls without disabilities.

We see that 95% of GEC-T girls with disabilities are enrolled in school at the time of baseline data collection, 4.2% were in community-based education (CBE), and only 0.5% of these girls were out of school. This is greater than the share of girls without disabilities, who are more likely to report being out of school. The GEC-T window primarily targets girls enrolled in school and supports girls with disabilities already in school. Only a few projects support out-of-school girls. This difference is likely due to sampling and targeting and differences across project designs.

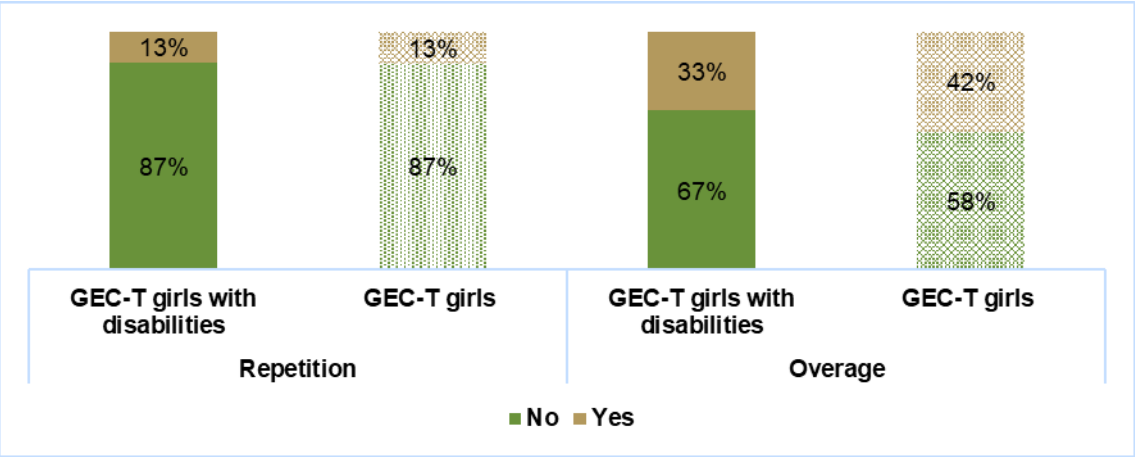
Figure 7: Enrolment status of GEC-T girls with disabilities and GEC-T girls



Note: This figure differs slightly from the figure presented in the main report as it compares GWDs to all girls in the portfolio rather than GWDs and girls without disabilities.

Figure 8 examines repetition and overage, showing that 13% of girls with disabilities were repeating a grade. Defining girls who were one or two years older than the official predicted age for the grade as "overage" girls, 33% of GEC-T girls with disabilities were overage. For repetition, this places girls with disabilities as equally likely to repeat, while they are less likely to be overage than non-disabled girls.

Figure 8: Repetition and overage status of GEC-T girls with disabilities and GEC-T girls



Note: This figure differs slightly from the figure presented in the main report as it compares GWDs to all girls in the portfolio rather than GWDs and girls without disabilities.

The table below shows some other characteristics of GEC-T girls with disabilities compared to the overall sample. Just 0.8% of the girls with disabilities are married, and 1.8% of them have given birth – lower or equal to the non-disabled girls. 19% of girls with disabilities are single orphans, while 4.7% of them have lost both the parents.

Table 13: Characteristics of GEC-T girls with disabilities

Characteristics	GEC-T girls with disabilities	GEC-T girls
Girl is married	1%	2%
Girls is a mother	2%	2%
Single orphan	19%	14%
Double orphan	5%	2%
Poor[37]	30%	24%

Disabled girls are more economically disadvantaged than non-disabled girls. Unpacking this, a greater proportion of disabled girls report going without cash income, food, and clean water most of the time, not being able to meet basic needs without charity, as well as having difficulty paying for school. The difference is greatest in ability of household to meet basic needs without charity, with 68% of disabled girls reporting not being able to meet basic needs. This is almost double the proportion reported by non-disabled girls. The difference is also large in terms of households not having enough clean water for use and difficulty paying to send girls to school. This shows that in GEC-T, disabled girls are more marginalised than non-disabled girls, and implies that disabled girls face more barriers to education than non-disabled girls.

37 Five different categories were reviewed to reflect the economic situation of the household. "Poor" indicates that a girl is from a household which has more than 50% of disadvantages out of these categories.

Table 14: Comparison between disabled and non-disabled GEC-T girls in being poor

	Household does not have land	Household has gone without cash income most times	Household has gone without enough clean water for use at home	Household has gone to sleep at night feeling hungry	Household has difficulty meeting basic needs without charity	Household has difficulty paying to send girl to school
Non-disabled mean	29%	21%	6%	3%	36%	55%
Disabled mean	32%	34%	9%	6%	68%	63%
Difference	-2%	-12%	-3%	-3%	-32%	-8%
P-value	0.08	0.00	0.00	0.00	0.00	0.00

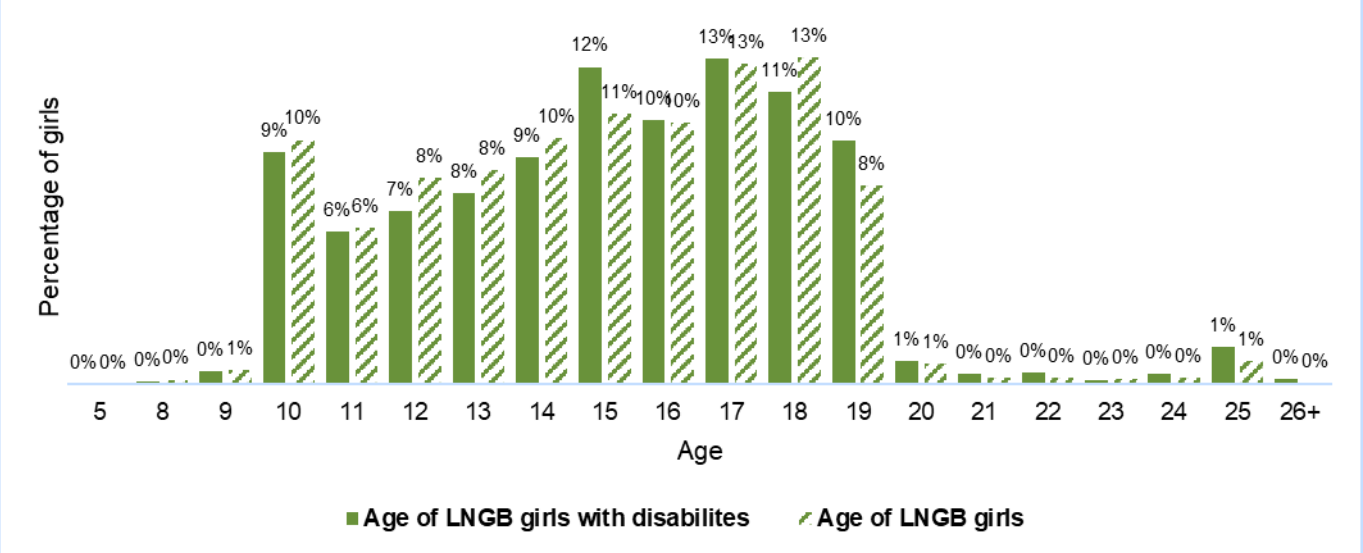
For a subsample of girls, we can see how the characteristics changed over time – and they change very little. The key status of important life events – being married, mother, and orphan do not change. Most poverty indicators do not change significantly over time for both disabled and non-disabled girls. One economic indicator, difficulty in meeting basic needs without charity, shows an increase of 5.0% for non-disabled girls and a decrease of 15% by non-disabled girls over time.

LNGB

The figure below shows the age distribution of LNGB girls with disabilities. These figures are based on 13 categories collected using CFM.

The average age of girls in the sample with disabilities is 15 years old. In the LNGB, the youngest girls with disabilities are 8 years old, which represents 0.5% of girls with disabilities. The oldest woman with disabilities is 35 years old, representing 0.1%.

Figure 9: Age distribution of LNGB girls with disabilities and LNBG girls



Note: This figure differs slightly from the figure presented in the main report as it compares GWDs to all girls in the portfolio rather than GWDs and girls without disabilities.

The table below shows the other characteristics of LNGB girls with disabilities.38 22% of LNGB girls with disabilities are married and 32% have a child or children. The figures are much higher for LNGB compared to GEC-T as LNGB projects particularly target girls who are married and have children. 24% of them are single orphans and 3.4% double orphans. Similar to the GEC-T girls (see Table 13), disabled girls are more likely to be orphans and poor compared to non-disable girls. 30% of LNGB girls are from a poor household.

38 Enrolment, repetition, and overage were not included in the description since LNGB Windows targets out-of-school girls.

In contrast to the GEC-T girls, disabled girls in this sample are more marginalised than non-disabled girls in a sense that they are also more likely to be married at a young age and be a mother.

Table 15: Characteristics of LNGB girls with disabilities

Characteristics	LNGB girls with disabilities	LNGB girls
Girl is married	22%	19%
Girl is a mother	32%	30%
Single orphan	24%	18%
Double orphan	3%	2%
Poor	30%	20%

Similar to GEC-T, in LNGB, girls with disability are more marginalised than girls without disability in terms of household economic situation39. A higher proportion of disabled girls has gone without cash, water, and food, and are less likely to meet basic needs without charity. They are, however, not poorer in terms of land ownership. More households of girls with disability own land, compared to households of girls without disability.

Table 16: Comparison between disabled and non-disabled LNGB girls in being poor

	Household does not have land	Household has gone without cash income most times	Household has gone without enough clean water for use at home	Household has gone to sleep at night feeling hungry	Household has difficulty meeting basic needs without charity
Non-disabled mean	44%	23%	16%	10%	48%
Disabled mean	34%	38%	24%	18%	57%
Difference	10%	-15%	-8%	-7%	-9%
P-value	0.00	0.00	0.00	0.00	0.00

Additional barriers to schooling faced by disabled girls

The LNGB data include a specific set of questions related to the barriers to schooling faced by girls. This set of questions were asked to caregivers, for a subset of girls. Table 17 shows reasons listed for girls not being in school. The reasons are similar among caregivers of disabled and non-disabled girls, with the top six reasons being the same for caregivers of disabled and non-disabled girls.

Overall, the biggest barrier to schooling is finance – the main reported reasons for disabled girls being out of school are that there is not enough money to pay for girls' education (71%) and that the girls have to work (32%). The main reasons are the same for both caregivers of disabled and non-disabled girls (albeit with differences in percentages). Following the financial reasons are reasons related to accessibility of the school. Lack of transport services and the school being too far away are listed by almost one in five caregivers, both of disabled and non-disabled girls.

The table also highlights several factors that do not seem to influence the caregiver's decision in girls' enrolment. The reasons with the least influence on enrolment are reasons related to school environment such as being unable to use the toilet or move around school. Reasons related to perception on girls' characteristics (girl being too old, had enough schooling, not mature enough) are also ranked low.

Financial constraints affect caregivers of girls with disability more than caregivers of non-disabled girls. There is a higher proportion (12 p.p. difference) of caregivers answering "There isn't enough money to pay the costs of (name)'s schooling" for disabled girls than caregivers of non-disabled girls. This resonates with the results from the previous section showing that disabled girls are more economically disadvantaged than non-disabled girls.

39 A question of whether household has difficulty in paying to send girls to school is omitted as most LNGB girls are out of school, and this question is not relevant.

Another reason that resonates more with caregivers of girls with disability than caregivers of non-disabled girls is "Girl has a health condition that prevents (her) from going to school", with 13% of disabled girls' caregivers answering yes compared to only 4.2% of non-disabled girls' caregivers.

Primary caregivers of disabled girls also reported other reasons in slightly higher proportion (statistically significant but smaller in magnitude) than primary caregivers of non-disabled girls. These include safety during travel to school, needing special services in school, teachers not knowing how to treat a childlike (name), girl too old, about to get married or have a child.

Table 17: Reasons for girl not attending school (reported by the primary caregivers)

Reasons	Disabled girls	Non-disabled girls
There isn't enough money to pay the costs of (name)'s schooling	71%	58%
Girl needs to work, earn money or help out at home	32%	28%
Transport services are inadequate	20%	18%
School is too far away	16%	17%
To attend school, girl needs special services or assistance such as speech therapist, support worker, sign language interpretation that not available	14%	11%
Girl is not interested in going to school	14%	11%
Girl has a child or is about to have a child	13%	10%
Girl has a health condition that prevents her from going to school	13%	4%
No one available to travel with girl to/from school	12%	12%
It is unsafe for girl to travel to/from school	12%	9%
To attend school, girl needs assistive devices/technology such as braille textbook, hearing aid, wheelchair, etc. that are not available	11%	9%
School does not help girl in finding a good job	10%	8%
It is unsafe for girl to be in school	9%	7%
The school does not have a program that meets girl's learning needs	8%	5%
Girl is too old to attend school	8%	5%
Schooling not important for girl	7%	6%
Girl is not mature enough to attend school	7%	5%
Girl is married or about to get married	7%	5%
Teachers do not know how to teach a childlike (name)	7%	4%
Girl says they are mistreated/bullied by other pupils	6%	4%
Girl has completed enough schooling	5%	3%
Child was refused entry into the school	4%	4%
Girl cannot move around the school or classroom	3%	3%
Girl cannot use the toilet at school	3%	3%

4. Examining changes in outcomes for girls with disabilities

4.1. Learning outcomes

GEC-T

GEC-T projects, unlike the LNGB, also include a non-treated sample of girls which allows for an estimate of the difference-in-difference – that is, how girls' learning outcomes change over time, compared to a non-treated group.

Looking at disabled girls, it is clear that the learning levels in baseline are low, and slightly unbalanced in some literacy subtasks between treatment and non-treated girls. While there is not a clear pattern with treatment girls scoring higher in some, and lower in other subtasks, overall, treatment girls achieved 4 p.p. more in overall literacy at the lower level (EGRA) relative to non-treated girls.

For numeracy, the baseline achievement in some subtasks is unbalanced, across the treatment and non-treated group of disabled girls, with non-treated girls scoring higher in the simpler tasks, but much lower in the more difficult subtraction tasks.

Table 18: GEC-T literacy learning levels of disabled girls in baseline

Baseline; Disabled	EGRA						SeGRA			
Learning levels in percentage points	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
Non-treated [mean]	30	55	43	60	30	40	46	24	15	28
Treatment [mean]	33	59	46	54	32	44	45	23	14	27
Difference	-3	-4*	-2	6**	-2	-3*	0	1*	1	1
T-statistics	-1.13	-1.65	-0.86	2.22	-1	-1.71	0.58	1.89	1.47	1.25
P-value	0.26	0.10	0.39	0.03	0.32	0.09	0.56	0.06	0.14	0.21

*Note: Difference 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.

Table 19: GEC-T numeracy learning levels of disabled girls in baseline

Baseline; Disabled	EGMA									SeGMA			
Learning levels in percentage points	Num. ident.	Quantity Discrim.	Missing Num.	Addition Level 1	Addition Level 2	Subtract. Level 1	Subtract. Level 2	Word Prob.	Overall EGMA	Mult./divis, fractions, geometry	Algebra	Sophist. Word problems	Overall SeGMA
Non-treated [mean]	81	79	41	69	64	59	48	45	59	26	11	5	14
Treatment [mean]	75	75	46	67	69	56	62	47	59	27	10	4	14
Difference	6**	4*	-5**	2	-5	3	-13**	-2	0	-1**	1	0	-1*
T-statistics	2.59	1.79	-2.25	0.95	-1.54	1.54	-3.49	-0.97	0.19	-2	0.92	0.8	-1.77
P-value	0.01	0.07	0.02	0.34	0.13	0.12	0.00	0.33	0.85	0.05	0.36	0.42	0.08

*Note: Difference 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.

We look at whether an average disabled treatment girl has learned significantly more than an average disabled non-treated girl between baseline and midline. We find that they have improved in literacy (EGRA and SeGRA) and numeracy (SeGMA) more than the average non-treated girl.

These improvements, measured by a difference-in-difference (DID) estimator, are statistically significant at the 95% confidence level for SeGRA (Panel P-value 0.009), and SeGMA (CS P-value 0.047), and at the 90% confidence level for EGRA (Panel P-value = 0.054) and SeGMA (Panel P-value = 0.056). The significant difference of these results, however, disappears when project-equal weights are applied (apart from EGRA for the cross-sectional sample which actually becomes significant with P-value = 0.016), suggesting that they stem from a few high performing projects with a high share of disabled girls (such as CAMFED).

The scale of girls' improvement is +4 p.p. and +3 p.p. in lower and higher-level literacy, respectively, and +3 p.p. in higher level numeracy for the panel sample (recontacted girls). It is about the same size in the cross-sectional sample (all girls) but in EGRA, where the difference is about twice as large, although not statistically significant when weights are not applied.

Table 20: Learning improvements of the average disabled GEC-T girls over and above non-treated girls

Difference-in-difference in percentage points		Literacy		Numeracy	
		EGRA	SeGRA	EGMA	SeGMA
Cross-sectional sample (all disabled girls; no weighting)	% correct score	+8	+3	+3	+3**
	P-value	0.101	0.191	0.325	0.047
Cross-sectional sample (all disabled girls; project-equal level weighting)	% correct score	+9**	0	+2	0
	P-value	0.016	0.964	0.592	0.948
Panel sample (recontacted disabled girls only; no weighting)	% correct score	+4*	+3**	+1	+3*
	P-value	0.054	0.009	0.560	0.056
Panel sample (recontacted disabled girls only; project-equal level weighting)	% correct score	+6	+2	+1	+2
	P-value	0.136	0.400	0.625	0.631

*Note: 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.

Table 20 shows that the observed increase in learning over and above the non-treated groups for the recontacted girls, is statistically significant, but is relatively small in size.

Disabled GEC-T girls who were recontacted have improved by 3-4 p.p. more than girls in the non-treated group over eighteen months40, 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 disabled GEC-T girls have improved (or not) in specific subtasks. Table 21 breaks down girls' increases over and above the non-treated group across all literacy subtasks.

Table 21: Literacy improvements of disabled GEC-T girls over and above non-treated girls - % correct scores

Difference-in-difference in percentage points		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
Cross-sectional sample (all disabled girls; no weighting)	**% correct**	+2	+2	+6	+16**	+2	+8	+2	+3**	+4**	+3
	P-value	0.689	0.666	0.140	0.015	0.574	0.101	0.662	0.351	0.041	0.191
	N (BL)	699	929	992	911	928	1,093	3,830	3,693	3,558	3,894
	N (ML)	699	1,220	1,238	1,345	1,301	1,383	2,290	2,227	2,120	2,290
Panel sample (recontacted disabled girls only; no weighting)	**% correct**	+4	+2	0	+5	+6*	+4*	+2	+5**	+5**	+3**
	P-value	0.222	0.535	0.862	0.158	0.095	0.054	0.216	0.001	0.007	0.009
	N (per round)	454	610	613	576	562	712	2,031	1,902	1,846	2,071

*Note: 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 overall EGRA increase reported in Table 20 is mostly a result of the increase in oral reading fluency, in which disabled GEC-T girls improved by +16 p.p. over and above non-treated girls and reading comprehension (Panel sample +6 p.p.). For SeGRA, the largest differences are observed in the subtasks of longer reading comprehension and short essay construction (+5 p.p.).

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

Table 22: Numeracy improvements of disabled GEC-T girls over and above non-treated girls - % correct scores

Baseline; Disabled		EGMA									SeGMA			
		Num. Ident.		Quantity Discrim.	Missing Num.	Addition Level 1	Addition Level 2	Subtract. Level 1	Subtract. Level 2	Word Prob.	Overall EGMA	Mult./divis., fractions, geometry	Algebra	Sophist. Word problems	Overall SeGMA
Cross-sectional sample (all disabled girls; no weighting)	**% correct**	+8	+7	+3	+7	-2	+10*	-11**		-2	+3	+4*	+4**	+3**	+3**
	P-value	0.117	0.285	0.373	0.113	0.257	0.046	0.012		0.446	0.325	0.054	0.050	0.004	0.047
	N (BL)	409	586	1,074	928	353	1,053	383		1,095	1,100	4,182	3,672	3,551	4,269
	N (ML)	393	559	1,264	1,154	728	1,269	741		1,324	1,324	2,586	2,179	2,035	2,586
Panel sample (recontacted disabled girls only; no weighting)	**% correct**	+4	+6	-1	0	-2	+2	-10		-2	+1	+6**	+4**	+2**	+3*
	P-value	0.230	0.184	0.835	0.823	0.196	0.432	0.220		0.397	0.560	0.001	0.001	0.005	0.056
	N (per round)	238	381	677	577	198	666	219		705	709	2,297	1,916	1,872	2,350

*Note: 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 increase in SeGMA is explained by improvements in all the subtasks, with largest increase (+6 p.p.) in multiplication, division, fractions, and geometry. The low and statistically insignificant increase in EGMA can be explained by a mixture of improvements in relatively easier subtasks including number identification, quantity discrimination, or subtraction level 1, unimprovement in subtraction level 2 and word problems relative to girls in the non-treated group.

The overall conclusion is therefore that the interventions supported disabled GEC-T girls in learning improvement over and above the non-treated group across the majority of subtasks and overall scores, although with a relatively small magnitude.

We next look at the GEC-T treatment girls only and analyse how disabled girls compare to non-disabled girls to determine whether any inequality in learning levels existed in baseline, and if so, whether it widens or shrinks over time.

Table 23: GEC-T literacy learning levels of treatment girls in baseline

Baseline; Disabled	EGRA						SeGRA			
Learning levels in percentage points	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
Non-disabled [mean]	40	58	47	62	39	49	50	28	20	33
Disabled [mean]	33	59	46	54	32	44	45	23	14	27
Difference	-8**	+1	-2	-8**	-7**	-5**	-5**	-5**	-6**	-6**
T-statistics	-4.33	0.53	-1.15	-5.93	-5.10	-4.63	-8.42	-9.05	-11.39	-13.37
P-value	<0.001	0.600	0.250	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001

*Note: Difference 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.

Table 24: GEC-T numeracy learning levels of treatment girls at baseline

Baseline; Disabled	EGMA									SeGMA			
Learning levels in percentage points	Num. ident.	Quantity Discrim.	Missing Num.	Addition Level 1	Addition Level 2	Subtract. Level 1	Subtract. Level 2	Word Prob.	Overall EGMA	Mult./divis, fractions, geometry	Algebra	Sophist. Word problems	Overall SeGMA
Non-disabled [mean]	77	80	54	65	73	54	61	50	62	34	15	6	21
Disabled [mean]	75	75	46	67	69	56	62	47	59	27	10	4	14
Difference	-3*	-5**	-8**	+2*	-3	+1	0	-3**	-3**	-7**	-5**	-2**	-7**
T-statistics	-1.80	-3.59	-6.06	+1.87	-1.54	0.94	+0.14	-2.38	-3.38	-12.04	-9.63	-7.28	-15.56
P-value	0.070	<0.001	<0.001	0.060	0.120	0.350	0.890	0.020	<0.001	<0.001	<0.001	<0.001	<0.001

*Note: Difference 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 learning levels of disabled GEC-T girls were significantly lower in baseline compared to non-disabled GEC-T girls. The learning gap between disabled and non-disabled girls is most apparent in letter sound identification and oral reading fluency (8 p.p.), and reading comprehension (7 p.p.) in literacy; and missing number (8 p.p.), multiplication, and division, fractions, geometry (7 p.p.) in numeracy. Overall, the gap between disabled and non-disabled GEC-T girls in baseline was slightly wider in numeracy (7p.p., SeGMA) than literacy (6p.p., SeGRA).

We then check to see if the average disabled GEC-T girl has learned more or less than the average non-disabled GEC-T girl between baseline and midline. This allows us to see if the learning gap between disabled and non-disabled girls widened or narrowed between baseline and midline.

Despite the lower start, we find that the average disabled GEC-T girl improved at the same pace as the average non-disabled GEC-T girl – while some of the coefficients are positive, none of them is statistically significant at the 90% confidence level.

Moreover, when panel sample of GEC-T girls is considered (with no weighting), the EGRA literacy coefficient is of greater magnitude, while those for numeracy remain at zero, although none are statistically significant. This result may be interpreted as a positive, as it suggests that the GEC-T disabled girls are not falling behind non-disabled girls over time. This is shown in Table 25 below.

Table 25: Learning improvements of the average disabled GEC-T girl over and above non-disabled GEC-T girls

Difference-in-difference in percentage points		Literacy		Numeracy	
		EGRA	SeGRA	EGMA	SeGMA
Cross-sectional sample (all treatment girls; no weighting)	% correct score	+1	+1	0	+1
	P-value	0.804	0.663	0.984	0.407
Cross-sectional sample (all treatment girls; project-equal level weighting)	% correct score	+2	+2	-1	0
	P-value	0.516	0.373	0.671	0.978
Panel sample (recontacted treatment girls only; no weighting)	% correct score	+3	+1	0	0
	P-value	0.169	0.564	0.755	0.825
Panel sample (recontacted treatment girls only; project-equal level weighting)	% correct score	+6	+2	0	-2
	P-value	0.157	0.419	0.938	0.633

*Note: 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.

Once again, the aggregate learning scores at the test level hide disparities in the way disabled GEC-T girls have improved (or not) in specific subtasks relative to non-disabled GEC-T girls. Table 26 breaks down disabled girls' increases over and above the non-treated group across all literacy subtasks.

Table 26: Literacy improvements of disabled GEC-T girls over and above non-disabled GEC-T girls - % correct scores

Difference-in-difference in percentage points		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
Cross-sectional sample (all treatment girls; no weighting)	**% correct**	-2	-1	0	+4	0	+1	-1	-1	+4**	+1
	P-value	0.607	0.765	0.938	0.549	0.979	0.804	0.571	0.583	0.003	0.663
	N (BL)	7,404	11,295	9.560	10,253	10,554	12,148	18,339	16,446	15,160	18,735
	N (ML)	7,580	13,164	11,150	14,683	13,697	15,224	20,897	19,478	16,893	20,899
Panel sample (recontacted treatment girls only; no weighting)	**% correct**	0	+3	-1	+5	+3	+3	+0	+0	+1	+1
	P-value	0.962	0.322	0.572	0.165	0.253	0.169	0.939	0.933	0.809	0.564
	N (per round)	5,490	7,935	6,841	6,997	6,988	8,400	9,490	7,867	7,298	9,743

*Note: 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 overall statistically insignificant literacy changes of disabled GEC-T girls over and above non-disabled GEC-T girls reported in Table 25 are therefore mostly a result of a mixture of increases of greater and lower magnitude across various literacy subtasks. While there is no statistically significant difference in literacy subtasks between disabled and non-disabled girls, there is a 4 p.p. increase in short essay construction. None of the results are statistically significant for the sample of recontacted girls.

Table 27: Numeracy improvements of disabled GEC-T girls over and above non-disabled GEC-T girls - % correct scores

Baseline, disabled		EGMA									SeGMA			
		Num. Ident.	Quantity Discrim.	Missing Num.	Addition Level 1	Addition Level 2	Subtract. Level 1	Subtract. Level 2	Word Prob.	Overall EGMA	Mult./divis, fractions, geometry	Algebra	Sophist. Word problems	Overall SeGMA
Cross-sectional sample (all treatment girls; no weighting)	**% correct**	+3	+2	+4	+2	-3	0	-6	-3	0	+2	0	+1	+1
	P-value	0.486	0.724	0.351	0.669	0.633	0.998	0.142	0.236	0.984	0.524	0.959	0.472	0.407
	N (BL)	5,207	6,202	11,753	9,091	3,721	10,299	5,325	12,138	12,156	22,202	16,138	13,436	22,733
	N (ML)	5,730	6,606	13,258	11,019	6,066	11,997	7,721	14,130	14,130	24,585	18,591	15,180	24,585
Panel sample (recontacted treatment girls only; no weighting)	**% correct**	+1	+3*	0	-1	+4*	-2	-3	-1	0	+1	-2*	-3**	0
	P-value	0.783	0.073	0.861	0.579	0.083	0.194	0.513	0.752	0.755	0.365	0.090	0.004	0.825
	N (per round)	3,725	4,563	8,072	6,303	2,136	7,244	3,117	8,355	8,371	12,100	7,746	6,838	12,426

*Note: 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.

A similar finding holds for numeracy, where a mixture of positive and negative coefficients for subtask changes in learning of disabled GEC-T girls relative to non-disabled girls appears. For CS sample, a negative coefficient for subtraction level 2 stems out, due to a decrease in learning levels of disabled relative to non-disabled GEC-T girls. The remaining coefficients for the CS GEC-T girls are not statistically significant. The sample of recontacted girls reveals statistically significant increases of disabled relative to non-disabled GEC-T girls in lower numeracy levels for quantity discrimination (+3 p.p.) and addition level 2 (+4 p.p.). While disabled GEC-T girls also improved in higher level numeracy subtasks including algebra and sophisticated word problems between baseline and midline, they did so at a slower pace relative to non-disabled girls, resulting in negative and statistically significant DID coefficients of –2 p.p. and –3p.p. for the two subtasks, respectively.

This suggests that the evolution of learning does not follow a simple pattern, with performance differing across the tasks. However, on average, between baseline and midline, the learning levels of both disabled and non-disabled GEC-T girls improved, and none of the aggregate learning scores (and only a few of sub-categories between disabled and non-disabled treatment girls) show statistically significant differences.

As such, it seems suggestive that the interventions seem to have managed to support disabled girls' learning at a similar pace relative to non-disabled girls, preventing disabled GEC-T girls from falling further behind the non-disabled girls.

LNGB

The LNGB projects generally only include treatment girls, with learning data for disabled girls from a non-treated group for projects with two evaluation rounds only available for Pin Nepal. As such, the difference-in-difference model cannot be used.

Instead, we report the baseline levels of all the disabled girls, as well as the simple difference for treatment disabled girls between baseline and follow-up.41 As the number of disabled girls who are recontacted are extremely low, we refrained from conducting a panel analysis (including only recontacted girls). Instead, the analysis focuses on a comparison of baseline and follow-up learning levels of cross-sectional samples (comparing average scores of girls at baseline and at midline/endline without distinguishing those who were recontacted). Another difference between the GEC-T and LNGB is that there are no secondary level assessments for literacy and numeracy in the latter portfolio of projects. As such, any reported results focus on literacy (EGRA) and numeracy (EGMA) at the lower proficiency level.

The baseline learning levels of treatment disabled LNGB girls are very low for both literacy and numeracy, with average treatment disabled girl having achieved 18 p.p. overall score in literacy and 33 in numeracy. An average treatment disabled LNGB girl did not achieve more than 50 p.p. in any of the literacy or numeracy subtasks in baseline. However, disabled LNGB girls' scores significantly improved between baseline and follow-up across all subtasks in both literacy and numeracy.

They did so more in overall literacy compared to numeracy which they scored lower in baseline to begin with, by 34 p.p. and 31 p.p., respectively. A similar result of a comparable magnitude is found when using the recontacted sample of girls – but this restricts the portfolio analysis to three projects, with about 80% of the sample being from one project (IRC Sierra Leone).

41 The term 'follow-up' is used here as some projects conducted baseline-midline, while others baseline-endline. IRC Sierra Leone cohort 1 is the only project which conducted all three evaluation rounds. The baseline to midline comparison is made for this project.

Table 28: LNGB literacy learning levels of treatment disabled girls

Treatment; Disabled	EGRA								
Learning levels in percentage points	Letter sound ident.	Familiar word	Oral reading fluency	Reading comp.	Listening comp.	Letter knowledge	Writing, dictation	Letter cluster identification	Overall EGRA
Baseline [mean]	25	13	7	8	40			3	18
Follow-up [mean]	72	45	34	49	85			28	52
Difference	+46**	+32**	+27**	+41**	+45**			+25**	+34**
T-statistics	+15.00	+15.96	+17.25	+17.95	+16.31			+4.87	+20.94
P-value	<0.001	<0.001	<0.001	<0.001	<0.001			<0.001	<0.001
No. of observations (BL)	435	568	556	540	385			31	576
No. of observations (Follow-up)	195	487	455	209	302			21	487

*Note: Difference coefficients with two asterisks are statistically significant at the 95% confidence level (P-value lower than 0.05 = 5%). The actual P-value is reported in the table. Results for 'letter knowledge' and 'writing, dictation' are not shown due to low number of observations (14 in baseline).

Table 29: LNGB numeracy learning levels of treatment disabled girls

Baseline; Disabled	EGMA											
Learning levels in percentage points	Num. ident.	Quantity Discrim.	Missing Num.	Addition Level 1	Addition Level 2	Subtract. Level 1	Subtract. Level 2	Word Prob.	Multiplication	Division	Fractions	Overall EGMA
Baseline [mean]	36	41	14	31	29	24	23	21	15	9	10	33
Follow-up [mean]	68	76	45	68	69	60	58	64	37	28	30	64
Difference	+32**	+35**	+31**	+37**	+40**	+36**	+35**	+43**	+22**	+19**	+20**	+31**
T-statistics	+15.02	+15.33	+11.88	+16.54	+6.72	+15.73	+5.9	+15.11	+6.01	+6.42	+4.90	+17.32
P-value	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
No. of observations (BL)	583	446	284	577	44	577	44	293	162	162	131	593
No. of observations (Follow-up)	492	407	250	476	331	477	331	476	90	91	69	493

*Note: Difference 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.

We then look at how disabled girls compare to non-disabled girls to determine whether any inequality in LNGB learning levels also exists in baseline, and if so, whether any improvement is made over time. Similar to findings from GEC-T, non-disabled girls have higher baseline levels in overall literacy and numeracy and in virtually all subtasks.

Table 30: LNGB literacy learning levels of treatment girls in baseline

Baseline; Disabled	EGRA								
Learning levels in percentage points	Letter sound ident.	Familiar word	Oral reading fluency	Reading comp.	Listening comp.	Letter knowledge	Writing, dictation	Letter cluster identification	Overall EGRA
Non-disabled [mean]	43	24	15	23	41			9	30
Disabled [mean]	25	13	7	8	40			3	18
Difference	-17**	-11**	-9**	-14**	-1			-6*	-12**
T-statistics	-8.52	-7.65	-7.78	-9.05	-0.25			-1.81	-9.54
P-value	<0.001	<0.001	<0.001	<0.001	0.800			0.070	<0.001

*Note: Difference 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. Results for 'letter knowledge' and 'writing, dictation' are not shown due to low number of observations (14 in baseline).

Table 31: LNGB numeracy learning levels of treatment girls in baseline

Baseline; Disabled	EGMA											
Learning levels in percentage points	Num. ident.	Quantity Discrim.	Missing Num.	Addition Level 1	Addition Level 2	Subtract. Level 1	Subtract. Level 2	Word Prob.	Multiplication	Division	Fractions	Overall EGMA
Non-disabled [mean]	54	60	24	53	52	46	44	55	14	10	22	48
Disabled [mean]	36	41	14	31	29	24	23	21	15	9	10	33
Difference	-18**	-19**	-10**	-23**	-23**	-22**	-21**	-34**	1	-1	-12**	-15**
T-statistics	-10.53	-9.75	-5.13	-12.61	-3.53	-11.96	-3.35	-12.60	0.53	-0.63	-3.57	-11.15
P-value	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	0.590	0.530	<0.001	<0.001

* Note: Difference 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.

We then check to see if disabled girls had learned at a similar pace to non-disabled girls, to see if the gap widens over time. The analysis here focuses on the cross-sectional sample. Positively, we find that disabled LNGB girls have improved at a greater pace than non-disabled girls, although the difference is statistically significant only for numeracy. Similar results are obtained when weights are accounted for, with the magnitude being lower by 4 p.p. The results are encouraging, but it is important to be cautious when interpreting this as the results are based on a subsample of disabled girls with both baseline and a follow-up. When weights are accounted for, the differences have become insignificant suggesting this is driven by a few high performing projects.

Table 32: Learning improvements of the average disabled LNGB girl over and above non-disabled LNGB girls

Difference-in-difference in percentage points		Literacy – EGRA	Numeracy – EGMA
Cross-sectional sample (all treatment girls; no weighting)	% correct score	+11	+16**
	P-value	0.105	0.011
Cross-sectional sample (all treatment girls; project-equal level weighting)	% correct score	+7	+14**
	P-value	0.258	0.009

*Note: 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.

Looking at the subtasks, in literacy (table 33), LNGB disabled girls have shown the most progress in letter sound identification, followed by letter cluster identification. In numeracy (table 34), the improvement is highest in word problems, followed by addition level 1, and subtraction level 1. In these categories, disabled girls started at a much lower learning level from non-disabled girls in baseline but managed to show significant improvement in the follow-up above and beyond non-disabled girls.

Table 33: Literacy improvements of disabled LNGB girls over and above non-disabled LNGB girls

Treatment; Disabled		EGRA								
		Letter sound ident.	Familiar word	Oral reading fluency	Reading comp.	Listening comp.	Letter knowledge	Writing, dictation	Letter cluster identification	Overall EGRA
Cross-sectional sample (all treatment girls; no weighting)	% correct	+36**	+9	+9	+16	+1			+14**	+11
	P-value	<0.001	0.303	0.219	0.209	0.885			<0.001	0.105
	N (BL)	2,564	4,850	4,364	2,183	2,698			400	4,941
	N (ML)	2,673	5,160	4,701	1,855	3,283			800	5,160

*Note: 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. Results for Writing/Dictation and Letter Cluster Identification were based on very small sample sizes (of 14 BL disabled girls), and results are not reported here.

Table 34: Numeracy improvements of disabled LNGB girls over and above non-disabled LNGB girls

Treatment; Disabled		EGMA											
		Num. ident.	Quantity Discrim.	Missing Num.	Addition Level 1	Addition Level 2	Subtract. Level 1	Subtract. Level 2	Word Prob.	Multiplication	Division	Fractions	Overall EGMA
Cross-sectional sample (all treatment girls; no weighting)	% correct	19**	19*	13**	23**	10**	22**	15**	24**	3**	-2	8**	16**
	P-value	0.021	0.056	0.002	0.026	0.000	0.018	0.000	0.011	0.000	0.197	0.000	0.011
	N (BL)	4,954	4,196	2,668	4,606	1,420	4,658	1,420	2,677	810	822	410	5,019
	N (ML)	5,153	4,444	2,804	4,851	1,161	4,949	1,161	2,804	1,207	1,240	407	5,251

*Note: Difference 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 results of regressions on learning outcomes from both the GEC-T and LNGB samples show some similarities. In both programmes, there is significant inequality in learning level at baseline. Disabled girls have started at a significantly lower learning level compared to non-disabled girls.

Both programmes show very encouraging results on learning trajectories of disabled girls. For GEC-T, where there is a non-treated group, we see increases compared to those who didn't receive support. In GEC-T, disabled and non-disabled girls learn at a similar pace, despite disabled girls starting at a lower level. In LNGB, disabled girls show large improvement in learning compared to baseline, and they were learning at a faster pace compared to non-disabled girls, ending up with higher literacy and numeracy scores in the follow-up period.

While the usual statistical caveats apply, given the small sample sizes, these results are suggestive of a positive impact on girls with disabilities.

4.2. Intermediate outcomes

Alongside the literacy and numeracy outcomes, we also looked at how intermediate outcomes have changed. For the GEC-T girls, we use the panel structure (treatment vs non-treated at baseline and midline) to estimate the difference-in-difference. We cannot replicate this for LNGB data as life skills data is available longitudinally for only 5 cohorts of 3 projects (SC Nepal, VSO, and WEI42).

Many of the intermediate outcomes (IOs) measured by projects (in Table 35) show positive changes between baseline and midline43, except for three of them: girls' not attending to school in the most recent school year (reported by the primary caregiver), girls' perception of doing well on test because of luck, and perception of unsafe travel to school (reported by the primary caregiver) show negative changes. It is important to note that the intermediate outcomes are presented as undesirable situations – e.g., feeling nervous when reading in front of others. Therefore, in Table 35, a positive coefficient should be understood as a negative change as it shows an increase in the number of girls reporting undesirable outcomes; and inversely, negative coefficients should be understood as a positive change.

In most cases, these significant differences are driven primarily by changes for girls with no disabilities. In slightly more than half of the IOs that have significant changes, disabled girls show no change in these variables from baseline to midline.

Teaching quality improved in a number of domains, but these differ for girls with and without disability.

Some aspects, around most gender-specific questions, adapting languages, interactive teaching, and teacher punishing students who get things wrong showed positive improvements for non-disabled children, but none for disabled children.

42 VSO and WEI are excluded from the previous descriptive and learning analyses as they do not have girl-reported disability. SC Nepal also has very low prevalence of disability.
43 The coefficients changed from study 3 as we made some changes outlined in the methodology section. Nonetheless, the statistical significance and magnitude remain roughly the same.

Disabled girls reported improvements in four domains of teaching quality: Treating boys and girls differently in class, Teacher's absences, Teacher used physical punishment on other students in the past week, and Teacher used physical punishment on girls in the past week. The magnitude of changes for these variables are higher for disabled girls compared to non-disabled girls.

Disabled girls' caregivers reported less difficulty in affording school. Disabled girls also felt less nervous when doing Maths in front of others, felt that they were more focused on a goal, and felt they can organise peers to do an activity. Their perception of safety improved, with 9 p.p. less girls reporting feeling unsafe.

However, for many other domains there are no significant changes in IOs for girls with disability.

Table 35: Change in intermediate outcomes of disabled and non-disabled GEC-T girls

Difference-in-difference for panel sample (recontacted girls), in *percentage points*		Variable	DID coefficient (all)	Non-disabled		Disabled	
				DID coefficient (non-disabled)	Sample size	DID coefficient (disabled)	Sample size
Attendance		Did not attend school on most days since the start of the most recent school year	+2** (0.004)	+1* (0.05)	8,366	+6** (0.003)	934
Economic empowerment		Difficult to afford to go to school	-5** (<0.001)	-5** (<0.001)	8,386	-11** (0.002)	1,021
Life skills	Academic self-confidence	Nervous when reading in front of others	-5** (<0.001)	-5** (<0.001)	20,050	-4	2,326
		Nervous when doing maths in front of others	-5** (<0.001)	-5** (<0.001)	20,046	-6** (0.037)	2,318
		Not confident answering questions	0	0	20,647	+1	2,354
		Does not ask teacher if does not understand	+1	0	20,536	+2	2,344
		Does not do well on test because of hard work	0	+1	5,363	-1	829
		Does well on test because of luck	+3** (0.012)	+3** (0.01)	4,866	+1	765
		Cannot describe thoughts when speaking	0	0	20,646	-3	2,329
	Self-efficacy	Not focused on a goal	-1* (0.056)	-1	5,980	-4** (0.007)	931
		Cannot put a plan in place and stick with it	0	-1	4,344	0	741
	Motivation / goal orientation	Does not want to continue studying	0	-1	19,698	0	2,333
	Decision making	Does not recognise how choices made today affect future life	+1	+1	15,130	0	2,054
		Cannot work well in a group	-2** (0.001)	-2** (<0.001)	20,766	0	2,365

Difference-in-difference for panel sample (recontacted girls), in *percentage points*		Variable	DID coefficient (all)	Non-disabled		Disabled	
				DID coefficient (non-disabled)	Sample size	DID coefficient (disabled)	Sample size
	Leadership & communication	Cannot organise peers to do an activity	-4** (<0.001)	-4** (<0.001)	20,661	-5** (0.008)	2,348
	School-based loneliness	Feels lonely at school	-4** (0.008)	-5** (0.004)	4,351	0	773
Teaching quality	Welcoming environment	Not made to feel welcome in the classroom	0	0	17,476	0	2,191
	Gender-sensitive pedagogy	Boys and girls are treated differently in the class	-3** (0.002)	-2** (0.023)	16,065	-6** (0.021)	2,153
		Teacher asks more questions to one gender	-2** (<0.001)	-2** (<0.001)	15,785	-2	2,102
		Teacher asks harder questions to one gender	-2** (<0.001)	-2** (<0.001)	15,307	-2	2,009
	Attendance	Teacher often absent from the class	-7** (<0.001)	-7** (<0.001)	17,345	-8** (0.001)	2,191
	Child-centred practices	Teacher does not use a different language to help understand	-5** (<0.001)	-5** (<0.001)	16,351	-4	2,061
		Teacher does not suggest ways to continue study after school	-1	-1	15,843	-1	1,980
	Interactive learning	Teacher does not encourage students to participate	-1* (0.055)	-1* (0.067)	8,857	-2	470
	School corporal punishment	Teacher punishes students who get things wrong	-3** (<0.001)	-3** (0.001)	16,115	-3	2,085
		Teacher used physical punishment on other students in the past week	-11** (<0.001)	-11** (<0.001)	12,484	-13** (<0.001)	1,653
		Teacher used physical punishment on girl in the past week	-7** (<0.001)	-6** (<0.001)	10,578	-10** (0.001)	1,574
School related gender-based violence		Unsafe for girls to travel to schools (PCG)	+4** (<0.001)	+4** (<0.001)	8,983	+7* (0.051)	1,074
		Feels unsafe travelling to and from school	-8** (<0.001)	-7** (<0.001)	14,000	-9** (<0.001)	1,535
		Feels unsafe at school	0	0	16,876	-1	2,133

Difference-in-difference for panel sample (recontacted girls), in *percentage points*	Variable	DID coefficient (all)	Non-disabled		Disabled	
			DID coefficient (non-disabled)	Sample size	DID coefficient (disabled)	Sample size
Parental attitudes	Girls should not stay at school until tertiary education	-13** (<0.001)	-14** (<0.001)	7,215	-2	1,015
	Does not listen to views of girl about her education	-2	-2	5,900	-4	318
	Not worth investing in girl's education when funds are limited	0	0	7,671	0	950
	Girl not as likely to use her education as a boy	-2** (0.002)	-3** (0.002)	8,489	-1	976

Note: All regressions use girls' fixed effects and robust standard errors. The table differs slightly to the one presented in the main report as it includes all intermediate outcomes, regardless of the results' statistical significance.

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 non-treated girls), orange shows negative and statistically significant DID (GEC-T girls' intermediate outcomes worsened significantly more than non-treated girls).

5. Conclusion

Profiling

Prevalence

As we are limited to the data on the samples, we cannot say how many girls' overall in the GEC II had a disability. However, 8.6% of the sampled GEC-T girls and 14% of sampled LNGB girls reported having at least one type of disability. The higher proportion of LNGB disabled girls can be explained by the use of a wider scale to measure disability compared to GEC-T.

The number and share of girls with disability vary by projects. Projects with a focus on disability (CSU and LC) report higher proportion of disabled girls (43% and 36% respectively), while other GEC-T projects are mostly single figures. Some projects, particularly CAMFED, report a comparatively high share of girls with disability. For LNGB, about half the projects reported single figures, and others reported more than 10% of girls being disabled. Link C3 reported the highest prevalence of 24%.

Even though a girl is categorised as disabled when she answered "a lot of difficulty" or "cannot carry out the action at all" and "daily" for feelings of anxiety and depression, looking beyond the disability threshold can shed light on the type of functioning difficulty girls are most likely to have. The most commonly reported functioning difficulty is having difficulty remembering and concentrating (19% and 12% of GEC-T and LNGB girls reporting some level of difficulty). For LNGB, the wider scale shows significance of mental health problems of anxiety and depression, with 57% of girls experiencing these to a varying extent, and about 4.3-5.9% of girls experiencing them daily.

Characteristics of girls with disabilities

Disabled girls are more marginalised than non-disabled girls in terms of being poor and orphaned. In LNGB, they are also more likely to be married and mothers. In GEC-T, disabled girls are less likely to be overage (compared to grade level). Disabled girls have similar characteristics to non-disabled girls in terms of age, enrolment status, and grade repetition.

In LNGB, both caregivers of disabled and non-disabled girls have highlighted financial reasons as main reasons for girls being out of school (not enough money to pay for schooling, and girl has to work). These reasons are more prominent in caregivers of disabled girls. Reasons related to disability (such as health condition preventing enrolment) are also reported more frequently by caregivers of disabled girls.

Learning outcomes

In both GEC-T and LNGB, girls with disability started at a lower baseline learning level compared to non-disabled girls. This is uniform across literacy and numeracy, as well as across subtasks.

By midline, girls in GEC-T have made a similar level of progress in learning relative to non-disabled girls. This result is encouraging as it shows the projects have managed to improve learning outcomes for both disabled and non-disabled girls, despite disabled girls being more marginalised. Disabled girls who received treatment in GEC-T had also showed learning gains above and beyond disabled girls who did not receive treatment (non-treated group).

For LNGB, by midline (or endline), disabled girls have made improvements in literacy and numeracy above and beyond non-disabled girls (though results are only statistically significant in numeracy). The improvements are most prominent in Letter Sound Identification in EGRA and Quantity Discrimination in EGMA. The progress stems largely from disabled girls making greater improvement in learning outcomes compared to non-disabled girls.

Intermediate Outcomes

Many intermediate outcomes show positive changes from baseline to midline for GEC-T girls. These positive changes, however, were driven primarily by non-disabled girls. For disabled girls, most intermediate outcomes show no change. A few areas that show improvement include economic empowerment, being more focused on goals, feeling less nervous when doing Maths, greater ability to organise peers to do an activity, some teaching quality indicators, and feeling safe travelling to school.

6. References

Mitra, S. (2017) Disability, Health and Human Development. rep. Palgrave p. 73. https://link.springer.com/chapter/10.1057/978-1-137-53638-9_4

Outhred, R. et al. (2022) Who are the GEC-T supported girls? Baseline analysis of the socioeconomic and demographic characteristics of girls supported by GEC-T projects. rep. Oxford Policy Management and Oxford MeasurEd. https://girlseducationchallenge.org/media/nasektyb/opm-who-are-the-gect-supported-girls.pdf

UNESCAP. (n.d.). *Results of the Testing of the ESCAP/WG Extended Question Set on Disability*. United Nations Economic and Social Commission for Asia and the Pacific. https://www.washingtongroup-disability.com/fileadmin/uploads/wg/Documents/Results_of_the_Testing_of_the_ESCAP-WG_Extended_Question_Set_on_Disability.pdf

Appendix 1. Measurement of disability in the GEC-T and the LNGB

To assess functional difficulties in different domains, girls and PCGs were asked a set of questions taken from the Short Set and/or the Child Functioning Module (CFM), developed by the Washington Group (WG) on Disability Statistics. Girls reported on their own level of functional difficulty, while primary caregivers reported on the level of functional difficulty of the girl.

Table 36: Disability tools used

Windows	WG report	Tools used	For data analysis
GEC-T	Girl-reported	WG short set (6 cats)	Used 6 cats WG short set
	PCG-reported	WG short set (6 cats)	Not used in analysis
		CFM (13 cats)	
LNGB	Girl-reported	CFM (13 cats)	Used 13 cats CFM
	PCG-reported	WG short set[44] (6 cats)	Not used in analysis
		CFM (13 cats)	

WG Short Set

The questions in the WG Short Set assess functional difficulty across six domains including vision, hearing, mobility, remembering, self-care, and communication. For each domain, respondents rate the girl's level of difficulty carrying out an action on the following continuum: no difficulty at all; some difficulty; a lot of difficulty; or completely unable to carry out the action.

Table 37: Washington Group Short Set questions

Washington Group Short Set questions		
Vision	[Do/Does] [you/he/she] have difficulty seeing, even if wearing glasses? Would you say…	1. No difficulty 2. Some difficulty 3. A lot of difficulty 4. Cannot do at all 98. Refused 99. Don't know
Hearing	[Do/Does] [you/he/she] have difficulty hearing, even if using a hearing aid(s)? Would you say…	
Mobility	[Do/Does] [you/he/she] have difficulty walking or climbing steps? Would you say…	
Remembering	[Do/does] [you/he/she] have difficulty remembering or concentrating? Would you say…	
Self-care	[Do/does] [you/he/she] have difficulty with self-care, such as washing all over or dressing? Would you say…	
Communication	[your/his/her] usual language, [do/does] [you/he/she] have difficulty communicating, for example understanding or being understood? Would you say…	

WG CFM

CFM questions consist of the core six disability domains mentioned above, as well as other developmental disabilities domains including learning, concentrating, accepting change, controlling behaviour, making friends, anxiety, and

44 Pin Nepal Cohort 3 only

depression. The rating system remains the same except for anxiety and depression where the domains have one of six response categories: daily, weekly, monthly, a few times a year and never.

Table 38: CFM questions

CFM questions		
Vision	CF1. [Do/Does] [you/he/she] wear glasses?	1. Yes 2. No (Skip to CF3)
	CF2. When wearing [his/her] glasses, [do/does] [you/he/she] have difficulty seeing? Would you say…	1. No difficulty 2. Some difficulty 3. A lot of difficulty 4. Cannot do at all 98. Refused 99. Don't know (Skip to CF4)
	CF3. [Do/Does] [you/he/she] have difficulty seeing? Would you say…	
Hearing	CF4. [Do/Does] [you/he/she] use a hearing aid?	1. Yes 2. No (Skip to CF6)
	CF5. When using [his/her] hearing aid, [do/does] [you/he/she] have difficulty hearing sounds like people's voices or music? Would you say…	1. No difficulty 2. Some difficulty 3. A lot of difficulty 4. Cannot do at all 98. Refused 99. Don't know (Skip to CF7)
	CF6. [Do/Does] [you/he/she] have difficulty hearing sounds like people's voices or music? Would you say…	
Mobility	CF7. [Do/does] [you/he/she] use any equipment or receive assistance for walking?	1. Yes 2. No (Skip to CF12)
	CF8. Without [your/his/her] equipment or assistance, [do/does] [you/he/she] have difficulty walking 100 yards/meters on level ground? That would be about the length of 1 football field. Would you say…	1. No difficulty 2. Some difficulty 3. A lot of difficulty (Skip to CF10) 4. Cannot do at all (Skip to CF10) 98. Refused 99. Don't know
	CF9. Without [your/his/her] equipment or assistance, [do/does] [you/he/she] have difficulty walking 500 yards/meters on level ground? That would be about the length of 5 football fields. Would you say…	1. No difficulty 2. Some difficulty 3. A lot of difficulty 4. Cannot do at all 98. Refused 99. Don't know
	CF10. With [your/his/her] equipment or assistance, [do/does] [you/he/she] have difficulty walking 100 yards/meters on level ground? That would be about the length of 1 football field. Would you say…	1. No difficulty 2. Some difficulty 3. A lot of difficulty (Skip to CF14) 4. Cannot do at all (Skip to CF14) 98. Refused 99. Don't know

CFM questions		
	CF11. Without [your/his/her] equipment or assistance, [do/does] [you/he/she] have difficulty walking 500 yards/meters on level ground? That would be about the length of 5 football field. Would you say…	1. No difficulty 2. Some difficulty 3. A lot of difficulty 4. Cannot do at all 98. Refused 99. Don't know
	CF12. Compared with children of the same age, [do/does] [you/he/she] have difficulty walking 100 yards/meters on level ground? That would be about the length of 1 football field. Would you say…	1. No difficulty 2. Some difficulty 3. A lot of difficulty (Skip to CF14) 4. Cannot do at all (Skip to CF14) 98. Refused 99. Don't know
	CF13. Compared with children of the same age, [do/does] [you/he/she] have difficulty walking 500 yards/meters on level ground? That would be about the length of 5 football fields. Would you say…	1. No difficulty 2. Some difficulty 3. A lot of difficulty 4. Cannot do at all 98. Refused 99. Don't know
Self-care	CF14. [Do/does] [you/he/she] have difficulty with self-care such as feeding or dressing him/herself? Would you say…	1. No difficulty 2. Some difficulty 3. A lot of difficulty 4. Cannot do at all 98. Refused 99. Don't know
Communication	CF15. When [you/he/she] speak(s), [do/does] [you/he/she] have difficulty being understood by people inside of this household? Would you say…	
	CF16. When [you/he/she] speak(s), [do/does] [you/he/she] have difficulty being understood by people outside of this household? Would you say…	
Learning	CF17. Compared with children of the same age, [do/does] [you/he/she] have difficulty learning things? Would you say…	
Remembering	CF18. Compared with children of the same age, [do/does] [you/he/she] have difficulty remembering things? Would you say…	
Concentrating	CF19. [Do/does] [you/he/she] have difficulty concentrating on an activity that he/she enjoys doing? Would you say…	
Accepting change	CF20. [Do/does] [you/he/she] have difficulty accepting changes in his/her routine? Would you say…	
Controlling behaviour	CF21. Compared with children of the same age, [Do/does] [you/he/she] have difficulty controlling [your/his/her] behaviour? Would you say…	
Making Friends	CF22. [Do/does] [you/he/she] have difficulty making friends? Would you say…	
Anxiety	CF23. How often [do/does] [you/he/she] seem very anxious, nervous or worried? Would you say…	1. Daily 2. Weekly 3. Monthly 4. A few times a year 5. Never
Depression	CF24. How often [do/does] [you/he/she] seem very sad or depressed? Would you say…	

CFM questions		
		7. Refused 9. Don't know

Appendix 2. Project-level availability of tools

GEC-T

As shown in the table below, all projects have collected disability information by using different tools at baseline. Varkey is the only project whose available datasets did not include such data. Out of 27 projects, 23 asked WG short-set questions via girls' surveys. The majority of projects asked PCGs either the WG-SS (17 projects) or CFM (10 projects), although four projects collected both.

Table 39: Availability of disability tools in GEC-T projects at baseline

GEC-T		WG short set		CFM
Project	Country	*Girls*	*PCGs*	*PCGs*
AKF	Afghanistan	Yes	Yes	No(1)
AVANTI	Kenya	No(2)	Yes	No
BRAC	Afghanistan	Yes	Yes	No(1)
Camfed	Tanzania	Yes	Yes	No
	Zambia	No	Yes	Yes
	Zimbabwe	Yes	Yes	No
Camfed (ex-BRAC)	Tanzania	Yes	Yes	No
CARE International	Somalia	Yes	Yes	Yes
Cheshire Services	Uganda	Yes	Yes	No
Childhope	Ethiopia	Yes	No(2)	No(2)
DLA	Ghana	Yes	Yes	No
	Kenya	Yes	Yes	No
	Nigeria	Yes	Yes	No
EDT	Kenya	Yes	No	Yes
HPA	Rwanda	Yes	Yes	No
I Choose Life	Kenya	Yes	Yes	No
Leonard Cheshire	Kenya	Yes	No	Yes
LINK	Ethiopia	Yes	No	Yes
Mercy Corps Nepal	Nepal	Yes	Yes	No
Mercy Corps Nigeria	Nigeria	Yes	No	No
Opportunity(3)	Uganda	Yes	No	Yes
PLAN	Sierra Leone	Yes	Yes	Yes
PEAS	Uganda	Yes	No	No
Relief International	Somalia	Yes	Yes	No
STC DRC	DRC	No	No	Yes
STC MOZ	Mozambique	Yes	Yes	No
Varkey Foundation(4)	Ghana	No	No	No
Viva	Uganda	No	Yes	No
VSO	Nepal	Yes	Yes	No

GEC-T		WG short set		CFM
Project	Country	*Girls*	*PCGs*	*PCGs*
World Vision	Zimbabwe	Yes	Yes	Yes
WUSC	Kenya	Yes	No	Yes

Note: (1) Invalid data; (2) Based on the reviews of the tools, the project has collected such data but there is no respective variable in the dataset and it was not reported in the baseline report; (3) Midline data was used as a baseline due to the data quality issue with baseline data for Opportunity and (4) While caregiver data was reported in the baseline report, there is no disability variable in the dataset.

LNGB

Unlike GEC-T, LNGB projects asked the CFM questions to girls. PCGs were asked a mix of WG short set questions and CFM questions at baseline. Out of 30 cohorts with available data (within 12 projects), 23 cohorts used CFM for girls surveys and 18 for PCG/HH surveys. There is one cohort (Pin Nepal Cohort 3) which collected only WG short set via PCG/HH surveys. ACTED Cohort 2 L&N do not have baseline data.

Table 40: Availability of disability tools in LNGB projects at baseline

LNGB			*CFM*		*WG short set*
Project	Country	Cohort	*Girls*	*PCGs*	*PCGs*
ACTED	Pakistan	C1 L&N	Yes	Yes	No
		C2 L&N	No baseline	No baseline	No baseline
		C4 L&N	Yes	Yes	No
		C1 ALP	Yes	Yes	No
ActionAid	Kenya	C1	Yes	Yes	No
		C3	Yes	No	No
Care	Somalia	C1 ABE	Yes	Yes	No
		C1 NFE	Yes	Yes	No
		C1 Formal	Yes	Yes	No
		C4 NFE	Yes	Yes	No
IRC Pakistan	Pakistan	Learn	Yes	Yes	No
		Earn	Yes	Yes	No
		Distant	Yes	No	No
IRC SL	Sierra Leone	C1	Yes	No	No
Link	Malawi	C1	Yes	Yes	No
		C3	Yes	Yes	No
Pin Ethiopia	Ethiopia	C1	Yes	Yes	No
		C3	Yes	No	No
Pin Nepal	Nepal	C1	Yes	No(1)	No
		C2	Yes	No	No
		C3	No	No	Yes
Plan	Zimbabwe	GS	Yes	No	No
		HH	No	Yes	No
Street Child	Nepal	C1	Yes	No	No

LNGB			*CFM*		*WG short set*
Project	Country	Cohort	*Girls*	*PCGs*	*PCGs*
		C2	Yes	No	No
		C3	Yes	No	No
VSO	Nepal	C1	No	Yes	No
WEI	Ghana	C1 Formal	No	Yes	No
		C1 Nonformal	No	Yes	No
		C2 Nonformal	No	Yes	No

Note: (1) Based on the reviews of the tools, the project has collected such data but there is no respective variable in the dataset and it was not reported in the baseline report.

Appendix 3. Comparison of girl- and PCG-reported disability

This section utilises baseline data from projects with data on both girl-reported disability and PCG-reported disability. It compares the prevalence and characteristics of girls categorised as disabled based on girl-reported data and those categorised as disabled based on PCG-reported data. As a result of this sample restriction, the reported figures differ from prevalence reported in the main text.

Disability prevalence differs by girl-reported and PCG-reported disability. In GEC-T, girls reported twice more cases of disability than the primary caregivers. In LNGB, girls also reported being disabled in a similar prevalence to their PCGs' reports. When 13 domains are used, girls reported higher prevalence of disability compared to the caregiver. Nonetheless, the difference is smaller than GEC-T. Additionally, when 13 domains were used instead of 6, both girls and PCGs reported more disability. This implies that when fixing the disability definition to include girl-reported (rather than PCG-reported) disability, the analysis includes more disabled girls than if PCG's reports were used.

Figure 10: High level disability status of GEC-T baseline treatment girls

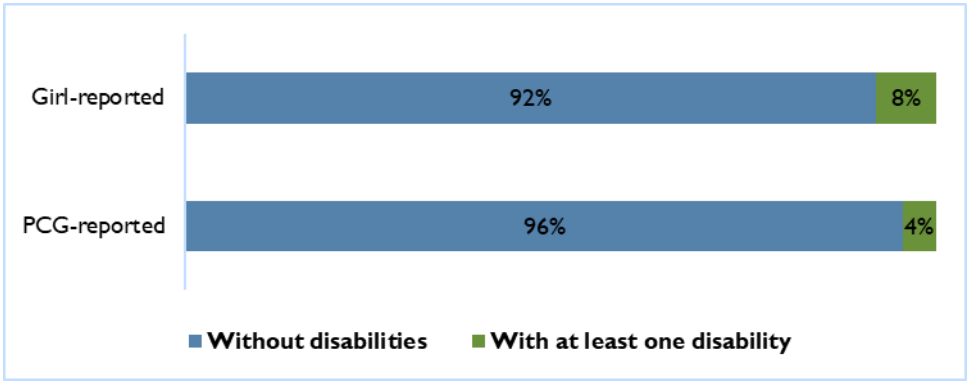
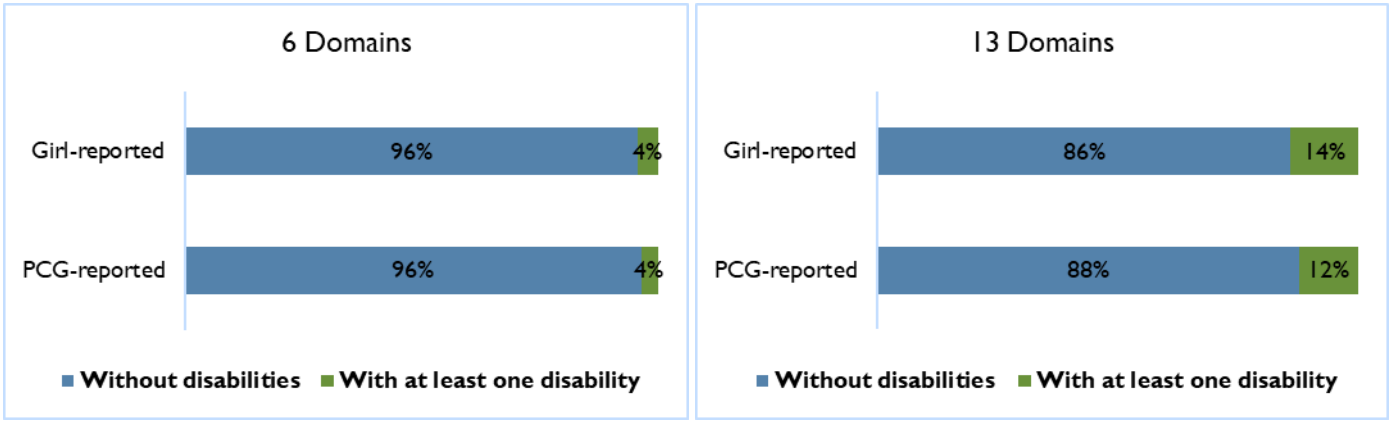
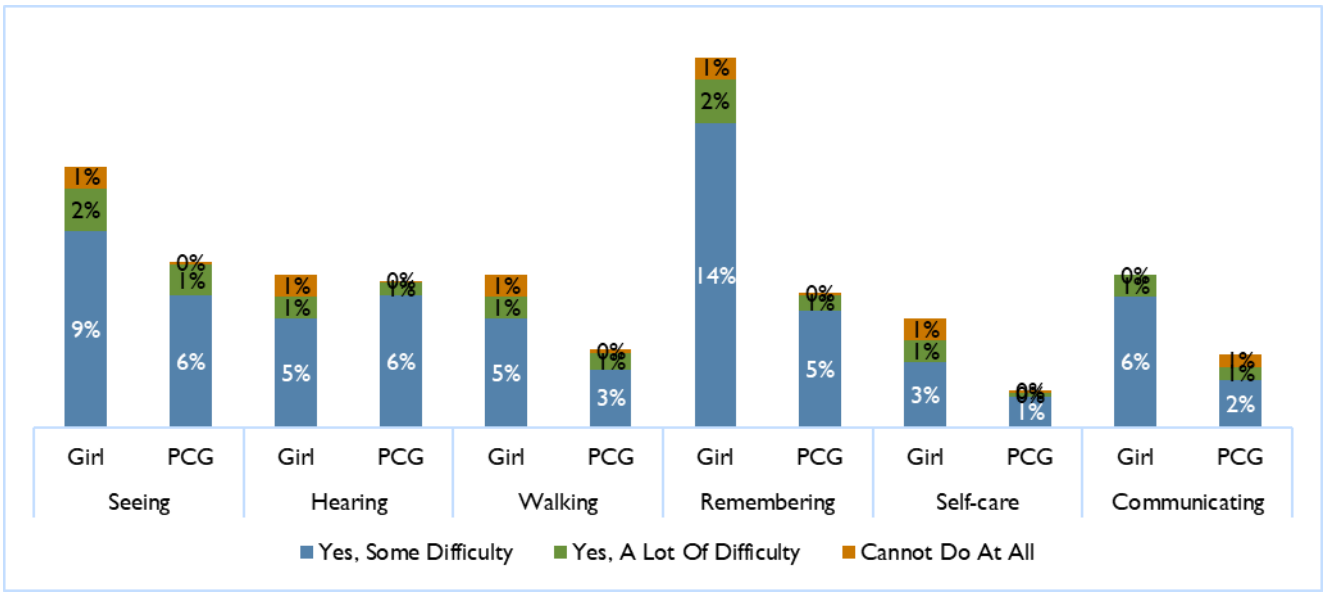


Figure 11: High level disability status of LNGB baseline treatment girls



The severity level for each domain differs between girl-reported and PCG-reported disability. Girls reported more instances of disability in all domains compared to the PCGs. Most notably, girls reported more instances of having some difficulty remembering and seeing than PCGs do. The agreement is highest in hearing disability. This further shows that the choice of using girl- or PCG-reported disability affects the number of disabled girls as well as the kinds of disability that are included.

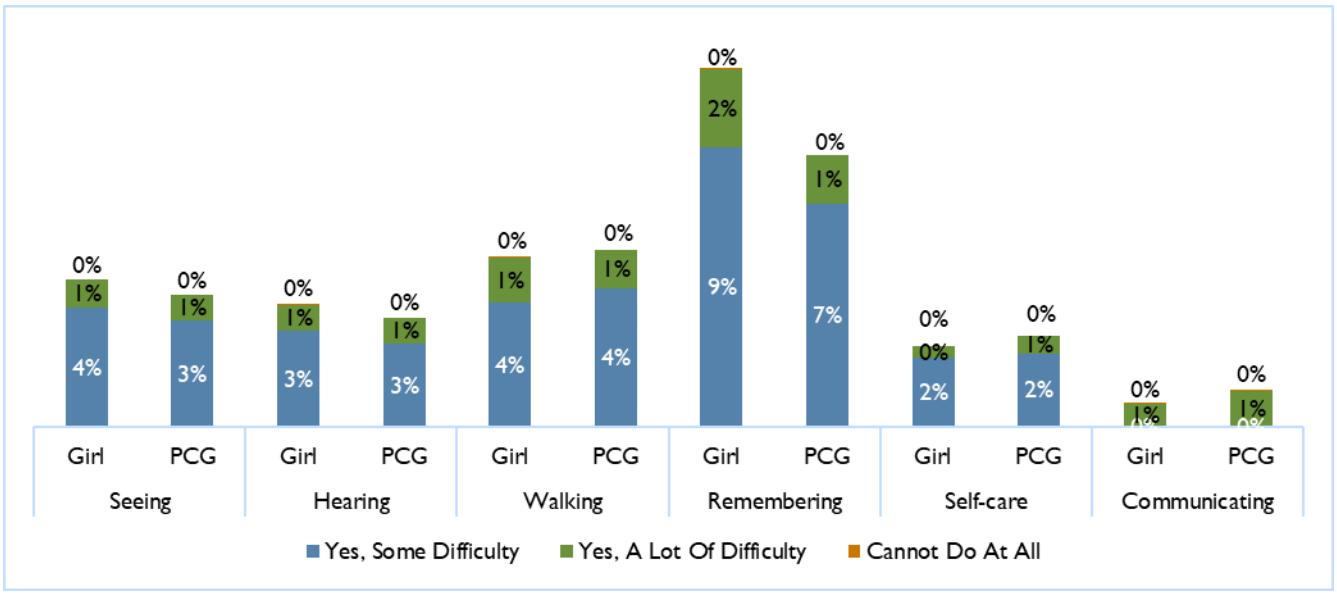
Figure 12: Disability status of GEC-T baseline treatment girls (Girl-reported compared with PCG-reported disability)



*Note: observations answering 'No, No Difficulty' were omitted.

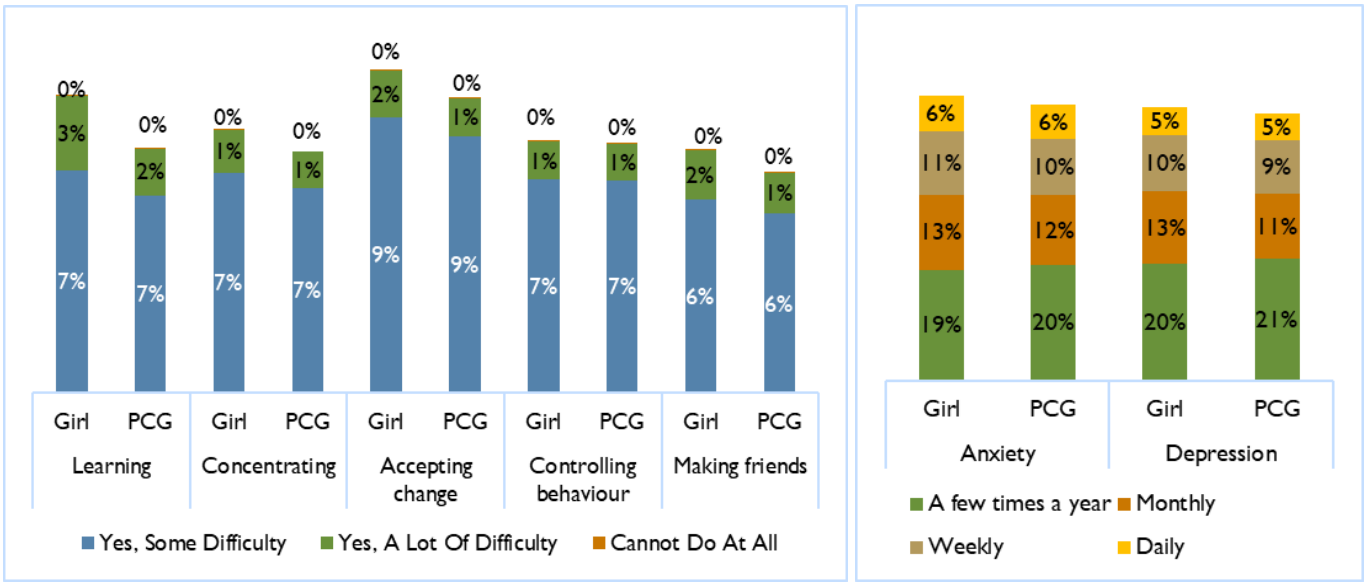
For LNGB, the differences by domain are less stark compared to GEC-T. PCG and girl show more agreement in reporting, with both agreeing the greatest difficulty for girls is in remembering (for 6 domains).

Figure 13: Disability status of LNGB baseline treatment girls – Six common domains (Girl-reported compared with PCG-reported disability)



*Note: observations answering 'No, No Difficulty' were omitted.

Figure 14: Disability status of LNGB baseline treatment girls– The other seven domains (Girl-reported compared with PCG-reported disability)



*Note: observations answering 'No, No Difficulty' and 'Never' were omitted.

Knowing that prevalence of disability differs when using girl and PCG-reported disability, the next step is to investigate whether the key sample characteristics differ for girls categorised as disabled based on girl-reported data and primary caregivers' report.

The table below shows the characteristics of GEC-T girls with disabilities, comparing disabled girls based on girl- and PCG-reported disability. Except for being poor, characteristics of disabled girls are similar across disabled girls based on the different definitions. Girls reported as being disabled by their caregivers are less likely to be poor compared to girls who girl-reported to be disabled.

Table 41: Characteristics of GEC-T girls with disabilities (Girl-reported compared with PCG-reported disability)

Characteristics	GEC-T girls with disabilities (Based on girl-reported disability)	GEC-T girls with disabilities (Based on PCG-reported disability)	GEC-T girls
Girl repeats a class	13%	14%	13%
Girl is overage	37%	35%	42%
Girl is married	1%	0%	2%
Girls is a mother	1%	1%	2%
Single orphan	21%	19%	14%
Double orphan	5%	5%	2%
Poor	38%	29%	24%

For LNGB, characteristics of disabled girls are similar for girls who reported being disabled and those disabled based on PCG's reports.

Table 42: Characteristics of LNGB girls with disabilities (Girl-reported compared with PCG-reported disability)

Characteristics	LNGB girls with disabilities (Based on girl-reported disability)	LNGB girls with disabilities (Based on PCG-reported disability)	LNGB girls
Girl is married	15%	13%	19%
Girls is a mother	22%	19%	30%
Single orphan	23%	21%	18%
Double orphan	1%	2%	2%
Poor	28%	27%	26%

Agreement

In projects where data is collected from both girls and PCG, there appears to be a high level of disagreement regarding which girls are disabled. For non-disabled girls, in most cases, girls and primary caregivers agree that the girl is not disabled (89% and 80% of total observation with available girl-reported and PCG-reported disability data for GEC-T and LNGB, respectively). However, in cases where the girl is reported to be disabled, there is large disagreement between the girl and primary caregiver's answers. In GEC-T, only 194 girls (12% of all reported disabled) are reported disabled by both girl themselves and the PCG. In 60% of cases, the girl reported being disabled while PCG reported them non-disabled. This discrepancy is reflected in all types of disability (but to a lesser extent in seeing and walking disability). In LNGB, the alignment is greater (32%), but remains high nonetheless. The agreement is lower in some types of disability including seeing, self-care, concentrating, and controlling behaviours. This implies that the sample of disabled girls varies significantly when girl-reported data is used instead of the PCG's.

Table 43: Comparison of girl and PCG answers in girl disability

	GEC-T			LNGB	
	PCG reported disabled	PCG reported non-disabled		PCG reported disabled	PCG reported non-disabled
Girl reported disabled	194	1,000	Girl reported disabled	369	445
Girl reported non-disabled	467	13,282	Girl reported non-disabled	345	4,656

Appendix 4: Robustness checks

The main measure of disability in this report is the girl-reported measure of functioning disability using the Washington Group sets of questions. As this can change over time, when panel analysis is used, we fix girls' disability status at baseline -so girls identified as having a disability at baseline are categorised as being disabled at midline, irrespective of if they report being disabled at midline or not.

While ideally, we would use these alternative measures of disability to check the robustness of our results doing so means reducing our sample noticeably, to the extent that the resulting sample sizes might be too small to be meaningful.

If we look to use only those girls who report being disabled at both baseline and midline, then of the 2,564 GEC-T girls who were identified as having a disability in baseline, only 431 were identified as having a disability at midline. For LNGB, of the 226 girls identified as having a disability at baseline, only 22 were identified as disabled at midline.

We can also look to include only those girls who report a disability in both the girls and PCG reports. If we do this and restrict the sample to (1) include girls who are identified as having a disability by both girls and PCG reports and (2) include girls that are identified as having a disability at both baseline and midline (girls reports) then we find a very small number of girls.

The results are reported below. Restricting our sample as in (1) yields a sample size that contains 100-162 girls and restricting as (2) yields a sample size that contains 151-386 girls.[45]

None of the estimates are statistically significant with these new samples, and the coefficients are inconsistent across tests. However, as the samples will not have sufficient power to identify any impacts, and the samples from each project are likely to be extremely small, we do not advise drawing any conclusions from this.

We did not conduct robustness checks for LNGB window as no non-treated group was used in those evaluations.

Table 44: Learning improvements of the average disabled GEC-T girl over and above non-treated girls (Full sample compared to various subsamples)

Difference-in-difference in percentage points		Literacy		Numeracy	
		EGRA	SeGRA	EGMA	SeGMA
Full panel sample (recontacted disabled girls only; no weighting)	% correct score	+4*	+3**	+1	+3*
	P-value	0.054	0.009	0.560	0.056
	Sample size (per round)	712	2,071	709	2,350
Subsample (PCG-reported disability) (recontacted disabled girls only; no weighting)	% correct score	-2	-3	0	-3
	P-value	0.406	0.364	0.966	0.453
	Sample size (per round)	109	142	100	162
Subsample (disability at baseline and midline) (recontacted disabled girls only; no weighting)	% correct score	+2	-2	+4	+1
	P-value	0.712	0.446	0.275	0.815
	Sample size (per round)	152	334	151	386

*Note: Difference 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.

45 When using the new measures, some projects have no observations, therefore they are dropped from the analysis altogether.

Appendix 5. Comparison against external evaluation reports

The table below summarises the proportions of GEC-T girls with disabilities based on girl-reported as per the WG-SS and flags the differences between what the evaluations team reported in baseline reports and the data we used for analysis.

Out of 27 projects, 16 projects and Camfed Tanzania and Zimbabwe reported girl-reported disability data in the baseline report. The estimates of girls with disabilities in the baseline report correspond to what we calculated with data, except for two projects (CSU and Childhope) showing greater than 10% of percentage difference. The difference for Childhope is due to their use of a broad definition of disability, which includes girls who stated "Yes, Some Difficulty". The reasons for discrepancies for CSU are due to the data quality issue.

Table 45: Discrepancies in proportion of GEC-T girls with disabilities

Project	Country	Data	Based on baseline report	
		Girls WG	Girls WG	Used data in baseline report
AKF	Afghanistan	1.2%	1.2%	Girls WG (Not specified in the report)
AVANTI	Kenya	N/A	N/A	Caregivers (3%)
BRAC	Afghanistan	2.5%	T1: 2% and T2: 2.9%	Girls WG (Not specified in the report)
CAMFED	Tanzania	17.3%	17.0%	Girls WG
	Zambia	N/A	N/A	Caregivers WG (4.5%)
	Zimbabwe	15.5%	15.0%	Girls WG
CAMFED (ex-BRAC)	Tanzania	17.5%	17.3%	Girls WG
CARE	Somalia	1.0%	N/A	Caregivers WG + CFM (6.9%)
Cheshire Services	Uganda	43.0%	50.0%	Girls WG
Childhope	Ethiopia	0.3%	6.1%	Girls WG
DLA	Ghana	10.3%	10.3%	Girls WG and Caregivers WG
	Kenya	5.7%	5.7%	Girls WG and Caregivers WG
	Nigeria	2.5%	2.5%	Girls WG and Caregivers WG
EDT	Kenya	7.9%	N/A	Caregivers (1%)
HPA	Rwanda	3.2%	3.2%	Girls WG (Not specified in the report)
I Choose Life	Kenya	3.8%	3.8%	Girls WG (Not specified in the report)
Leonard Cheshire	Kenya	35.6%	N/A	Caregivers CFM (39%)
LINK	Ethiopia	25.1%	23%	Girls WG and Girls CFM
MC Nepal	Nepal	0.1%	N/A – they only reported per disability category by IS/OOS[46] status	Girls WG
MC Nigeria	Nigeria	3.8%	ISG 4.8% and OOSG 2.8%	Girls WG

46 In-School and Out-of-School status

Project	Country	Data	Based on baseline report	
		Girls WG	Girls WG	Used data in baseline report
Opportunity	Uganda	2.2%	2%	Girls WG and Caregivers WG
PLAN	Sierra Leone	3.2%	JSS[47]: 3.3% and Primary 9.5%	Girls WG
PEAS	Uganda	2.6%	2.6%	Girls WG
Relief	Somalia	1.9%	1.9%	Girls WG
STC DRC	DRC	N/A	N/A	Caregivers CFM (34%)
STC MOZ	Mozambique	1.6%	N/A	Caregivers WG (1.89%)
Varkey	Ghana	N/A	N/A	Caregivers WG (7.5%)
Viva	Uganda	N/A	N/A	Caregivers WG (2%)
VSO	Nepal	2.5%	N/A	Combined Caregivers WG + Girls WG (3.4%)
World Vision	Zimbabwe	7.7%	N/A	Caregivers WG (10.1%)
WUSC	Kenya	2.4%	2.6% (treatment and non-treated)	Unclear

Note: It is marked in red if the percentage difference between the baseline report and the data is greater than 10%. The grey highlighted ones are the projects that did not report girl-reported data in the report. The baseline reports of the five projects highlighted in yellow presented the data separately only (e.g., ISG and OOSG separately) or treatment and non-treated together, so the direct comparison was not feasible.

Out of 30 LNGB cohorts, 14 cohorts defined a disability status by using CFM data collected from girls in the baseline report. Overall, the baseline report's estimates of girls with disabilities correspond to what we calculated using data. Some discrepancies were found in Acted C1 ALP and Pin Ethiopia C1. However, it cannot be clearly explained why the discrepancies occur based on the report and data.

Table 46: Discrepancies in proportion of LNGB girls with disabilities

Project	Country	Cohort	Data	Based on baseline report	
			Girls CFM	Girls CFM	Used data in baseline report
ACTED	Pakistan	C1 L&N	27%	28.3%	Girls CFM
		C2 L&N	N/A	N/A	Not collected
		C4 L&N	7%	7.7%	Girls CFM
		C1 ALP	14%	9.6%	Girls CFM
ActionAid	Kenya	C1	33%	36%	Girls CFM
		C3	17%	16.9%	Girls CFM
Care	Somalia	C1 ABE	4%	N/A	Caregivers CFM (11.9%)
		C1 NFE	3%		
		C1 Formal	3%		
		C4 NFE	11%		
IRC Pakistan	Pakistan	Learn	8%	N/A	Caregivers CFM (15.12%)
		Earn	10%		
		Distant	21%		
IRC SL	Sierra Leone	C1	15%	14.6%	Girls CFM

47 Junior Secondary Schools

Project	Country	Cohort	Data	Based on baseline report	
			Girls CFM	Girls CFM	Used data in baseline report
Link	Malawi	C1	34%	34.39%	Girls CFM
		C3	41%	40.3%	Girls CFM
Pin Ethiopia	Ethiopia	C1	15%	5.4%	Girls CFM
		C3	20%	Not in the report	Unclear
Pin Nepal	Nepal	C1	8%	Not in the report	Unclear
		C2	2%	Not in the report	Unclear
		C3	N/A	18.25%	Girls CFM
Plan	Zimbabwe	GS	28%	29.57%	Girls CFM
		HH	N/A	N/A	Caregivers CFM (27.42%)
Street Child	Nepal	C1	0%	0%	Girls CFM
		C2	0%	0%	Girls CFM
		C3	0%	0%	Girls CFM
VSO	Nepal	C1	N/A	N/A	Caregivers CFM (13.93%)
WEI	Ghana	C1 Formal	N/A	N/A	Caregivers CFM (13%)
		C1 Nonformal	N/A	N/A	Caregivers CFM (9.4%)
		C2 Nonformal	N/A	N/A	Caregivers CFM (8%)

Note: It is marked in red if the percentage difference between the baseline report and the data is greater than 10%. The grey highlighted ones are the projects that did not report girl-reported disability data in the report.

Appendix 6. Projects included in the analysis

The following tables show which projects were included and excluded for each analysis. Projects without girl-reported disability data are excluded from all analyses, as this study focuses on disability. Two projects (CAMFED Zambia and Viva) have PCG-reported disability, but not girl-reported disability. Due to reduction in sample sizes discussed in section 2, these projects are also excluded from our analysis. Other projects are excluded for panel analysis when they do not collect midline data (such as Link). Projects that present data quality issues (in midline) are excluded including Childhope, Plan, and Opportunity.48

Table 47: Included/ excluded GEC-T projects for Study 4 quantitative analysis

Project	Country	Descriptive	Learning		Intermediate outcomes
		Baseline data	*Cross-sectional*	*Panel*	*Panel*
AKF	Afghanistan	Yes	Yes	Yes	Yes
AVANTI	Kenya	No girl-reported disability data Has PCG-reported disability	No girl-reported disability data No midline	No girl-reported disability data No midline	No girl-reported disability data No midline
BRAC	Afghanistan	Yes	Yes	Yes	Yes
CAMFED	Tanzania	Yes	Yes	Yes	Yes
CAMFED	Zambia	No girl-reported disability data Has PCG-reported disability (but small n)	No girl-reported disability data Has PCG-reported disability (but small n)	No girl-reported disability data Has PCG-reported disability (but small n)	No girl-reported disability data Has PCG-reported disability (but small n)
CAMFED	Zimbabwe	Yes	Yes	Yes	Yes
CAMFED (ex-BRAC)	Tanzania	Yes	Yes	Yes	Yes
CARE	Somalia	Yes	Yes	Yes	Yes
Cheshire Services	Uganda	Yes	Yes	Yes	Yes
Childhope	Ethiopia	Yes	Baseline learning data invalid	Baseline learning data invalid	Baseline learning data invalid
DLA	Ghana	Yes	Yes	Yes	Yes
	Kenya	Yes	Yes	Yes	Yes
	Nigeria	Yes	Yes	Yes	Yes
EDT	Kenya	Yes	Yes	Yes	Yes
HPA	Rwanda	Yes	Yes	Yes	Yes

48 The baseline data from Opportunity International was shown to include "irregularities in unique IDs, data entry errors, difficulties to merge datasets." (Opportunity International UK's midline evaluation report. A new external evaluator was contracted at midline, and comparison between baseline and midline was deemed inconclusive "due to baseline data issues, changes to sample and tools." Therefore, midline data was used as baseline for descriptive analysis. Learning data from the three projects were deemed as invalid. Hence, learning analysis were not conducted with these three projects.

Project	Country	Descriptive	Learning		Intermediate outcomes
		Baseline data	*Cross-sectional*	*Panel*	*Panel*
I Choose Life	Kenya	Yes	Yes	Yes	Yes
Leonard Cheshire	Kenya	Yes	Yes	Yes	Yes
LINK	Ethiopia	Yes	No midline data	No midline data	No midline data
Mercy Corps Nepal	Nepal	Yes	Yes	Yes	Yes
Mercy Corps Nigeria	Nigeria	Yes	Yes	Yes	Yes
Opportunity	Uganda	Yes	Baseline learning data invalid	Baseline learning data invalid	Baseline learning data invalid
PLAN	Sierra Leone	Yes	Baseline learning data invalid	Baseline learning data invalid	Baseline learning data invalid
PEAS	Uganda	Yes	Yes	Yes	Yes
Relief International	Somalia	Yes	Yes	Yes	Yes
STC DRC[49]	DRC	No girl-reported disability data Has WG_CF data asked to PCG	No midline data	No midline data	No midline data
STC MOZ	Mozambique	Yes	No midline data	No midline data	No midline data
Varkey Foundation	Ghana	No disability data available (girl nor PCG)	No disability data available (girl nor PCG)	No disability data available (girl nor PCG) Unique ID data quality issues	No disability data available (girl nor PCG) Unique ID data quality issues
Viva	Uganda	No girl-reported disability data Has PCG-reported disability	No girl-reported disability data Has PCG-reported disability	No girl-reported disability data Has PCG-reported disability	No girl-reported disability data Has PCG-reported disability
VSO	Nepal	Yes	Yes	Unique ID data quality issues	Unique ID data quality issues
World Vision	Zimbabwe	Yes	Yes	Yes	Yes
WUSC	Kenya	Yes	Yes	Yes	Yes

49 STC DRC collects girl-reported disability based on 13-categories CFM scale, but not 6-categories WG scales as other projects. For consistency in reporting disability, STC DRC was excluded from all analyses.

Project	Country	Descriptive	Learning		Intermediate outcomes
		Baseline data	*Cross-sectional*	*Panel*	*Panel*
Number of projects (out of 31)		26	21	20	20
Number of projects (counting Camfed and DLA as single project – out of 27)		23	18	17	17

There are more restrictions to LNGB data, compared to GEC-T. Projects are generally excluded because they do not have valid follow-up data (midline or endline). Street Child is included in the descriptive analysis but not learning as it does not have disaggregated learning data.

Table 48: Included/ excluded LNGB projects for Study 4 quantitative analysis

Project	Country	Cohort	Descriptive	Learning
			Baseline data	*Cross-sectional*
ACTED	Pakistan	C1 L&N	Yes	Yes (as baseline to C2)
		C2 L&N	No baseline	Yes[50] (as endline to C1)
		C4 L&N	Yes	Yes (endline)
		C1 ALP	Yes	No follow-up
ActionAid	Kenya	C1	Yes	Yes (endline)
		C3	Yes	No follow-up
AKF	Afghanistan	C1	No data submitted at the time of writing	No data submitted at the time of writing
Care	Somalia	C1 ABE	Yes	Yes (midline)
		C1 NFE	Yes	Yes (midline)
		C1 Formal	Yes	Yes (midline)
		C4 NFE	Yes	No follow-up
IRC Pakistan	Pakistan	Learn	Yes	No follow-up
		Earn	Yes	No follow-up
		Distant	Yes	No follow-up
IRC SL	Sierra Leone	C1	Yes	Yes (midline and endline)
Link	Malawi	C1	Yes	No disability data available at follow-up (endline)
		C3	Yes	No follow-up
Pin Ethiopia	Ethiopia	C1	Yes	No follow-up
		C3	Yes	No follow-up
Pin Nepal	Nepal	C1	Yes	Yes (endline)
		C2	Yes	Yes (endline)
		C3	No disability data with 13 cats available	No follow-up No disability data with 13 cats available
Plan	Zimbabwe	GS	Yes	No follow-up No learning data

50 Due to programme design changes, ACTED C2 endline data is collected instead of C1 endline. This data is meant to be compared with C1 baseline.

Project	Country	Cohort	Descriptive	Learning
			Baseline data	*Cross-sectional*
		HH	Has PCG-reported disability	Has PCG-reported disability No follow-up No learning data
Population Council	Ethiopia	C1	No data submitted at the time of writing	No data submitted at the time of writing
Street Child	Nepal	C1	Yes	No detailed learning data available at baseline and follow-up (endline)
		C2	Yes	No detailed learning data available at baseline and follow-up (endline)
		C3	Yes	No detailed learning data available at baseline and follow-up (endline)
VSO	Nepal	C1	No girl-reported disability data Has PCG reported disability	Follow-up (midline) available No girl-reported disability data Has PCG-reported disability
WEI	Ghana	C1 Formal	No girl-reported disability data Has PCG-reported disability	Follow-up (midline) available No girl-reported disability data Has PCG-reported disability
		C1 Nonformal	No girl-reported disability data Has PCG-reported disability	No follow-up
		C2 Nonformal	No girl-reported disability data Has PCG-reported disability	No follow-up
Number of projects (out of 14)			10	5
Number of cohorts within 14 projects (out of 32)			23	10

Appendix 7. Attrition bias

Table 49 shows analysis of attrition bias for GEC-T. As systematic attrition (characteristics of girls lost in attrition differ from those recontacted) can bias estimates in change over time analysis, it is crucial that attrition bias is investigated. In the table, baseline level of literacy and numeracy are presented for disabled and non-disabled girls who were lost to attrition and those recontacted at midline, as well as for treatment and non-treated group. The table also highlights where the difference between groups is statistically different from zero.

In the first instance, when comparing between disabled girls in the treatment and the non-treated group, there is evidence of attrition bias. In the treatment group, disabled girls who were lost to attrition performed 4.75 p.p. worse in numeracy compared to disabled girls who were recontacted. In literacy, lost and recontacted girls performed at a similar level. This pattern is not reflected among the disabled girls who were in the non-treated group. In the non-treated group, disabled girls lost to attrition performed at about the same level in numeracy with those recontacted but performed 3.62 p.p. higher in literacy. This shows that there may be some evidence of attrition bias.

The pattern is more similar when comparing between disabled and non-disabled girls. Within the treatment group, non-disabled girls who were lost to attrition performed worse in numeracy than those recontacted (by 8.03 p.p.) while the score difference in literacy is not statistically different. This is the same pattern as with disabled girls discussed above. Within the non-treated group, non-disabled lost girls perform 2.29 p.p. higher than recontacted non-disabled girls, same pattern as disabled girls in the non-treated group. However, non-disabled girls who were recontacted also reported higher level of numeracy (results are not statistically significant for disabled girls).

Lastly, the pattern is also similar when comparing across disabled and non-disabled girls. In both treatment and non-treated groups, both lost and recontacted disabled girls performed worse in literacy and numeracy than lost and recontacted non-disabled girls.

Attrition bias analysis is conducted for GEC-T window only as we do not conduct a panel analysis for the LNGB window.

Table 49: Attrition bias overview between treatment and non-treated group and disabled and non-disabled girls (GEC-T)

		Overall literacy	Overall numeracy
Treatment	Disabled lost girls	32.28%	22.61%
	Disabled recontacted girls	31.55%	27.36%
	Non-disabled lost girls	40.03%	32.67%
	Non-disabled recontacted girls	40.19%	40.69%
	Difference between disabled lost and disabled recontacted girls	0.73%	-4.75%
	P-value	0.402	<0.001
	Difference between non-disabled lost and non-disabled recontacted girls	-0.16%	-8.03%
	P-value	0.623	<0.001
	Difference between disabled lost and non-disabled lost girls	-7.75%	-10.06%
	P-value	<0.001	<0.001
	Difference between disabled recontact and non-disabled recontacted girls	-8.65%	-13.34%
	P-value	<0.001	<0.001
Non-treated	Disabled lost girls	32.47%	20.96%
	Disabled recontacted girls	28.84%	21.68%
	Non-disabled lost girls	37.57%	28.40%

		Overall literacy	Overall numeracy
	Non-disabled recontacted girls	35.28%	35.62%
	Difference between disabled lost and disabled recontacted girls	3.62%	-0.72%
	P-value	<0.001	0.465
	Difference between non-disabled lost and non-disabled recontacted girls	2.29%	-7.22%
	P-value	<0.001	<0.001
	Difference between disabled lost and non-disabled lost girls	-5.11%	-7.44%
	P-value	<0.001	<0.001
	Difference between disabled recontact and non-disabled recontacted girls	-6.44%	-13.94%
	P-value	<0.001	<0.001
Comparison between treatment and non-treated group	Difference between treatment and non-treated disabled girls who were lost	-0.18%	1.65%
	P-value	0.859	0.115
	Difference between treatment and non-treated disabled girls who were recontacted	2.70%	5.68%
	P-value	0.001	<0.001
	Difference between treatment and non-treated disabled non-disabled girls who were lost	2.46%	4.27%
	P-value	<0.001	<0.001
	Difference between treatment and non-treated non-disabled girls who were recontacted	4.91%	5.08%
	P-value	<0.001	<0.001

Appendix 8. Profiling tables by project

The tables below show proportion of treatment girls with and without disability by project, as reported by girls. For GEC-T, the six domains were used. For LNGB, disability using both 6 and 13 domains is reported. The sample sizes and proportion of disabled girls vary by projects, as well as definitions of disability used (6 or 13 domains). Projects with a focus on disability (CSU and LC) report higher share of girls with disability in their samples.

Table 50: Proportion of girls with disabilities by project (GEC-T)

Project	Country	%	*Girl-reported disability*	
			Without disabilities	*With at least one disability*
AKF	Afghanistan	%	99%	1%
		No. of observations	1,409	17
AVANTI	Kenya	%		
		No. of observations	0	0
BRAC	Afghanistan	%	98%	2%
		No. of observations	948	24
CAMFED	Tanzania	%	83%	17%
		No. of observations	3,417	716
CAMFED	Zambia	%		
		No. of observations	0	0
Camfed	Zimbabwe	%	84%	16%
		No. of observations	2842	522
Camfed (ex-BRAC)	Tanzania	%	83%	17%
		No. of observations	1672	354
CARE International	Somalia	%	99%	1%
		No. of observations	490	5
Cheshire Services	Uganda	%	57%	43%
		No. of observations	151	114
Childhope	Ethiopia	%	100%	0%
		No. of observations	672	2
DLA	Ghana	%	90%	10%
		No. of observations	900	103
DLA	Kenya	%	94%	6%
		No. of observations	1,156	70
DLA	Nigeria	%	98%	2%
		No. of observations	1,112	28
EDT	Kenya	%	92%	8%
		No. of observations	4542	387
HPA	Rwanda	%	97%	3%
		No. of observations	422	14

Project	Country	%	*Girl-reported disability*	
			Without disabilities	*With at least one disability*
I Choose Life	Kenya	%	96%	4%
		No. of observations	1,730	68
Leonard Cheshire	Kenya	%	64%	36%
		No. of observations	206	114
LINK	Ethiopia	%	75%	25%
		No. of observations	403	135
Mercy Corps Nepal	Nepal	%	100%	0%
		No. of observations	749	1
Mercy Corps Nigeria	Nigeria	%	96%	4%
		No. of observations	888	35
Opportunity	Uganda	%	98%	2%
		No. of observations	571	13
PLAN	Sierra Leone	%	97%	3%
		No. of observations	701	23
PEAS	Uganda	%	97%	3%
		No. of observations	1,274	34
Relief International	Somalia	%	98%	2%
		No. of observations	1,579	30
STC DRC	DRC	%		
		No. of observations	0	0
STC MOZ	Mozambique	%	98%	2%
		No. of observations	658	11
Varkey Foundation	Ghana	%		
		No. of observations	0	0
Viva	Uganda	%		
		No. of observations	0	0
VSO	Nepal	%	98%	2%
		No. of observations	1,067	27
World Vision	Zimbabwe	%	92%	8%
		No. of observations	699	58
WUSC	Kenya	%	98%	2%
		No. of observations	846	21
Portfolio level		%	91%	9%
		No. of observations	31,104	2,926

Table 51: Proportion of girls with disabilities by project (LNGB) – 6 and 13 domains

Project	Country	Cohort	%	*Girl-reported disability (6 domains)*		*Girl-reported disability (13 domains)*	
				Without disabilities	*With at least one disability*	*Without disabilities*	*With at least one disability*
ACTED	Pakistan	C1 L&N	%	95%	5%	73%	27%
			No. of observations	218	12	167	63
		C2 L&N	%				
			No. of observations	0	0	0	0
		C1 ALP	%	99%	1%	93%	7%
			No. of observations	204	2	192	14
		C4 L&N	%	97%	3%	86%	14%
			No. of observations	423	13	375	61
ActionAid	Kenya	C1	%	87%	13%	67%	33%
			No. of observations	369	55	282	142
		C3	%	94%	6%	83%	17%
			No. of observations	482	30	426	86
Care	Somalia	ABE	%	100%	0%	96%	4%
			No. of observations	484	0	467	17
		NFE	%	99%	1%	97%	3%
			No. of observations	512	3	499	16
		Formal	%	99%	1%	97%	3%

Project	Country	Cohort	%	*Girl-reported disability (6 domains)*		*Girl-reported disability (13 domains)*	
				Without disabilities	*With at least one disability*	*Without disabilities*	*With at least one disability*
			No. of observations	417	4	410	11
IRC Pakistan	Pakistan	C4 NFE	%	96%	4%	89%	11%
			No. of observations	883	33	818	98
		Learn	%	99%	1%	92%	8%
			No. of observations	783	9	726	66
		Earn	%	98%	3%	90%	10%
			No. of observations	429	11	396	44
		Distant	%	97%	3%	79%	21%
			No. of observations	769	23	627	165
IRC SL	Sierra Leone	C1	%	96%	4%	85%	15%
			No. of observations	1,885	74	1,663	296
Link	Malawi	C1	%	83%	17%	66%	34%
			No. of observations	301	60	238	123
		C3	%	76%	24%	59%	41%
			No. of observations	195	63	152	106
Pin Ethiopia	Ethiopia	C1	%	96%	4%	85%	15%
			No. of observations	1,008	46	896	158
		C3	%	89%	11%	80%	20%

Project	Country	Cohort	%	*Girl-reported disability (6 domains)*		*Girl-reported disability (13 domains)*	
				Without disabilities	*With at least one disability*	*Without disabilities*	*With at least one disability*
			No. of observations	944	122	850	216
Pin Nepal	Nepal	C1	%	97%	3%	92%	8%
			No. of observations	389	11	369	31
		C2	%	99%	1%	99%	2%
			No. of observations	396	4	394	6
		C3	%				
			No. of observations	0	0	0	0
Plan	Zimbabwe	GS	%	82%	18%	71%	29%
			No. of observations	339	76	297	119
		HH	%				
			No. of observations	0	0	0	0
Street Child	Nepal	C1	%	100%	0%	100%	0%
			No. of observations	404	0	404	0
		C2	%	100%	0%	100%	0%
			No. of observations	406	0	406	0
		C3	%	100%	0%	100%	0%
			No. of observations	404	0	404	0
VSO	Nepal	C1	%				

Project	Country	Cohort	%	*Girl-reported disability (6 domains)*		*Girl-reported disability (13 domains)*	
				Without disabilities	*With at least one disability*	*Without disabilities*	*With at least one disability*
			No. of observations	0	0	0	0
WEI	Ghana	C1 Formal	%				
			No. of observations	0	0	0	0
		C1 Nonformal	%				
			No. of observations	0	0	0	0
		C2 Nonformal	%				
			No. of observations	0	0	0	0
Portfolio level			%	95%	5%	86%	14%
			No. of observations	12,644	651	11,458	1,838

Annex D: Details of Case Study Projects

Details of case study projects

*Note: the following project descriptions contain details of interventions that relate to the education of girls with disabilities.

Project 1: VSO Nepal ENGAGE

Organisation Name: Voluntary Service Overseas

Project Name: Empowering a New Generation of Adolescent Girls through Education in Nepal (ENGAGE)

Country: Nepal

Project Type: Leave No Girl Behind (LNGB)

Project Duration: 2017 - 2023

Project districts: Banke, Parsa and Sarlahi

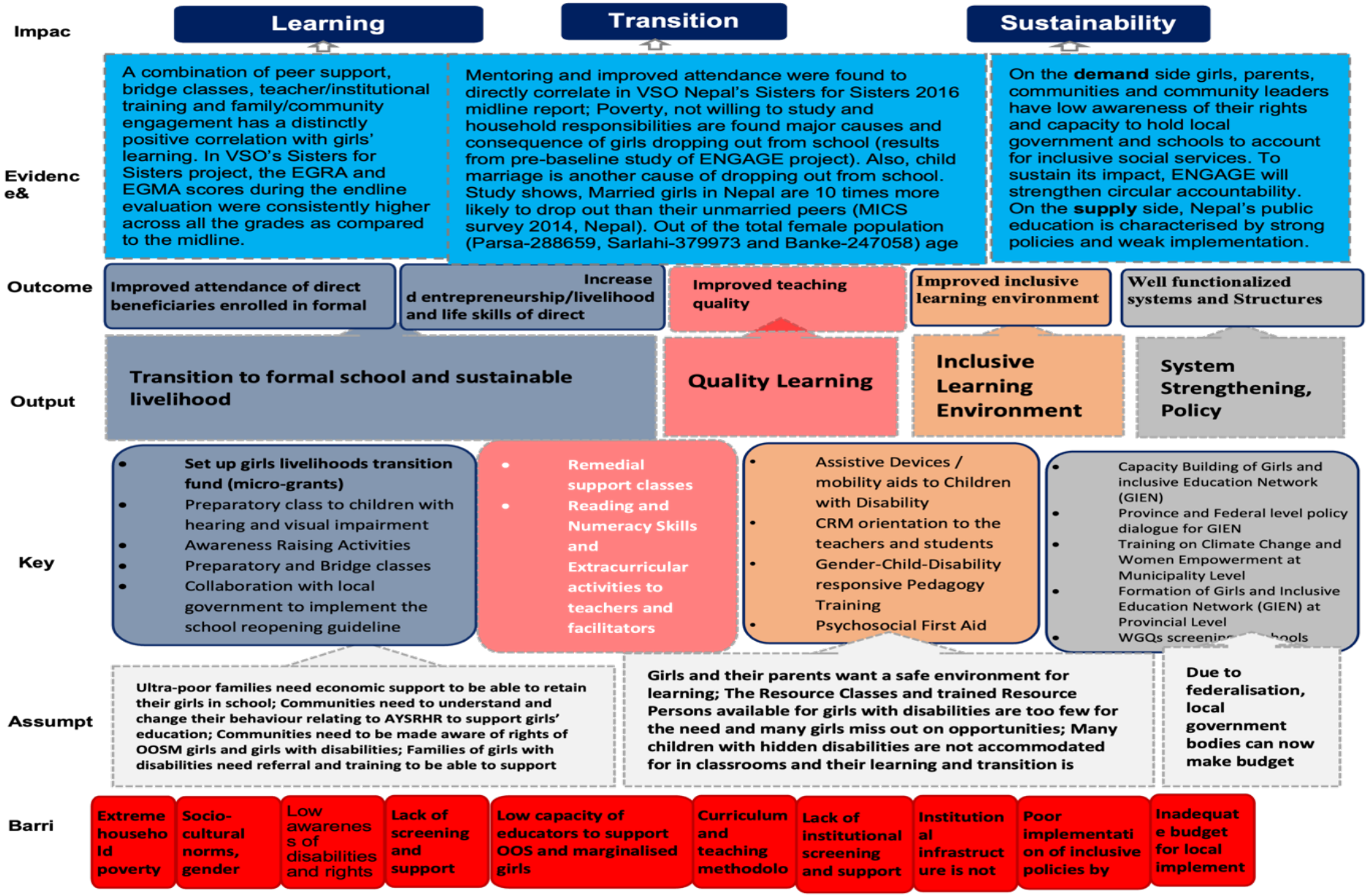
For further details: https://girlseducationchallenge.org/projects/project/empowering-a-new-generation-of-adolescent-girls-with-education-engage/

Aim of the Project

The central aim of the project is to empower OOS marginalized girls and OOS girls with disabilities through education. Various activities, as outlined below, are being implemented by the project to achieve the expected outcomes (see also the project's Theory of change[1]).

[1] VSO Nepal Engage, Baseline Evaluation Report, page ii. https://girlseducationchallenge.org/media/hyqblxdy/engage-lngb-baseline-evaluation.pdf

Figure 1: Theory of Change



Specific Project Components:

Identification and assessment of disabilities of girls

This identification of GWDs is done through community outreach while Washington Group questions are administered to the girls. The project also supports Government education officials to adopt the Washington Group set of questions for the identification of disabilities among children.

Bridge, Preparatory Classes and Resource Schools:

The project aims to provide basic literacy and numeracy skills to girls through the provision of 9-month bridge classes to help acclimate girls to education and accelerate their re-entry into formal education. Additional preparatory classes are given to children who require remedial classes. The bridge and preparatory classes also provide skills related to the pre-identified enablers of learning empowerment: training on Financial Literacy, Adolescent Sexual and Reproductive Health (ASRH), and Self Efficacy. Preparatory and bridge classes are given to girls with hearing and visual impairments after which they are transferred to resource schools. Resource schools are special education schools/ learning hubs for children with disabilities.

The project has also categorized the beneficiaries under sub-groups like OOS male siblings of the girls within the age group of 10-19 years of age who received learning interventions and other boys with disabilities who fall under the same age category. Children with severe and profound disabilities are transitioned into life skills programmes (rather than formal schooling) covering seven domains involving self-care and daily living skills.

Note: Bridge classes were operating in the 1st / 2nd years of the project with the beneficiaries already having transitioned to formal schooling or livelihoods, or life skills (for children with severe and profound disabilities).

Vocational Training

The project provides vocational training to older girls (15-19 years old), including GWDs, to transition to safe employment/entrepreneurship. This also involves in-kind support to the primary actors based on their approved business plan. Those who transition to livelihood have been provided training and the various professions they have transitioned into include auto-rickshaw driving; beauticians; doll-making; farming.

Disability-friendly infrastructure:

The project includes support for schools through disability-friendly infrastructure e.g., ramps, disability-friendly toilets, provision of IEC materials in disability inclusion, provision of learning materials- sign language, Braille etc. The project also focuses on the provision of assistive devices (wheelchairs, hearing aids etc.) in both bridge/ preparatory classes and resource schools. Physical Rehabilitation Therapy (PRT) is also included for CWDs.

Interactive programs and workshops with parents

The project has further supported parents through interactive programs and workshops focused on increasing parental support in girls' learning by addressing the barriers related to girls' excessive engagement in household chores. Parents/caregiver support training (for children with disability) is also provided. The project also prepared sign language videos to inform parents on safeguarding/ child protection/ health safety etc.

Teacher training

School teachers/ educators are trained in gender-responsive pedagogical approaches, psychosocial first aid training, distance teaching methods and effective school improvement plan (SIP) formulation. Teachers have also received capacity development training on disability accessibility and disability-friendly working environment in the intervention schools. Teachers and bridge class facilitators were also trained in reading and numeracy skills and extracurricular activities were incorporated into their teaching/ learning mechanism.

Peer to Peer education through Big Sisters

Big sisters are community volunteers (from the same community as the beneficiaries) who provide coaching/ mentoring for the “little sisters” i.e., the beneficiaries who are a part of the project. Big sisters also help in raising awareness among parents and community members and motivate parents to send girls to school.

Response to Pandemic

In response to the pandemic, the project had run distance teaching learning (DTL comprising peer-to-peer education, group discussion and one-to-one coaching through household visits, radio schooling programs, and video dissemination), and psychosocial support.

Community Awareness

The project also focuses on community-based attitudinal and behavioural changes to promote the inclusion of girls in education (particularly girls with disabilities), with a focus on parents and caregivers but also extending to community awareness programmes.

Formation of the Girls Inclusive Education Network (GIEN)

Additionally, VSO formed the GIEN as a way of continuing the work done to train and support the Big Sisters. The aim of the network now expanded to include all other GEC projects and organisations outside of the GEC, is to raise the voices of marginalised girls, lobby local authorities, influence policies, promote girls’ rights to education and protection, and raise awareness on climate change.

Project 2: Team Girl Malawi

Organisation Name: Link Education International

Project Name: Transformational Empowerment for Adolescent Marginalised Girls in Malawi (TEAM Girl Malawi)

Country: Malawi

Project Type: LNGB

Project Duration: 2017 - 2023

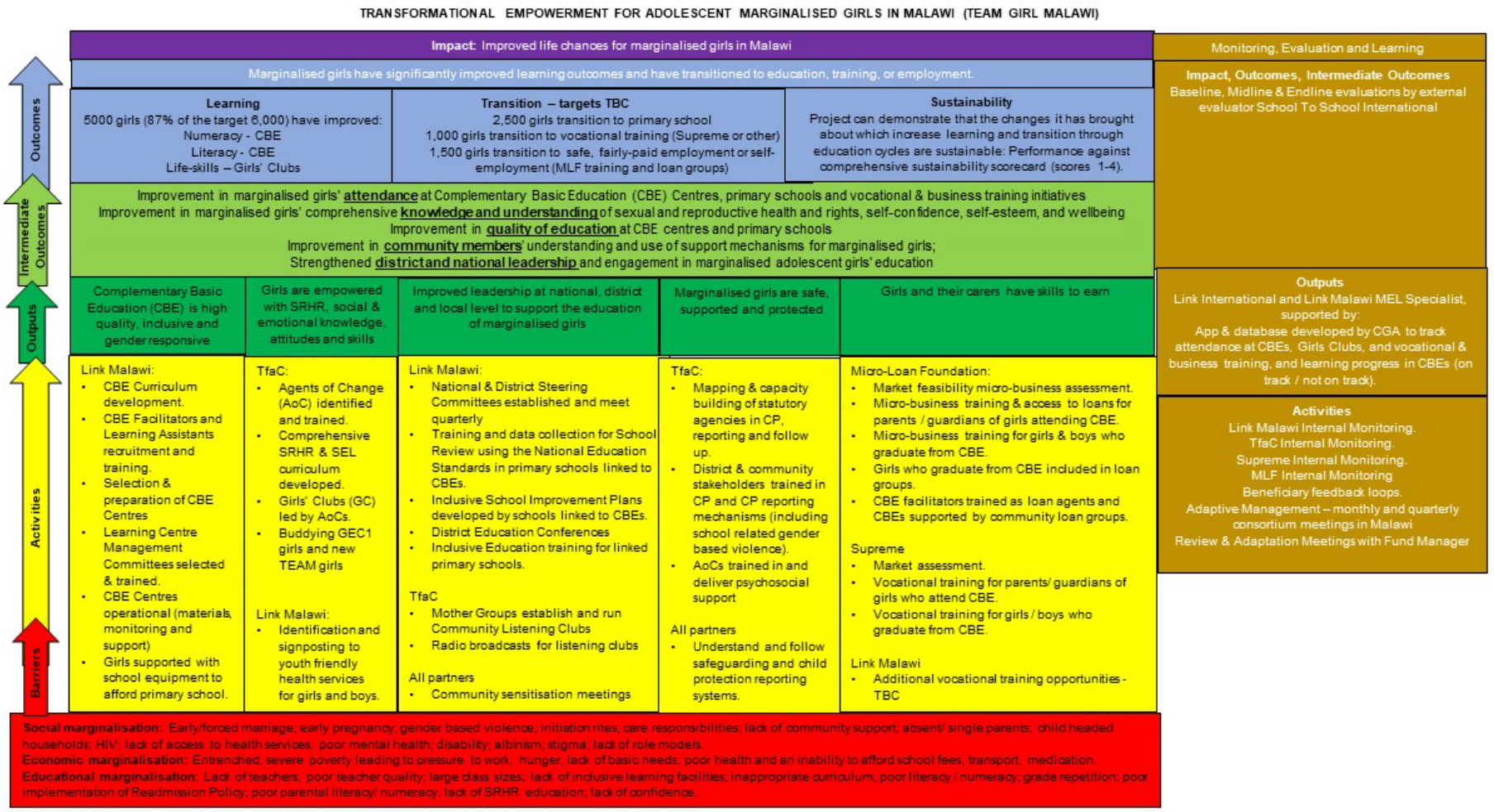
Project districts: Lilongwe urban, Dedza, Mchinji

For more details: https://girlseducationchallenge.org/projects/project/team-girl-malawi/

Aim of the Project

The central aim of the project is to address the barriers obstructing the participation in, and successful completion of, education of 10 to 19 year old marginalised girls who have never been to school or who dropped out of school without gaining functional literacy and numeracy skills, to not only improve literacy and numeracy skills but also to support their longer-term development and secure their economic independence.

Figure 2: Theory of Change[2]



[2] Link Education International Malawi, Baseline Evaluation Report, page ii. https://girlseducationchallenge.org/media/fahfjpv4/team-lngb-baseline-evaluation.pdf

Specific Project Components:

Community-based complementary basic education centres (CBEs)

Literacy and numeracy and life skills education is delivered through 105 Complementary Basic Education Centres which are employing trained community volunteers. Career development and work-based learning opportunities are also provided in these CBEs, as well as catch-up classes and learning assistance in literacy and numeracy.

Girls' Clubs located in primary schools in the same communities

The project has developed and delivered a full sexual reproductive health and rights curriculum through Girls' Clubs (safe spaces for girls to interact and learn informally) to improve girls' knowledge and understanding of these issues. Children, including girls with disabilities, therefore learn their rights and responsibilities in protecting themselves and their peers in 'Safe Clubs'.

Support for transition into primary school, vocational training and business training supported by micro-loans located in select communities

The project also offers financial literacy, vocational, and micro-business training to girls, including girls with disabilities, who have been supported to join village savings and loan associations to access loans for business ventures.

Training of Educators

Training on Inclusive Education has been delivered to a team of facilitators who are responsible for teaching through Complementary Basic Education (CBE) classes and Girls' Clubs. Based on the inclusive learning approach, teachers are trained to teach learners with specific challenges and to be problem solvers in the learning environment, making adaptations to pedagogy, the timing of lessons, and modifications to learning spaces to take account of multiple barriers faced by their students[3].

Support to families and community members

The project promotes good safeguarding practices and builds the capacity of community and district structures to support reporting and victim support of children who experience abuse. It works with communities and schools/learning centres to promote girls' education, including the education of girls with disabilities, and strengthen school governance to improve knowledge and understanding and work towards gender equitable education in Malawi.

Project 3: VIVA CRANE Uganda's Building Girls to Live, Learn, Laugh and SCHIP, in Strong, Creative, Holistic, Inclusive, Protective, Quality Education

Organisation Name: VIVA CRANE

Project Name: Building Girls to Live, Learn, Laugh and SCHIP, in Strong, Creative, Holistic, Inclusive, Protective, Quality Education Malawi (TEAM Girl Malawi)

Country: Uganda

Project Type: Girls Education Challenge-Transition (GEC-T)

Project Duration: 2017 - 2024

Project districts: Kampala, Wakiso, Mukono, Nakaseke, Buikwe

For further details: https://girlseducationchallenge.org/projects/project/building-girls-to-live-learn-laugh-and-schip-in-strong-creative-holistic-inclusive-protective-quality-education/

[3] https://girlseducationchallenge.org/blogs/blog-article/all-means-all-what-happens-when-you-include-girls-with-disabilities/

Aim of the Project

The central aim of the project is to continue the support provided to 9,890 marginalised girls in the first phase of the Girls' Education Challenge. Girls and their families are supported through mother-daughter and peer clubs, and 18 creative learning centres (CLCs).

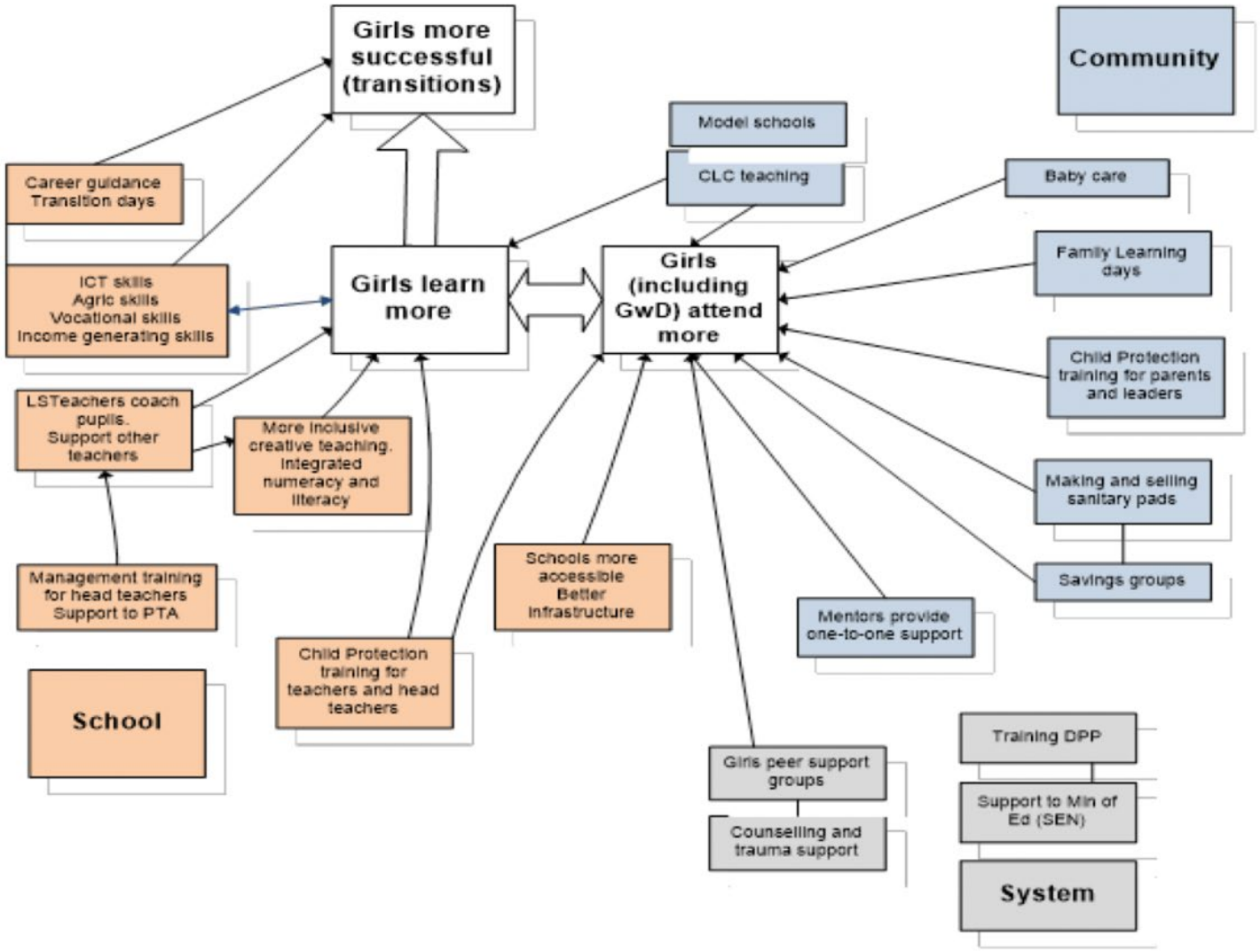
Theory of Change

The project theory of change revolves around four issues that affect girls' education:

- ***Live:*** Girls will break the cycle of abuse, violence, exclusion, child marriage, poverty, weak parenting, broken education, and limited literacy. Girls will develop strategies for success and overcoming life barriers through competency-based skills development. Community mentors and engaged parents will support this life journey. Parents and schools will form local clusters that build extra-curricular learning opportunities for children.
- ***Learn:*** Girls will achieve enhanced learning outcomes in numeracy, literacy and competency-based skills that orientate them towards purposeful life pathways. Girls will learn new and higher skills inside and outside of the classroom through creative, engaging acquisition of knowledge. Their experience in school will set them on a life pathway of achieving gender equity in the classroom and into adulthood. Girls will be supported by responsible parents, innovative and creative teachers, committed mentors, skilled counsellors and inspiring peers. Their teachers will deliver exciting, quality, inclusive education in child-centred environments that use a variety of methodologies that suit different learning styles, with additional learning support for those who need it.
- ***Laugh:*** Girls will overcome the shattering impact of abuse, rejection and failure as psychosocial support and learning therapy builds resilience and confidence. Their schools will do no further harm and help children and families learn how to build safe communities. Local parents' groups will train in holistic well-being, parenting, household strengthening, and adult literacy to help provide a smooth transition through to higher education.
- ***SCHIP:*** Learning will happen in partner schools where GEC girls have gone to help them become SCHIP schools that provide Strong, Creative, Holistic, Inclusive, Protective Quality learning environments with stronger educational and technical foundations that enable accelerated literacy, numeracy and competency-based learning that promote confidence and resilience. Girls will be helped to achieve gender equity in the classroom and into adulthood[4]."

[4] Viva and CRANE, Project Proposal, September 2016, p4.

Figure 3: Theory of Change[5]



[5] VIVA and CRANE, Baseline Evaluation Report, page ii. https://girlseducationchallenge.org/media/nijfm1nh/schip-gect-baseline-evaluation.pdf

Specific Project Components:

Mentors

A key element of the project is the targeting carried out by the project mentors for girls who were most likely to miss out on schooling. This targeting started at the beginning of GEC Phase I and was subsequently expanded to include sisters of the original selection who benefited from attending a Creative Learning Centre.

Creative learning centres (CLCs)

Creative learning centres involve numeracy and literacy programmes that relate to the girls' lives and employment opportunities. At the CLCs, the girls, including girls with disabilities, receive personalised support for learning and assessments and their involvement can range between six months to two years. ICT skills development lessons are also provided in the CLCs. After they complete the CLC programme, the girls transition to mainstream schools or home schools. Mainstream schools include learning support from teachers who are involved with the CLCs.

Life, career development and vocational skills training

Vocational training is provided to the beneficiaries, including girls with disabilities. Income-generating activities such as training on savings and financial literacy, and starting their own businesses are also encouraged.

Learning support teachers

The project provides support to over 50 different schools which includes Learning Support Teachers who provide targeted teaching and teacher training.

Mother-Daughter Clubs

The project includes mother-daughter clubs to help improve family communication skills and develop family plans. Additionally, home schools which involve vocational/skills development that the family is involved in are initiated so the girls, including girls with disabilities, can support their independent living (usually these girls are older and can't transition to a mainstream school).

Support to families, community members and government staff

The project also focuses on strengthening links between the school, community, and local government to protect children from abuse. Positive Parenting training is available to caretakers and parents as well as bringing parents of GWDs into group-based income-generating activities through Village Savings and Loan Associations (VSLA). Community mapping to identify marginalised girls, including girls with disabilities, who are not in school takes place with family discussions/ mentoring to bring them into the programme. The project also engages with the government through its partnership with the city council which helped in revamping an educational assessment centre.