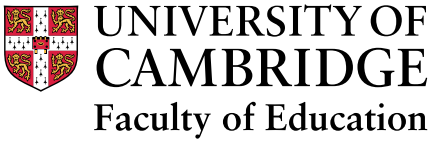

Report

Girls' Education Challenge Research Feasibility Study

August 2020

Authors:

This Report was undertaken by the Research for Equitable Access and Learning (REAL) Centre at the University of Cambridge. Dr Monazza Aslam and Dr Shenila Rawal led the review of the quantitative data and writing of the sections of the report that involved the assessment of these data, working with Joe Watson who was responsible for the detailed assessment of the quantitative datasets and providing analysis for the write-up. Rebecca Gordon was responsible for the literature scoping, interviews with implementing partners, and assessment of qualitative datasets, and wrote the sections of the report related to these aspects. Rebecca also compiled the report. Dr Phoebe Downing provided support to the compilation of datasets. Professor Pauline Rose led the process, including overseeing the different components, and providing overall quality assurance. To contact the authors of this report please e-mail: REALCentre@educ.cam.ac.uk.

Acknowledgements:

We are grateful to Alison Pollard and the UK's Department for International Development (DFID) Regional Girls' Education Advisors for their support and engagement throughout the process of preparing the report. We are also very thankful to the Girls' Education Challenge (GEC) Fund Manager and Evaluation Manager for the generosity of sharing their expertise and data, and answering our queries. The implementing partners of the GEC have also been key to the report, in terms of sharing their ideas as well as their data and other documentation.

The authors are solely responsible for the report's content and recommendations.

Suggested citation:

Aslam, M., Gordon, R., Rawal, S., Rose, P., and Watson, J. 2020. *Girls' Education Challenge Research Feasibility Study*. Report for DFID. REAL Centre, University of Cambridge.

Version 1: August 2020

Table of Contents

Acronyms

DFID	Department for International Development
DQA	Data Quality Assessment
EM	Evaluation Manager
FGD	Focus Group Discussion
FM	Fund Manager
GEC	Girls' Education Challenge
IDP	Internally Displaced Person
IMF	International Monetary Fund
IP	Implementing Partner
NGO	Non-Governmental Organisation
PPP	Public Private Partnership
VfM	Value for Money

Introduction

The Girls' Education Challenge (GEC) was launched in 2012 with a commitment to ensure up to one million of the world's most marginalised girls completed a full cycle of either primary or secondary education. The GEC is the largest girls' education fund in the world. Phase I (2012-2016) was funded by the Department for International Development (DFID) with £355 million targeting 1.4 million marginalised girls through 37 different projects delivered by non-governmental organisations and private sector organisations across 18 countries prioritised by DFID.[1] Aggregated analysis of project-level evaluation datasets of the first phase (GEC-1) of the programme by the Fund Manager (FM) identified that over 800,000 girls demonstrated measurably improved learning as a result of their participation in the projects funded under GEC. This finding was validated by the independent Evaluation Manager (EM) (Griffiths & Batran, 2017; Griffiths, Bisiaux & Di Paolo, 2017; Griffiths, Poli & Amili, 2017). Phase II (GEC-T) of the programme is currently operating (2017-2025) and has a total value of £500 million which is expected to support approximately 40 projects in the provision of quality education for up to 1.5 million marginalised girls aged 10 to 18 years.

Given the broad range of interventions covered under the programme, and the evaluation of these interventions that have taken place in Phase I and continue to take place in Phase II, a high volume of data is being generated as part of GEC. These data, if they are of sufficient quality, create opportunities to undertake research and answer specific research questions which can contribute to global evidence on what works to improve learning outcomes for the most marginalised girls in the world. Moreover, given the range of innovative interventions being adopted across GEC projects, including in contexts facing considerable challenges, further data collection could enable an even more in-depth understanding of how to address intransient problems in achieving gender equality in education.

The overall aim of this Research Feasibility Scoping Study is to determine how research related to the GEC Programme can be used to advance the global evidence base on what works to improve learning for marginalised girls, including identifying:

- what kinds of research questions could be answered by existing GEC data
- what additional data could be collected to further fill gaps in the global evidence base, and
- what methodologies might be appropriate for this future research.

With respect to existing GEC data, it should be noted that this report assesses the data based on its quality and relevance for future research purposes. It does not intend to provide an assessment on the quality of the data for the purpose of evaluation and recognises that this was the original purpose for which data were collected. Additionally, this review focused on data available at the time of the review, namely GEC-1 and the baseline of GEC-T. It should therefore be noted that there is likely to be further potential for data to be used to answer research questions once the GEC-T midline and endline data become available.[2]

It is hoped that the findings from this study will be used to support DFID and independent researchers to commission and implement robust and relevant research studies in response to gaps in the global evidence.

There have been four key stages of work as part of this Research Feasibility Study.

1) Developing the thematic approach to the data assessment

- Scoping of the global research literature on girls' education and identifying the gaps in the current evidence base of relevance to the GEC scope of work
- Design of key informant interview strategy and discussions with key informants
- Design of Implementing Partner (IP) interview strategy and consultation with IPs

2) **Collating data from the FM, EM and IPs**
   - Initial engagement to assess data availability
   - Data collation from the FM and EM
   - Data collation from the IPS

3) **Developing methodologies for the assessment of the scale, scope and quality of GEC datasets**
   - Approach to the assessment of the quantitative and qualitative data
   - Initial assessment of the two project datasets
   - Ongoing data collation for the purposes of assessment

4) **Data assessment and ongoing interviews**
   - Assessment of quantitative and qualitative datasets
   - Further interviews with IPs

Section I of this report presents the thematic approach to the data assessment. Section II outlines the methodologies used for the assessment of the scale, scope and quality of GEC datasets. Section III summarises the assessment for the quantitative and qualitative data. The recommendations are presented in the final section.

SECTION I: Thematic Approach to the Data Assessment

Our approach to the assessment of GEC datasets (both quantitative and qualitative) is grounded in scoping of the available research literature, discussions with key DFID education advisors involved in the GEC and discussions with other key informants.

Scoping of the available research literature

There is a growing global evidence base both on barriers to girls' education and what works for girls' education. This has been captured in a number of large-scale reviews, including recent ones involving authors of this report (Lloyd, 2011; Sperling & Winthrop, 2016; Unterhalter et al., 2014; Evans & Yuan, 2019 – but see Rose and Yorke's (2019) critique of this; Gordon et al., 2019). We also note that currently the Population Council is updating and expanding the evidence base on girls' education by: 1) systematically mapping the ecosystem of policymakers, practitioners, researchers and advocates working in global girls' education; 2) synthesising the evidence on what works; and 3) identifying opportunities to scale up successful interventions and investments (Psaki, 2019).

Part of our scoping process involved reading key existing reviews (and the original evidence) to consider the information in the existing evidence base. Together with discussions with DFID education advisors, engagement with key informants from the GEC FM, GEC EM and GEC IPs, this led to the development of the key themes noted in Table 1 of the report. In order to explore the extent to which data from the GEC has been used thus far to contribute to the global evidence base, we also undertook a specific key search for literature that has emerged based on GEC data (see Appendix 1).

Table 1 also presents a summary of information compiled from the scoping of existing literature, particularly that highlighted by the reviews mentioned above. This is structured according to the key themes that have been identified.

Table 1: Summary of existing literature according to key themes

Theme	Questions
1. Equality and equity	**1a) Which interventions are successful in reaching the most marginalised girls? [including related to poverty, disability, ethnicity, geography, nomadic/pastoralist populations; Internally Displaced Persons (IDP)/refugees etc.]** • **What has enabled or facilitated their success?** • **How does this vary according to context (e.g. in conflict affected states)?** There is some, but quite limited, evidence on the extent to which interventions are effective for the most disadvantaged girls (taking account of intersecting disadvantage). For example, Unterhalter et al. (2014) note only 22 studies which directly focused on girls from marginalised communities in their review, and very few studies which focus on conflict affected areas. However, there is some evidence that the introduction of village-based schools has improved access to education for girls in rural and pastoralist communities (evidence cited in Sperling & Winthrop, 2016). Providing safe spaces for girls in conflict settings was noted as being particularly important in supporting access and attendance (evidence cited in Gordon et al., 2019). There is also emerging evidence on the specific pathways needed for inclusive education for girls with disabilities, such as the use of peer learning approaches (evidence cited in Gordon et al., 2019).
2. Access and attendance	**2a) Which interventions improve access and attendance for girls? In what ways?** • **In which contexts?** • **How do patterns of attendance/non-attendance affect learning and outcomes?** There is a wide range of evidence on interventions that support improved access and attendance for girls. The evidence points to the need to target interventions towards adolescent girls, where dropout rates and absenteeism are higher (Gordon et al., 2019). There is far more limited evidence that explores the specific impact of attendance or non-attendance on learning outcomes. All reviews highlighted the importance of eliminating cost barriers for girls' education, leading to increased access and attendance. There is a wide range of evidence in different contexts that have shown that eliminating cost barriers has a positive impact on enrolment, as well as reducing teen pregnancy and early marriage. Approaches to eliminating cost barriers noted in the evidence base include large scale stipend programmes, targeted fee elimination (at the secondary level), bursaries, cash transfers and scholarships. These have been found to have a significant impact on enrolment and attendance. In kind-support (such as free uniforms) has also been found to impact on girls' enrolment, absenteeism and learning (studies cited in Sperling & Winthrop, 2016; Unterhalter et al., 2014; Gordon et al., 2019). There are some emerging and growing areas of evidence. For example, Lloyd's (2011) review noted research which explored the provision of toilet facilities as being an area

	where more research was needed. Gordon et al (2019) cite evidence that an increase in the number of toilets for girls reduced sexual assault and resulted in a significant decline in absenteeism. Gordon et al (2019) refer to recent evidence showing the increase in school attendance rates for girls as a result of the provision of sanitary pads in multiple contexts. Yorke and Rose (2019) highlight the evidence on the importance of providing contextually-relevant sanitary supplies for a positive impact. There has been a rich tradition of non-governmental organisations (NGOs) providing complementary non-formal education such as in rural Bangladesh (for example by BRAC), which has provided out-of-school girls with an opportunity to re-enter the formal system (cited in Lloyd, 2011). Recent evidence of the Complementary Basic Education programme in Ghana finds the benefits of this programme in terms of learning once children enter formal school, although low-achieving girls do not benefit to the same degree as their peers, suggesting specific interventions are needed to support these girls in particular (Carter et al., 2020). Reducing the time and distance to school as well as building more schools has been found to have an impact on girls' enrolment, including through community schools (Unterhalter et al., 2014; Sperling & Winthrop, 2016; Gordon et al., 2019). Providing bicycles or other modes of transport for girls to get to school can improve enrolment and attendance for marginalised girls (evidence cited in Gordon et al., 2019). There is also information on the positive impact of interventions that target parents to support girls' education. For example, providing information about employment returns to schooling has a positive effect on girls' enrolment (evidence cited in Unterhalter et al., 2014; Gordon et al., 2019).
3. Learning	**3a) What are effective approaches to measuring learning outcomes for marginalised girls?** **3b) Which interventions have a positive impact on (marginalised) girls' learning?** • **What is the effective/are different approaches to improving literacy and numeracy for girls (and boys)?** • **How do these approaches differ according to different population sub-groups, and for literacy or numeracy?** **3c) What factors might influence slow progress in learning outcomes?** There is a growing body of literature that provides evidence of interventions having a positive impact on girls' learning, although the number of studies on targeted interventions remain limited (Evans & Yuan, 2019; Rose & Yorke, 2019). In particular, Unterhalter et al (2014) emphasise the importance of policy and institutional cultures for enhancing girls' learning, based on 37 studies. Lloyd's (2011) review identified that promoting early learning in the first years of primary school, more instruction time, merit scholarships, alignment of age with appropriate grade and cash on delivery could lead to improvements in learning. While most studies find that financial interventions are more likely to have an impact on access than learning (Evans & Yuan, 2019), there is growing evidence that learning

	outcomes can also benefit in some circumstances. One study found that merit-based scholarship led to increased test scores for girls and targeted fee elimination increased exam scores for the most marginalised girls. Another study found that recipients of conditional cash transfers outperformed girls who did not have conditions in tests of reading comprehension, however there were also other outcomes as a result of conditions, such as higher pregnancy and marriage rates (evidence cited in Gordon et al., 2019). Additionally, other reviews have emphasised the overall limited evidence of cash transfers on learning outcomes (Rose & Yorke, 2019). Most studies focus on the impact of single interventions. However, there is evidence that multi-pronged interventions which tackle multiple dimensions of disadvantage can be particularly promising in improving learning, as evidence from the NGO, CAMFED, highlights (evidence cited in Gordon et al., 2019). There is some evidence that a gender-friendly school environment can have positive outcomes on girls' literacy results. Gender sensitive pedagogy has been found to result in a change in the gender dynamics of the school, which has led to improved learning outcomes for girls (evidence cited in Gordon et al., 2019). There are a few other interventions which have shown positive outcomes, although with a more limited range of evidence. One evaluation found that a teacher training programme at an Early Childhood Care and Development centre led to higher learning scores for girls, particularly where training had supported both teachers and parents (evidence cited in Gordon et al., 2019). Unterhalter et al., (2014) also cite studies which demonstrate that group learning and positive teacher training can lead to increased participation and learning outcomes for girls. Additionally, two studies found that the presence of female teachers had a positive effect on improving test scores for girls (evidence cited in Gordon et al., 2019).
4. Non-cognitive skills	**4a) What are effective approaches to measuring non-cognitive skills (E.g. self-esteem)?** **4b) How do different interventions improve non-cognitive skills?** **4c) What effect do non-cognitive skills have on learning outcomes?** Girls' clubs have been found to have an impact on non-cognitive skills, although there is still a relatively small amount of evidence available (Unterhalter et al., 2014; Sperling & Winthrop, 2016; Gordon et al., 2019). Peer learning is emerging as an intervention which has positive impacts on girls' self-esteem and confidence (evidence cited in Gordon et al., 2019). There were also emerging reports of the impact that integrating girls within school governance structures can have on girls' confidence (cited in Unterhalter et al., 2014). Community-based programming has led to benefits for girls' education, including increased commitment to study and self-efficacy (evidence cited in Gordon et al., 2019). However, overall, there is limited evidence on the effect that non-cognitive

	skills have on learning outcomes and a discussion about which approaches were most effective to measuring non-cognitive skills (there was quite an even mix of qualitative and quantitative studies).
5. Gender-sensitive pedagogy	**5a) What are effective approaches to gender-sensitive pedagogy?** • **What has been the effect on teaching?** • **What has been the effect on learning and other outcomes for girls and boys?** **5b) How can this inform wider government system reform?** There is reasonably strong evidence that interventions that engage with teacher education, training, attitudes and levels of support can yield positive outcomes for girls. For example, a number of studies have found that training teachers with regard to subject knowledge, pedagogy and gender equality can play a significant role in reducing girls' drop out and improving learning outcomes (studies cited in Unterhalter et al., 2014). There are also a number of studies which show the positive impact on interactive and inclusive teaching strategies, although these are not necessarily focused on gender (Sperling & Winthrop, 2016). There are fewer studies that focus on interventions that have tried to shift or change teachers' attitudes to gender; however, some studies have shown positive impacts on teachers' attitudes (cited in Unterhalter et al., 2014).
6. Social norms	**6a) Which interventions are most effective in tacking social norms such as associated with early marriage, early childbirth and gender-based violence? How are they effective?** There is limited availability of evidence on the impact of interventions on social norms. However, there are some studies that provide relevant insights. For example, political leadership, particularly women's leadership, was found to have the potential to change attitudes and aspirations for girls' education (Unterhalter et al., 2014). Life skills and community-based education programmes for girls during adolescence were found to help to challenge harmful gender norms, reduce child marriage and teenage pregnancy in a few separate contexts (evidence cited in Gordon et al., 2019). There is also strong evidence that learning spaces, girls' clubs and sex education approaches are effective at building knowledge and confidence (evidence cited in Unterhalter et al., 2014). One study which aimed to reduce harmful societal gender norms increased adolescents' support for gender equality and led to more gender equitable behaviour (evidence cited in Gordon et al., 2019). Whole-school approaches to targeting school-related gender-based violence can improve girls' feelings of safety at school and reported changes in the environment of the school (Sperling & Winthrop, 2016; Gordon et al., 2019). However, this does not necessarily imply wider social norm change. Indeed, Unterhalter et al. (2014) note that there were still few rigorously evaluated programmes that work on gender-based

	violence, but that those studies that exist showed a change in attitudes and knowledge about gender and violence. Women's literacy programmes can be a key area for transforming gender norms and identities, particularly if they allow women and girls the chance to develop gender awareness, and interventions that seek to enhance the capacity of poor or marginalised women to participate could have potential; however, there are not many studies that investigate this linkage as a result of interventions (Unterhalter et al., 2014).
7. Transitions	**7a) To what extent do interventions help smooth transitions for girls between:** • **Non-formal into primary/secondary schools** • **Primary to secondary** • **Secondary to technical and vocational/higher education** • **Secondary (or primary) to work** Overall, there is quite limited evidence on the extent to which interventions smooth transitions for girls. However, one study found that early childhood education could influence girls' retention and learning in the long term. Additionally, a stipend programme was found to make it more likely that girls transitioned to middle and high school (evidence cited in Gordon et al., 2019). Lloyd's (2011) review identified the importance of continuing educational programmes for older girls, including through non-formal education programmes, to support the transition to the job market (see also Gordon et al., 2019). However, there are not many studies which focus on these types of interventions and their long-term impacts on transitions. Some studies have explored the impact of interventions which target the school to work transition. Overall, there is a need for more evidence of the impact of interventions aimed at improving access and learning for marginalised girls for their future life outcomes, including in relation to work opportunities (Gordon & Rose, forthcoming).
8. Boys	**8a) What can we learn from the interventions about the (spill over) effects for boys?** • **To what extent have GEC activities been successful in improving access, attendance and learning outcomes for boys?** • **How can this contribute to countering potential backlash about a specific focus on girls' education?** **8b) If boys do need (additional) support, what type under what circumstances?** Overall, there is limited evidence showing the spill over effects of interventions on boys. Scholarships for girls in Kenya were found to increase boys' test scores as well. Similarly, a female stipend programme in Pakistan also led to an increase in boys' enrolment. There is also some limited evidence that mainstreaming gender in the classroom improves classroom participation for both boys and girls (evidence cited in Gordon et al., 2019). There are also a small number of studies which focused on interventions with boys to support gender equality which demonstrated positive impacts (evidence cited in Unterhalter et al., 2014).

9. Programme change	**9a) How has evidence (such as from monitoring and evaluation) been used to adapt interventions?** **9b) What are the barriers and levers to evidence being used to adapt interventions in different contexts and why?** From our scoping, there is very little in the global evidence base about the use of evidence to adapt interventions.
10. System change	**10a) How do lessons from GEC projects inform national system reform in ways that benefits retention and learning for (marginalised) girls?** • **What are the differences in success of implementation and outcomes between interventions working within and outside government provision?** • **What are levers and barriers to successful influencing of national systems in different contexts?** **10b) To what extent can GEC projects be scaled up to reach a larger number of girls and communities within a country, and/or replicated to other settings?** There is extremely limited availability of evidence of the impact that organisations or interventions have had on influencing system change. Some reviews have highlighted the evidence that legislative change is vital to support girls' education (Unterhalter et al., 2014). Gordon et al's (2019) review identified the importance of changing harmful laws and policies that restrict girls' education and supporting existing laws with enabling strategies (e.g. girls' re-entry to school after pregnancy). This review also cited the importance of embedded gender equality within policy and that gender mainstreaming ensures that commitments are put into action.
11. Cost-effectiveness/Value for Money (VfM)	**11) What are the most cost-effective ways (with equity) of reaching marginalised girls?** There is an overall gap in the area of evidence of cost-effective approaches to reaching marginalised girls. One study identified has shown the potential for cost-effective approaches to support the most marginalised girls (Alcott, Rose & Sabates, 2016).

What are the gaps in the evidence base?

Based on the reviews of evidence noted above, we identified the following gaps in the evidence base in particular:

- **Equality and equity:** There is limited existing evidence on whether interventions are successful in reaching the most marginalised girls (e.g. girls with disabilities, girls living in rural areas, girls from ethnic and linguistic minorities etc), and if so, how they do this.
- **Access and attendance:** There is limited research that has focused on the linkages between attendance and learning outcomes.
- **Social norms:** Evidence from community-based interventions shows promising practice on changing attitudes towards girls' education, but there is a need for further research in this area, particularly given deeply entrenched patriarchal norms.
    - In particular, research on interventions to change social norms in the community often focus on attendance, more evidence is needed to know the impact on learning and gender equality/empowerment.
    - Further information is needed about the impact of interventions in education to transform gender social norms and how this links to empowerment/gender equality more broadly. As noted by Rose & Yorke (2019), this is likely to require longer-term research, as social norms can take a long time to shift.
- **System change and cost-effectiveness/VfM:** There are large gaps in evidence on wider systems-strengthening and on the politics of system change processes. Furthermore, there is limited evidence on the cost-effectiveness of reaching the most marginalised girls in an equitable way.

Based on the review of the evidence, we also identified a lack of longitudinal studies and mixed methods studies. In addition, existing reviews gave some insight into the research methods that could be used to expand the global evidence. These are summarised by Psaki (2019):

- Administering assessments (e.g. testing literacy or numeracy skills)
- Examining school records (e.g. of attendance, test results)
- Asking students, parents, or teachers questions (e.g. what are the main reasons you did not attend school last week?)
- Conducting observations (e.g. classroom, household).

One of the key questions that Psaki (2019) poses is: 'which parts of the programme are most effective?' Thus, future studies on girls' education ideally need to prioritise understanding on which aspects of interventions have the most positive impact on girls' education. It is important to note also, however, that reviews commonly highlight the importance of multi-dimensional interventions to tackle the different forms of disadvantage that girls face, so this needs to be taken into account in research design.

SECTION II: Methodologies for the Assessment of the Scale, Scope and Quality of GEC Datasets

Following substantial work collating data provided by the FM, EM and IPs from GEC I and II, we identified projects for which quantitative and qualitative data were available for further assessment of their quality. Firstly, our aim was to review the scale, scope and quality of these data. Secondly, we aimed to identify the kinds of research questions that existing data can help to answer, and the potential methodologies that could be used for future research. This section outlines the approach adopted for assessing the quality of the quantitative and qualitative data.

Approach to the assessment of the quantitative data

A two-stage process was used to assess the quality of the quantitative data available for the projects that have participated in Phase I and Phase II of the GEC:

1. **Stage one** – an initial assessment of all projects' data.
2. **Stage two** – An in-depth quality assessment of those project data where a final dataset for that project are identifiable and available.

The process broadly followed the general principles of data quality assessment developed by the International Monetary Fund (IMF) (presented in the 2012 generic Data Quality Assessment (DQA) Framework[3)], along with other quality assessment frameworks.[4] Adherence to the principles of the DQA Framework is reflected by, for example, investigation into the conditions under which data were collected. Initially these approaches were piloted by applying them to two example projects. These were shared with DFID in an initial Scoping Report. After agreement with DFID, this approach was then applied across all GEC project data.

Stage one

Stage one of the quantitative data assessment focused on an initial scoping exercise to provide a summary overview for each project in the GEC portfolio giving relevant information, particularly on the project's data collection process, scope of data, availability of data etc. This was achieved through an examination of project endline reports (written by the independent evaluators of each project) and quantitative review templates (produced by the FM).

In the next instance, all available quantitative datasets (whether final or non-final) were counted and recorded. This involved searching through all potential data sources – i.e. those provided by the FM, EM (for GEC-1 data sources), independent evaluators and individual IPs. This exercise resulted in a list of all available quantitative data that could undergo a more rigorous quality assessment by the team.

Specifically, these available quantitative datasets were examined in detail in an attempt to identify which were the final usable datasets. For project datasets to be considered final and usable they needed to meet the following criteria:

- Be the sole project dataset available on a particular set of measures or be clearly identifiable as the most recent project dataset available on a particular set of measures.[5]
- Be associated with an adequate codebook or have variables within the dataset itself that are sufficiently labelled.

Projects for which the identification of final quantitative datasets was not possible are not included in the full DQA process. Consequently, they also do not form part of the thematic evaluation exercise conducted that aims to identify key themes across data sets and projects. Following from this identification, the DQA process was conducted for all completed GEC-T projects (29 of 29) and all but 5 completed GEC-1 projects (33 of 38).

Stage two

The second stage of assessment involved a more rigorous and in-depth examination of final data sets. This included an assessment of the nature of the sampling procedure, the extent of sample representativity, characteristics of the data collection tools and methodology, as well as data management and quality control as could be determined from the existing documentation available to the REAL Centre team. This process has been largely conducted with recourse to endline reports (supported by initial investigation of quantitative review templates produced by the FM). Additionally, written information regarding data management concerns such as data processing and data editing were also verified by examining the final datasets.

Key facets of survey sampling, sample representativity, data collection tool and methodology characteristics, data management and quality control were given traffic light colour coding (green, amber and red) to reflect the evaluation team's assessment of the quality of that particular aspect of the data set. For example, a green code within a data set may be given in an instance where a particular project provides high quality data (specific indicators or range of indicators that are reliable/relevant etc.). The quality assessment uses both objective as well as subjective criteria. The determination of which colour is most applicable was conducted in the most transparent manner possible, with justifications provided in all cases (contact authors for more information).

Furthermore, Stage two of the quantitative data quality assessment also included identifying the presence of indicators that are relevant to the key themes of interest. For each of these themes, information on up to fourteen indicators was sought by looking through final datasets (contact authors for more information). The presence of indicators relevant to the key themes is also measured using a traffic light system. In this instance, the colours are used to show where a dataset provides accurate information directly relevant to an indicator (green), information on a proxy or related indicator (amber) or no measure of direct or related relevance to an indicator (red). This entire process helped inform a more macro assessment of whether the GEC data are adequate, relevant and useful for the purpose of undertaking future research within the key themes identified above.

Approach to the assessment of the qualitative data

Qualitative insights around experiences, perceptions and behaviours are important and valuable evidence for informing policy decisions in education provided they are derived from high-quality studies. To assess the quality of qualitative data collected as part of the GEC, we used the BE2 guidance note[6] on *Qualitative Research in Education: Considerations for Best Practice*. This note provides guidance on how to assess the quality of qualitative research from its early design stages through implementation, data collection, data analysis and dissemination.[7] Based on this, we used the approach set out below to guide our assessment of the qualitative research produced under the GEC:

Identifying best practice

The main principles that should be considering when assessing whether research is of high-quality include:

1. **Systematic:** Quality research does not appear to be a 'cherry-picked' collection of interviewees or data that supports pre-existing ideas about the answers. Instead, it reports and discusses the full range of evidence to show balance.
1. **Credible/appropriate:** From the research question, the choice of methodology, to the method, the type of questions asked, to those who have asked them, and the ways in which they are asked, should generate credible and relevant accounts of phenomena.
2. **Transparent:** Quality research methods and protocols for analysis are documented so that others can see exactly how the data were collected and analysed.

Questions for the assessment

Prior to the research design

- Was there sufficient formative research done to determine what was not already known?
- Were the aims of the research clearly drawn up and based on the existing evidence base?
- Were the research questions worth asking, and was this discussed and articulated? In whose interests are the questions being asked?

During the planning phase

- Was there consideration of whether the methodology works with questions to produce the knowledge that is needed?
- Was there sufficient consideration of the methods (tools) that can be used to acquire the knowledge required?
- Were these paradigms and considerations articulated and scrutinised?
- Did the protocols for gathering data follow from a clear description of the study purposes and the main questions and frameworks guiding the study?
- Did the pilot enable the testing and adaptation of this protocol?

During the data collection phase

- Were sampling decisions made based on the purpose of the study and theoretical assumptions?
- Do the data collection protocols clearly identify who will provide data, how data collection processes will be undertaken and who the team who will collect data is?
- Is information provided on the training of the team who undertook data collection?
    - Is information provided on discussions and awareness of the reflexivity and positionality of the research team?
- Was the timeframe for data collection adequate for the purpose of the research?
- Is there a thorough data log that keeps types of data and related information, such as the date and location where data were collected, format of the data, duration, research responsible, note taker, original language and critical identifiers from the respondent, among others?

During the analysis

- Did the data analysis process follow the purpose of the study and the research questions?

Throughout – ethics

- Did the research clearly follow ethical guidelines including the 5 Rs (relationships, respect, relevance, responsibility and reciprocity)?
- Were the harms and benefits of the research adequately considered?

- How did the research team ensure privacy and confidentiality if required, and how was this articulated?
- Which ethics committees and institutional review boards were consulted prior to the research?
- How have dissemination processes been decided upon?
- Have ethical processes and decisions been communicated openly and transparently?
- Is there evidence of member checking and/or peer debriefing?

These questions formed the basis for our assessment of the qualitative data.

Initial scoping

Prior to undertaking an assessment of the qualitative data, an initial scoping exercise was conducted to provide a summary overview for each project in the GEC portfolio giving relevant information, particularly on the data collection process undertaken for each project, scope of data, availability of data etc. This was achieved through an examination of project endline reports (written by the independent evaluators of each project) for GEC-1 and baseline reports for GEC-T. Subsequently, all available data sources (e.g. transcripts, data protocols) were recorded, which involved searching from data provided by the FM, EM and IPs. This exercise resulted in a list of all available qualitative data that could be used by the evaluation team to undergo a more rigorous quality assessment process. A full analysis was conducted for all projects where we had access to transcripts, which was 11 projects overall for both GEC-1 and GEC-T.

SECTION III: Data Assessment

Assessment of the quantitative data

Introduction

This section presents the findings from the quantitative DQA process conducted on the 33 of 38 GEC-1 and 29 GEC-T projects for which data were available. [8]

A single DQA table was completed for each of the projects. Creating DQA tables involved reading through various project reports. For GEC-1, this included reading at least one endline report for each project, and for GEC-T a baseline report for each project. Additionally, initial investigation of datasets selected to be final was carried out. In doing so, 68 final GEC-1 datasets were examined (an average of around two datasets per project, with GEC-1 projects having a minimum of one and a maximum of six final datasets) and 150 GEC-T baseline datasets were examined (an average of around five per project, with GEC-T projects having a minimum of one and a maximum of 18 final datasets).[9] For each project, the datasets selected as final included at least one dataset focused on learning in all instances. When more than one final dataset was included for any GEC-1 or GEC-T project, other datasets provided information depending on information collected by projects, that could include, for example, classroom observations, classroom headcounts, teacher and headteacher surveys and school surveys. The datasets selected as final for GEC-T projects also typically included a transitions-focused dataset.

Process for the completion of DQA forms: Identification of final datasets and evaluation reports

Initial consultations were held with key stakeholders to identify individuals able to support in the collation of the datasets and reports required for assessment. Collaboration between the REAL Centre, the FM, EM and Oxford Policy Management (responsible for the analysis of baseline datasets from GEC-T) enabled the team to understand how to gain access to project and programme level data. Challenges were faced in collating the data provided by the FM and EM and, therefore, direct contact with IPs was made to ensure that final versions of the endline data were accessed for this evaluation. Final project endline reports were also accessed for all GEC-1 projects and, additionally, some endline data were made available for each GEC-1 project by any one or a combination of the aforementioned sources.

A key challenge relating to the GEC-1 data was the identification of final quantitative datasets. The information provided by both individual and separate sources frequently included multiple datasets. Selecting a final dataset from all possible options was additionally hindered by confusing dataset titles. Datasets labelled as final were sometimes followed by more recently produced datasets (that were not labelled as final). Indeed, the selection of appropriate final datasets was not a trivial matter, given that the contents of some key variables appeared to change between these datasets. *Therefore, the choice of final dataset used in any research exercise would need careful assessment, as using different ones could result in very different findings.*

In selecting final datasets for GEC-1, a dataset was most frequently treated as being final if it was sourced directly from the FM, because it had been used by the FM for their own attempts to replicate the results detailed in project endline reports, and thus could be considered final. Additionally, FM data was found to be relatively well cleaned. The dataset shared by the EM was the next GEC-1 data source most often treated as being final. This is because datasets shared by the EM were more likely to have been cleaned than those shared by IPs.

The sourcing of final baseline datasets and reports proved to be more straightforward for GEC-T compared to GEC-1, and were mostly available directly from the FM.

Completing DQA forms using datasets and evaluation report information

The initial sections of DQA forms provide information on:

- the nature of the survey (target population, sampling design and sampling procedure),
- whether the sample was representative at any geographical level and if it included boys,
- tool piloting, the enumeration process, dataset length, translation and the study methodology,
- field work and data management concerns (quality control, data processing and data editing).

For GEC-1 projects, this information was obtained primarily through searching the appropriate sections of GEC-1 endline reports. For GEC-T projects, this information was largely sourced from the relevant sections of GEC-T baseline reports. For both GEC-1 and GEC-T, information from reports was supported by examining the datasets selected to be final. In all instances, this included checking the number of variables and observations in each final dataset and checking dataset labelling to gauge whether information had been translated to English.

Each subsection within this initial part of the DQA form was given a traffic light colour during the REAL Centre assessment process. These colours indicate the presence of high project data strength (green), a less strong part of the project data (amber) and a weaker component of project data (red). The determination of which colour is most applicable was conducted in the most transparent manner possible, with justifications provided in all instances. The colour coding assessment is based on both the availability of relevant information about the evaluation design in documentation as well as on a judgement of the quality of existing data (which would include how it was collected, the suitability of the indicator etc.). It should be noted, however, that colour selection inevitably retains an element of subjectivity, given the multiple issues to consider. DQA forms with respect to specific GEC-1 or GEC-T projects can be provided on request.[10]

Findings on the strength and quality of research design

Table 2 summarises the findings of the DQA analysis conducted across all GEC-1 and GEC-T 'final' datasets available to the research team. This table shows the number of project datasets and their traffic-light assessments under quality assessment criteria. As mentioned before, the coding system here is based on whether relevant information on the evaluation design is available, whether it can be relied on as being accurate and relevant, disclosure and full information pertaining to the various aspects on which assessment is undertaken (e.g. sample design, questionnaires, representativity etc.).

Table 2: DQA criteria analysis for GEC-1 and GEC-T

Criteria	Sub-theme	GEC 1			GEC-T			All GEC		
		Green total	Amber total	Red total	Green total	Amber total	Red total	Green total	Amber total	Red total
Survey	Target population	33	0	0	29	0	0	62	0	0
	Sampling design	29	4	0	26	3	0	55	7	0
	Sampling procedure	25	8	0	27	2	0	52	10	0
Sampling representativity*	National	0	0	33	0	1	28	0	1	61
	Province	0	0	33	0	4	25	0	4	58
	District	0	0	33	0	4	25	0	4	58
	Boys attainment and engagement	3	3	27	7	1	21	10	4	48
Questionnaire	Pilot testing	14	13	6	28	1	0	42	14	6
	Methodology	28	5	0	29	0	0	57	5	0
	Length	29	4	0	29	0	0	58	4	0
	Translation	30	3	0	27	2	0	57	5	0
	Enumeration process	24	9	0	29	0	0	53	9	0
Field work and data management	Quality control	14	18	1	24	5	0	38	23	1
	Data processing	8	22	3	22	7	0	30	29	3
	Data editing	5	23	5	20	9	0	25	32	5

*Sampling representativity is a key criterion on which quality of data sets are commonly assessed. To ensure a comprehensive analysis, the DQA undertaken here also considers this key criterion. This important to ensure that any future researchers using these data are aware that their analysis cannot, therefore, be used to generalise across wider populations. However, it should be noted that, given the targeted nature of GEC projects, the data collected are not expected to be representative.

Future efforts by secondary researchers to work with the GEC data will require engagement with the FM, EM and IPs to access final datasets. Researchers should take care when selecting final datasets from all possible data sources and are encouraged to consult the DQA forms associated with this report for guidance.

As noted above, it is challenging to select a dataset from multiple versions which sometimes come from different sources (whether shared by the FM, EM or IP). Future researchers seeking to identify final datasets can request project DQA forms from the evaluation team, which feature information on which datasets we selected as final for each project. It is hoped that in the future final datasets will be made open access. Where this has not happened, it will be vital to liaise with the FM and IPs to ensure access to the appropriate and relevant data sets.

The possibility of choosing differing versions of GEC-1 datasets from multiple possible data sources will have affected the identification of information relevant to particular themes (see discussion on thematic mapping below). This limitation may have affected the findings of our thematic analysis and will also potentially affect the findings of any future research which may be influenced by the dataset choice decision made at the outset of any research initiative using GEC data.

Future research endeavours using GEC data need to be mindful of the fact that these data have been collected based on specific programme aims and objectives on a specific sub-set of target beneficiaries within a given context.

With interventions focused on delivering particular objectives amongst specific groups, the GEC evaluations have focused on sampling associated with targeted beneficiaries. As evaluations were concerned with these project beneficiaries, dataset samples have not typically been representative at the national, provincial or even district levels. However, robust statistical techniques such as randomised controlled trials and experimental designs can mitigate biases within the sample that may otherwise reduce the robustness of analysis. Indeed, across the entire GEC portfolio, projects have frequently been assessed using a range of robust quantitative methodologies to collect the data (e.g. randomised controlled trial design, quasi-experimental designs, cohort tracking studies etc.).

In some cases, samples coverage and size might impede further analysis of the data. This was an issue noted during interviews with IPs, some projects mentioned that the GEC evaluation sample only covered a small subset of project' work, meaning they had to use their internal monitoring data to compare effectively between all schools over a longer period. Another IP noted that their GEC-1 evaluation appeared to be inconclusive because of an insignificant sample size. Challenges in sampling were particularly noted by groups targeting specific girls. One project originally had an external control group. However, it was found to be very challenging to include girls with disabilities within this because of the very high dropout rate and low retention levels of girls with disabilities in school. Additionally, many girls with a disability do not transition to secondary school, so they felt that it would be extremely difficult to find a sufficient sample size of girls in secondary schools from the treatment group as the programme developed.

Evaluations of GEC-1 projects have often focused on examining improvements in learning and retention outcomes amongst intended beneficiaries only. This is advantageous as analysis of the data can be presumed to be directly concerned with intended beneficiaries, and the impact of the interventions on them. However, such sampling strategies could present limitations on whether further research conducted using these data could then be generalised across other population groups. Given the specificity of the samples, making generalisations at the national, provincial or even district level may, therefore, not be meaningful. This means that replication of GEC interventions in

different contexts might produce varying results (as could be the case with any intervention assessed using a sample that was not representative of the new context in which an intervention was to be implemented). However, this is a limitation of any evaluation of interventions that needs to be considered when that intervention is going to be scaled up or implemented in an alternative context, and so is not unique to the GEC datasets.

In order to be able to identify progress in learning robustly, it is necessary to track the same children over time. During interviews with IPs, a number of projects highlighted that they faced difficulties in tracking those in their programmes. For one, this was particularly the case due to transitioning communities as a result of Ebola. They noted that this had led to a new approach for GEC-T, introducing a tracking model using laminated cards with a barcode, photo and signature display to be able to verify and validate data from girls in their programme. For another, a challenge in tracking girls occurred because the project changed evaluators and didn't have any way to identify girls who had been part of the evaluation for GEC-1. They also noted issues with dropout and attrition, and so 'being able to follow a story of girls over four years is really difficult.' Tracking once girls had left school was also identified as a considerable challenge, as girls disperse and are difficult to re-contact if projects didn't work or engage with them as soon as they leave. The sampling approach adopted across the two phases has differed which has meant that girls could not be tracked over time in any of the projects.

The DQA has revealed that both GEC-1 and GEC-T projects on the whole provide good quality information regarding the process of data collection. GEC-T projects also provide good information with regards to data management and processes.

The DQA has revealed that the majority of GEC-1 programmes provide relatively comprehensive information regarding the process of data collection such as the target population, sample design, whether the questionnaires were piloted, the methodology used, time frame within which data were collected etc. Several projects indicated the use of tablets for data collection which would suggest higher levels of fidelity in the data that was then generated (e.g. "Discovery Project Kenya" (Discovery Communications) and "Child-Centred Schooling: Innovation for the Improvement of Learning Outcomes for Marginalised Girls in Zambia" (CAMFED) and "New Equilibrium for Girls" (CAMFED)).

However, on the whole, there does not appear to be comprehensive information relating to data cleaning, missing/incomplete data and the treatment thereof within several project datasets for the GEC-1 data. Projects were found to have missing or incomplete data without the requisite accompanying details on how this arose (and, relatedly, was accounted for in external evaluation). In the absence of this information, it is posited that data quality concerns could have arisen through missed questions or data being incorrectly entered. Without clear guidance in project reports, researchers are likely to find it difficult to be able to analyse the data in a meaningful way without further engagement and clarifications from the FM, IP or the independent evaluators. GEC-T projects on the whole have been deemed to be much stronger in both data collection and in data management and data processing.

In relation to data processing, the DQA investigation has revealed some discrepancies between what is observed within the available datasets and what is reported within the evaluation documents.

Baseline/endline reports for multiple projects at GEC-1 and GEC-T provided detailed information on the steps that have been taken to ensure the production of clean and accurate data. However, this information conflicts in several instances with the state of the final dataset(s) for projects as observed by the evaluation team.

More generally, in interviews with IPs, some recognised the strength and expertise of evaluators, acknowledging the importance of having an external person coming in and giving a different perspective on whether the programme is on track in a way that complemented their ongoing monitoring. Overall, IPs mentioned that they thought GEC generally has a higher standard of evidence than is required by most, 'the GEC was noble and unique in coming up with an ambitious plan in a relatively rigid and consistent data strategy in very different contexts, age groups, doing different things, and see if you can compare apples to oranges on the same metrics.' It was also considered that these standards were helpful in allowing projects to develop their monitoring and evaluation skills. In some cases, external evaluators have given data collection training which project staff have been able to attend, and so improve their own monitoring capacity.

During interviews with IPs, some raised concerns about the complexity of their project design and the way in which evaluations did not always capture this. Some IPs have had several different evaluators over GEC-1 and GEC-T, which added to some of the difficulties. Some noted that the overall advice and support from the FM for the evaluation was helpful, but that further support could be beneficial for ensuring that the evaluation design recognised and captured the full impact of the projects.

Potential themes for future research using existing quantitative data

Searching through datasets for topics of interest

The second part of the DQA form aimed to collect information on the presence of indicators relevant to key themes of interest. These themes cover eleven thematic areas as identified in Table 1. Information on 39 individual indicators across the eleven thematic areas were sought by looking through final datasets.

The search process involved identifying whether or not information on specific topics was available from final data sets (68 datasets for GEC-1 and 150 at GEC-T). Due to the large number of datasets, this required a partially automated search process. To implement this, a script was created in the programming language, R. This script searched variable names, labels and values for information of relevance to different theme indicators. For example, the indicator of disability was investigated by searching variable names, labels and values for each of the following character strings: 'disab', 'hearing', 'eyesight', 'visual', 'mobility', 'impairment', 'respir', 'illness'.[11]

Partially automating the searching approach allowed all datasets that were reviewed to be searched systematically within the timeframe for this report. Using a script is also likely to have reduced error that could otherwise have arisen through potential researcher mistakes given the large number of datasets and variables. However, this search method has its own limitations that the research team recognise. For some projects, datasets were not well labelled. This meant that manual searchers of individual codebooks (where available) were required to support automated topic searching. Furthermore, the character strings searched for might have omitted key terms in some instances. For example, the search strings searched for to identify the presence of a geography-related topic did not include the potentially relevant term, 'slum'. (Instead, searching was restricted to the following strings, 'location', 'longi', 'latitud', 'urban', 'rural' and 'geography'.) As such, it is possible that some information available in project datasets was missed.

The scripted search did, however, lead to the identification of many variables of relevance to different sub-themes. Indeed, time constraints meant that it was not possible to fully investigate all variables of relevance to every theme indicator. Nevertheless, Table 3 does provide a landscape view of some themes that have emerged and that could provide future researchers with potential avenues for

further exploration. As with DQA criteria, the presence of indicators relevant to key themes was also measured using a traffic light system. In this instance, the colours were used to show where datasets provide information directly relevant to an indicator (green), information on a proxy or related indicator (amber) or no measure of direct or related relevance to an indicator (red) through the search conducted.

Findings theme 1: Equity and equality

Project datasets for GEC-T are particularly strong in providing valuable information on data disaggregated by different forms of disadvantage, related to theme 1. Some GEC-1 project data provide relevant indicators under this theme, however, the strength and frequency of this is greater in GEC-T project datasets.

Theme 1 relates to which interventions can be analysed for their success in reaching the most marginalised girls with respect to access and learning (related to themes 2 and 3). Marginalisation in many contexts manifests itself through socio-economic status, disability, gender, geography, ethnicity, parental education etc. For future research to understand whether these programmes have been successful in reaching these most marginalised girls, data need to have been collected across these dimensions.

For GEC-1 datasets, approximately half appear to have variables pertaining to disability, whilst all of the GEC-T datasets have a variable relating to disability (see Table 7). In general, across the GEC-1 and GEC-T portfolios, basic data appear to have been collected in relation to disability, although the reliability of these data in GEC-1 would need further investigation in a number of cases. For example, several datasets include binary (yes/no) variables concerning disability relating to sight, hearing, vision and mobility and long-term illness. GEC-T projects provide more nuanced information on disability, including using the Washington Group short set of questions. For example, the "Promoting Advancement of Girls' Education in Mozambique" (Save the Children) collects information on the use of hearing aids and disability assistance devices. Similarly, the "Educating Nigerian Girls in New Enterprises" (Mercy Corps) programme collects information not only on girls' disabilities but also information on the availability of assistive devices such as Braille textbooks.

More than half of the GEC-1 datasets collect data on a proxy measure of socio-economic status and all of the GEC-T project datasets examined in this report appear to be collecting data in the form of some proxy indicator. Across both the GEC-1 and GEC-T portfolios, information on socio-economic status has mainly been proxied by the highest level of parental/household/caregiver education/literacy and/or occupation. However, a few projects have collected information on various assets that may allow the computation of an asset index e.g. "The Child-centred Schooling: Innovation for the Improvement of Learning Outcomes for Marginalised Girls in Zambia" (CAMFED) includes an asset index in their data collection.

Information relating to geography and ethnicity varies across programmes in the portfolio. The data for some programmes provides information on geographical location and religion/ethnicity. Both these topics are, for example, addressed in the "Kenya Equity in Education Project" (WUSC). Overall, 5 GEC-1 and 12 GEC-T projects directly covered geography, while 11 GEC-1 and 12 GEC-T projects gave information of direct relevance to ethnicity.

The DQA found only a very small number of projects providing data on topics concerning migration, with respect to children's refugee/nomadic/pastoralist/IDP status. One of 33 projects in GEC-1 and four of 29 projects in GEC-T give information of direct relevance to this indicator.

Table 3: DQA theme analysis for GEC-1 and GEC-T

Criteria/Theme	Sub-theme	GEC 1			GEC-T			All GEC		
		Green total	Amber total	Red total	Green total	Amber total	Red total	Green total	Amber total	Red total
Theme 1: Equality and equity	Disability (from measures concerning individual children)	17	2	14	27	2	0	44	4	14
	Poverty/ Socio-economic status	12	7	14	13	16	0	25	23	14
	Gender	3	3	27	7	1	21	10	4	48
	Geography	5	3	25	12	3	14	17	6	39
	Nomadic/Pastoralist/IDP/ Refugee	1	2	30	4	4	21	5	6	51
	Ethnicity	11	0	22	12	4	13	23	4	35
	Parental literacy	9	6	18	9	17	3	18	23	21
Theme 2: Access and attendance	Disability (from measures concerning an entire school/project)	3	7	23	15	13	1	18	20	24
	Language of instruction/home language (and/or refugee status, if available)	4	11	18	24	4	1	28	15	19
	Economic barriers	14	3	16	25	1	3	39	4	19
	Supply-side constraints	9	10	14	26	2	1	35	12	15
	Violence	12	3	18	27	1	1	39	4	19
	Completion rates	2	14	17	5	8	16	7	22	33
	Dropout rates	11	8	14	13	6	10	24	14	24
	Enrolment rates	16	4	13	24	3	2	40	7	15
	Attendance rates	16	7	10	26	2	1	42	9	11
	Teachers number and quality	11	8	14	25	3	1	36	11	15
	Deployment of female teachers	7	5	21	12	1	16	19	6	37
	Fragile and conflict affected states	2	2	29	17	6	6	19	8	35
	Geography	6	2	25	13	2	14	19	4	39

Theme 3: Learning	Literacy scores	32	1	0	26	3	0	58	4	0
	Numeracy scores	32	1	0	26	3	0	58	4	0
	Other student learning outcomes	1	0	32	3	0	26	4	0	58
Theme 4: Non-cognitive skills	Non-cognitive skills	6	2	25	17	0	12	23	2	37
Theme 5: Gender-sensitive pedagogy	Teacher number and quality	7	11	15	26	2	1	33	13	16
	Perceptions on gender equality in schools (teacher/headteacher attitudes)	0	5	28	1	1	27	1	6	55
	Classroom observation data	4	0	29	12	3	14	16	3	43
Theme 6: Social norms	Cultural norms (1)	4	14	15	16	11	2	20	25	17
Theme 7: Transitions	Transition rates	1	0	32	19	0	10	20	0	42
	Girls beginning work	0	2	31	3	3	23	3	5	54
Theme 8: Boys	Data on boys	3	3	27	7	1	21	10	4	48
Theme 9: Programme change	Any relevant programme change indicator	0	0	33	0	0	29	0	0	62
Theme 10: System change	Institutional attitudes towards gender	0	5	28	1	4	24	1	9	52
	Role of women	1	4	28	1	12	16	2	16	44
	Changing attitudes	0	1	32	0	0	29	0	1	61
	Cultural norms (2)	0	0	33	0	1	28	0	1	61
Theme 11: Cost-effectiveness/ VfM	Finance data	0	2	31	0	0	29	0	2	60

Based on the available data for Theme 1, GEC-1 and GEC-T project datasets may allow researchers to explore patterns on 'who are the GEC girls?' and may allow for simple descriptive statistics and analyses across various contexts to understand whether the programmes have been able to reach those who they are targeting. These studies could adopt a regional or a more thematic focus (e.g. given the high number of projects that provide information on parental/household literacy/occupation, the pattern between this and different variables relating to the girl-child may be explored further using the existing data).

Findings theme 2: Access and attendance

The DQA revealed that a large number of GEC-T datasets provide information of relevance to the majority of the 14 indicators relevant to theme 2. GEC-1 projects also provide information across this theme, but the coverage is more variable.

Theme 2 examines the availability of data that allows an assessment of which interventions improve access and attendance for girls, what the patterns of attendance and non-attendance are and what some of the barriers that affect access and attendance might be for girls' schooling. A wide range of variables exist across the portfolio datasets that collect information on this theme. There are some projects that provide very comprehensive information pertinent to the indicators within this theme. For example, the GEC-T "Discovery Project, Ghana" (Discovery Communications) provides information of direct relevance to 11 of the 14 indicators comprising theme 2.

As with theme 1, GEC-1 datasets tend to provide less information of direct or partial relevance to theme 2 indicators. There are, of course, exceptions to this pattern. One exception is provided by the GEC-1 project, "A New 'Equilibrium' for Girls" (CAMFED). The data for this project give information of at least partial relevance to 9 of the 14 theme 2 indicators related to access and attendance.

The GEC portfolio collects very rich data on various aspects that can prevent girls' accessing school or attending regularly. Table 7 indicates that 17 GEC-1 projects have variables pertaining to economic barriers for children attending school; for GEC-T, 26 projects collect data on economic barriers to attending schools (school fee costs and uniform costs), distance to school, and safety issues in travelling to school and whilst in school.

Further, information of direct relevance to the barriers to schooling due to conflict and (in and out of school) violence is often available for GEC-T projects. GEC-T project datasets include information on violence and conflict in 27 and 17 (of 29) cases, respectively. The GEC-T projects for which data gives information on both sub-themes include "Girls' Education Promotion Project (SOMGEP)" (CARE). In this project, variables cover the safety of male and female pupils in school, and provide caregiver-derived information on conflict, violence and open fighting. The data available for GEC-1 provides information directly relevant to violence in fewer instances, (12 of 33 projects) and conflict (2 of 33 projects).

Multiple GEC-1 and GEC-T projects have collected rich information on girls' enrolment, completion and dropout rates. For example, the GEC-T project "Educate Girls, End Poverty" (Relief International) collected data relevant to each of these sub-themes. Amongst the relevant data, information on enrolment status, reasons for dropout and factors positively influencing school completion are available in the transitions-focused dataset for this project, while teacher survey responses give information on various barriers to girls' school completion. The majority of projects at both GEC-1 and GEC-T also provide attendance information. In certain cases, different measures of attendance were employed to collect data on attendance, potentially allowing for triangulation of information. To provide an example, the data for the GEC-1 project "Supporting Slum and Homeless Street Girls with

Disabilities in Kampala City to Access Quality Primary Education" (Cheshire Services) featured information on attendance history (with the endline report also including information from school registers and attendance spot checks). The availability of these data presents an opportunity for future researchers to explore the *patterns of girls' participation in school through thematic or regional studies.*

A sub-set of projects permit education participation-related indicators to be linked to key factors influencing participation. These factors include economic barriers to school attendance and the language of instruction (e.g., the GEC-T project "Community Based Education for Marginalised Girls in Afghanistan" (BRAC)).

It is also worth noting that GEC data offer an opportunity for country or regional comparisons. For example, with several interventions being implemented by projects in Kenya, many of which have collected rich information on indicators related to access and attendance, a study on Kenyan girls and issues pertaining to their access and attendance in school would be insightful. The following projects in the GEC portfolio offer this opportunity: "Discovery Project Kenya" (Discovery Communications), "the iMlango Project" (Avanti), "Kenya Equity in Education Project" (WUSC), "Education Strategies for Disabled Girls in Kenya" (Leonard Cheshire Disability), "Improved School Attendance and Learning for Vulnerable Kenyan Girls through an Integrated Intervention" and "Wasichana Wote Wasome" (Education Development Trust).

Girls' school access and attendance may be severely impacted by disability. As noted above, whilst GEC-1 projects tend to ask more generic questions such as 'does the child have a disability', 'does the child wear glasses' etc., these types of questions do not allow for nuanced discussions of incidence of disability. However, GEC-T projects have asked more detailed questions. For example, a programme in Uganda with a specific focus on disability "Supporting Slum and Homeless Street Girls with Disabilities in Kampala Cities to Access Quality Primary Education" (Cheshire Services) asks about perceptions of disability as well as questions such as 'have you ever helped someone who is disabled?' Some projects with a less specific focus on disability also provide relevant information. For example, a household-directed survey for "Securing Access and Retention into Good Quality Transformative Education" (ChildHope) considers perceptions about the rights of children with disabilities to attend school. The projects that include more nuanced data would allow for a study that could *explore the relationship between disability and girls' attendance and access to schools* across the contexts covered within the GEC portfolio. As with other research areas, a regional or country specific study may also be possible given the diversity of projects in specific regions or countries that have collected information on girls' disability status.

Another promising avenue for future research pertains to *the role that sanitation and toilet facilities play in girls' school attendance*. Of particular note are issues relating to menstruation, an under-researched area currently. Several projects across the GEC portfolio collect data on school sanitation facilities, provision of menstruation supplies and menstrual education and whether these aspects have impeded school attendance. For example, at GEC-1 "Empowering Pioneering Education Strategies for Disabled Girls in Kenya" (Leonard Cheshire Disability) provides information concerning sanitary pads and their effect on girls' attendance. This is based on three questions, which provide information on whether the child has 'started her monthly period', 'been supplied with sanitary pads by Leonard Cheshire Disability' and if 'using the sanitary pads improved … school attendance'. "Promoting Advancement of Girls' Education in Mozambique" (Save the Children) also gives data on school attendance whilst girls are menstruating as well as the provision of toilet facilities. Further, "Educating Nigerian Girls in New Enterprises" (Mercy Corps) collects rich information on schooling access indicators as well as a measure of who provided the girls with information on menstruation.

The assessment has also revealed another promising avenue for future research could be an investigation into girls' home language, the language of instruction and delivery, and the relationship that these two may potentially have with not only with the efficacy of the programme but also with other outcomes. This potential area of investigation appears particularly viable in GEC-T projects: 24 of the 29 GEC-T projects gave information of direct relevance to this topic. The availability of a variable on caregiver's abilities in the language of instruction could also be a potential area for further research.

The role of teacher quality has been shown by extensive research efforts to be a critical determinant of a child's educational outcomes (including access, learning and even non-cognitive outcomes), but less so from a gender perspective. Across the entire GEC portfolio, there is rich information on teacher gender, pupils' perceptions of teacher quality as well as indicators of gender-sensitive pedagogy. For example, several programme data collection efforts ask children about whether teachers treat girls and boys differently, how many female teachers there are in their schools, how often teachers are absent and whether teachers mark homework. Amongst the data selected as final, datasets that solely concerned classroom observation were available for 12 GEC-T and four GEC-1 projects. Classroom observation data could be used to *explore the relationship between teacher quality and pupil outcomes for girls* (access and learning) across a range of contexts.

Findings theme 3 and 4: Learning outcomes and non-cognitive skills

Final datasets selected for GEC-1 and GEC-T consistently provide valuable information on literacy and numeracy scores. Additionally, information for a good proportion of both GEC-1 and GEC-T give information on other student learning outcomes and non-cognitive skills.

All the final GEC portfolio datasets have collected rich information on learning outcomes (mostly literacy and numeracy outcomes). It is commendable that this key donor requirement has been met given that such information was not as commonly available in project evaluations prior to GEC.

The availability of rich learning outcomes data offers a useful opportunity for a future researcher to *map the learning outcomes across the entire portfolio and identify key determinants thereof*. However, given that the learning outcomes data have been collected for girls at various levels (primary/secondary) of education or through different tools (e.g. EGMA, EGRA, SEGMA, SEGRA and Uwezo-based tools), the resulting research may face some limitations in comparing the learning outcomes as they are based on different benchmarks. Further work would be needed, for example, to identify if there are anchor items across the different datasets to allow comparison. Nevertheless, this type of research does contribute to the evidence on the patterns of learning for marginalised girls and the diversity of these outcomes across a range of contexts (Outhred et al., 2019). It should be noted that whilst several of the projects assessed learning outcomes through EGRA and EGMA tools, cross-country (and sometimes even within country) comparisons are not advisable given that the format, content and recipients of these assessments differed on a project by project basis. The primary intention for using these tools was their extensive and accepted use across many contexts at that time and the fact that they were easily available in many languages. Whilst comparability would have been a desirable outcome, the requirements for such analysis to be robust are probably not met for these types of comparisons to be made.

Many recent research efforts are aimed at measuring non-cognitive skills due to the fact that traits such as motivation, aspirations, self-esteem, peer-relations etc. can be related to improvements in outcomes including educational, socio-emotional and labour market outcomes. These could also be hypothesised to have different affects by gender. Projects across the GEC portfolio have collected data on non-cognitive outcomes; however, these are less commonly available in GEC-1 than GEC-T: six projects in GEC-1 and 17 in GEC-T include information of direct relevance to this indicator. It is not

surprising that there is less information on this theme given it is an area for which data are not as commonly collected. As such, it would be of interest to explore in more detail the lessons from collecting these data for future studies.

GEC-1 projects that have collected data are mainly related to marginalised girls' self-esteem (for example, "Securing Access and Retention into Good Quality Transformative Education" (ChildHope) and "MGCubed" (Varkey Foundation). GEC-T data on non-cognitive outcomes appear to be richer and more nuanced particularly on aspects such as girls' aspirations, peer-relations, life-skills and self-confidence (for example all the "Discovery Projects" (Discovery Communications) in Kenya, Ghana etc; "Innovating in Uganda to Support Educational Continuation by Marginalised Girls/Girls Education Finance: Empowerment for Girls' Education" (Opportunity International)). *These data would allow an exploration of the patterns of marginalised girls' non-cognitive outcomes and whether they differ by context or other factors (such as socio-economic status where data are available, by disability etc.). Where the research design has comparison control and intervention groups (or where cohort designs have been adopted) research could also examine whether interventions have improved girls' non-cognitive skills*.

Interviews with IPs identified that there was some difficulty in having common tools that can be used to measure certain outcomes that translate effectively across different contexts, or that can be used to inform systems changed. One IP noted their experience of developing a new learning assessment for their evaluation. In this case, it was felt important to develop a learning assessment in line with the national examination council in the country. This meant that the government would be more likely to accept the results, and so be able to use the findings to influence policy, which would not have been possible if they had used the standard FM's guidelines.

Finding theme 5: Gender-sensitive pedagogy

Information on two of the three indicators related to gender-sensitive pedagogy was frequently found to be available for GEC-T projects. However, the data for GEC-1 projects typically provided information on fewer indicators related to this theme.

In addition to indicators of teacher gender discussed under theme 2, information on teacher quality, teaching approaches and gender-sensitive pedagogy (including from classroom observations) was frequently found to be available for GEC-T projects. For example, datasets for the GEC-T project "Supporting Marginalised Girls in Sierra Leone to Complete Basic Education with Improved Learning Outcomes" (Plan International) give information from:

- Questions asked to children concerning whether their teachers were frequently absent from class and if teachers 'treat boys and girls differently in the classroom'.
- Classroom observation (from a dataset devoted to this topic).

However, the data available for this project, like a number of other GEC-I and GEC-T projects was not identified to provide information of direct or partial relevant to the indicator, 'Perceptions on gender equality in schools.' It should be noted that this aspect is very difficult to capture unless it is the specific focus of a given research initiative. Two projects that give some (if limited) information in both GEC-1 and GEC-T on this indicator are "MGCubed" (Varkey Foundation), which includes one variable from a school survey, labelled 'Does this school encourage student-centered, gender-sensitive education' and "Equal Access to Education for Nomadic Populations in Northern Afghanistan Project" (ChildFund) asks whether pupils perceive teachers as treating boys better than girls.

Findings theme 6: Social norms

Indicators relating to social norms were not sufficiently prevalent across the data sets, given their importance for understanding education experiences of marginalised girls. This theme should be a priority for future research.

Theme 6 pertains to the social norms that can play a key role in holding back girls' education that cannot be underestimated (see Section I). Therefore, collecting data on this theme is a vital component of any research effort aiming to identify how to improve the educational outcomes of marginalised girls. This theme includes indicators that cover information on aspects such as early marriage, female genital mutilation, menstruation and child work. Table 7 indicates that nearly every dataset in GEC-T (except two) collected some sort of indicator pertaining to social norms whereas in GEC-1 slightly more than half did, with a majority of these being deemed partial (amber) according to the assessment. Information considered to be partially relevant to theme 6 included variables concerning child work, such as girls' participation in agricultural, family business and house-based work (from the GEC-1 project "Supporting Slum and Homeless Street Girls with Disabilities in Kampala City to Access Quality Primary Education" (Cheshire Services).

One area that could form the focus of future research efforts from existing data would be *an examination of the work children do (particularly girls' domestic work) and the impact that this has on their ability to attend school and learn*. Some GEC projects collect data on girls' domestic work as a barrier to education (attendance) such as "Girls' Enrolment, Attendance, Retention and Results" (PEAS) in GEC-1 and GEC-T, "Securing Access and Retention into Good Quality Transformative Education" (Childhope, GEC-1 and GEC-T) and "Supporting the Education of Marginalised Girls in Kailali District" (Mercy Corps, GEC-T). Several projects would allow descriptive statistical analysis of this theme, however, some of the richer and higher quality data sets could allow for further, more robust, econometric techniques to be implemented.

Other social norms that could be explored due to their potential relationship to whether girls attend school relate to aspects such as early marriage and female genital mutilation. Although these indicators have not been collected extensively in the quantitative data, there is some scope for analysis exploring the relationship between these aspects and girls' outcomes particularly from the GEC-T project data. For example, datasets for the project "Improving Girls' Access through Transforming Education (IGATE)" (World Vision) provided information on the age of marriage, and child work (as a reason for school absence). Related to this, datasets for a selection of other GEC-T projects provided information on perceptions about the importance of secondary education/marriage. For example, perceptions on whether 'It is better for girls to get married than complete secondary education' were captured from a household survey used to assess the GEC-T project "Community Based Education for Marginalised Girls in Afghanistan" (BRAC). Given that many girls in GEC-T were older, and potentially at an age when they might drop out to get married or give birth, future rounds of GEC-T could provide an opportunity to identify which girls are most affected by this, provided those out of school are tracked. More limited information is available in the GEC-1 projects. For example, data for "Pastoralist Afar Girls' Education Support Projects (PAGES)" (Save the Children) included two variables on the perceptions of girls' marriage: 'Girls should be married once they go through menarche (first menstruation)'; and, 'Absuma marriage is very important for the wellbeing of girls and women.'

Little information was identified on female genital mutilation. No variables relevant to this topic were identified amongst selected GEC-1 datasets. Similarly, the term was only found to feature as a potential response option in the data for one GEC-T project, "Wasichana Wetu Wafaulu" (Education

Development Trust). In this project, households were asked '… which violence/harms against children are you aware of in this community?' Household response choices included female genital mutilation, which was presented in the data alongside other options including teenage pregnancy, child marriage and child labour.

Findings theme 7: Transitions

Information on transitions has been collected in a specific dataset across the majority of the GEC-T datasets. However, it is not possible to actually track children across GEC-1 and GEC-T datasets in general.

Research has indicated that girls' dropout rates tend to peak during the adolescent years and particularly at key transition points such as between primary and secondary school (see Section I). The GEC-1 projects provided no information of relevance to theme 7 in all but one instance. This lack of coverage was perhaps because children were younger at this stage. As most GEC-T projects target girls at adolescence, collecting data on transitions in GEC-T projects is important. GEC-T datasets generally provided information on two key indicators pertaining to this theme: 'girls beginning work' and 'transition rates'. Transition rates are examined following the definition of transition used in specific projects (including grade progression), while excluding any information on intention to transition (namely with respect to aspirations). All GEC-T datasets provide some sort of information on these two indicators. For example, the "Step Towards Afghan Girls' Education Stages (STAGES)" (Aga Khan Foundation) gave work-based information through an entire dataset giving information on transitions. This dataset includes variables based on head of household responses concerning whether the girl is enrolled in school, was recently enrolled, is in employment, or is participating in non-formal training.

In addition to data on transition rates noted above, some projects also collect information about girls' aspirations relating to work once they complete schooling. This would potentially allow analysis on the extent to which aspirations are associated with actual outcomes for girls from different backgrounds, for example. For example, the "Community Based Education for Marginalised Girls in Afghanistan" (BRAC) asks the girls about their ambition after completing schooling and the type of work they would aspire to do. Another project "MGCubed" (Varkey Foundation) asks the caregivers their perceptions of what types of work the girls might do after completing schooling. *These types of data could, therefore, allow researchers to examine patterns in girls' transition rates and could potentially allow for the exploration of reasons behind these patterns*.

Findings theme 8: Boys

It is not generally possible to examine boys' attainment and engagement with GEC projects.

The learning outcomes-focused datasets in GEC-1 do not typically include boys. This was a gap recognised by some IPs in our interviews with them. For example, one IP noted that although not including boys in the evaluation was a conscious choice, linked to the resources available to conduct this evaluation, it meant they were not able to make any comparisons on how boys have progressed in comparison to girls, which was seen as a shortcoming. Another IP also noted that attendance spot checks only looked at girls, and not boys, taking a comparison from one year to the next.

More GEC-T projects gave information on boys, with relevant child/adolescent-focused datasets identified for eight projects. Even where they do include data on boys, the proportion of boys in the samples are generally far smaller, potentially limiting analysis. One example of a dataset that includes information on boys is GEC-T project "Educate Girls, End Poverty" (Relief International). For this dataset, only 398 of 2919 children are male.

A key recommendation of this assessment is for future data collection efforts to address this limitation by collecting data on boys either as indirect beneficiaries of programmes or to understand some of the unintended consequences of programme implementation on boys.

Findings theme 9-11: Programme change, system change and cost-effectiveness/VfM

Indicators on themes 9-11 were particularly limited across the GEC portfolio.

There was comparatively little information on themes 9-11 compared to other themes within the available datasets. Indeed, information relevant to themes 9 and 11 (programme change and cost-effectiveness/VfM) was not identified for almost all projects (across GEC-1 and GEC-T). *The collection of new qualitative or quantitative data might be most appropriate for furthering understanding of programme change, system change and cost-effectiveness/VfM, given that existing datasets do not include much information relevant to these themes***.** It might also be possible for future researchers to draw on pre-existing data through communication with IPs. This was evidenced by further analysis of the "A New Equilibrium for Girls" (CAMFED) project in Tanzania, which included cost-effectiveness findings based in part on cost data sourced from CAMFED (Sabates et al., 2018). In addition, IPs noted during our interviews with them that information on how data have influenced their programme, as well as their influence on national education systems was sometimes available in other documentation such as their monitoring reports and other reports to the FM. They proposed that it would be valuable for researchers to access this information as a basis for analysis on these topics.

<u>Overarching comments</u>

- The GEC portfolio has generally collected some good quality data on a range of indicators that could be used to evaluate project-specific outcomes, more notably in GEC-T datasets. Whilst there are some exceptions and variation across projects, there are several projects that provide very rich and high-quality data. However, several project datasets suffer from limitations due to poor data collection, management and processing (e.g. cleaning, labelling, how missing variables are treated etc.).
- The GEC portfolio, through the richness of its quantitative data, allows many opportunities to explore various themes pertaining to girls' education, particularly those who may be the most marginalised. However, an important caveat is that there are limitations of the extent to which generalisations can be made across contexts (as is the case with other research using similar methodology).
- Whilst GEC-1 data allows for investigation of patterns that emerge across the themes, the richness of GEC-T data allows for more nuanced exploration.
- In GEC-1, themes 1 (equality and equity), 2 (access and attendance) and 3 (learning) are generally well covered. GEC-T also contains information on these themes 1, 2 and 3.
- GEC-T projects also have information on themes 4 (non-cognitive skills), 5 (gender-sensitive pedagogy), 6 (social norms) and 7 (transitions).
- For the themes where there is information in both GEC-1 and GEC-T, the latter provides more in-depth information with improvements in both the number of variables data are collected on as well as the quality of the resultant data. GEC-T projects collect a wide range of rich data from a variety of instruments (e.g. more usage of classroom observation tools).
- Themes 8 (boys), 9 (programme change), 10 (system change) and 11 (cost-effectiveness/ VfM) have very limited data across all projects in the entire GEC portfolio. These all deserve to be a focus of future data collection as part of a research programme.

Recommendations on projects which have data for further analysis that is of good quality, and covers important thematic areas, can be found in Appendix 3.

Assessment of the qualitative data

Introduction

This section presents the findings from the DQA process conducted on the qualitative data from GEC-1 and GEC-T. As noted at the outset, this assessment does not seek to determine the quality of the data based on its original purpose (project evaluation) but rather aims to assess the quality of the data for potential future research.

We had access to transcripts from 11 of the 35 projects for GEC-1, which we assessed using the BE2 guidelines (see Section II). For GEC-T, we had access to transcripts for 11 of the 29 projects. For these projects, a full assessment was completed.

It should be noted that while our assessment scrutinised all transcripts provided by the projects to the extent that they were made available, access was not provided to all transcripts for these projects. This was often due to logistical constraints in evaluators having the time or budget to fully transcribe and translate all interviews. In other cases, sample transcripts were provided due to difficulties in anonymising the research, making it less common for the qualitative data to be shared. Further analysis of qualitative data could be done in some cases, but it would require in depth engagement with the IP and external evaluator, and potentially funding to support them to anonymise further transcripts before they can be shared.

While there was less information available to us for a full assessment of qualitative data, its value is evident from our interviews with IPs. These highlighted that the outcome-level data requested by GEC often could not provide an understanding of the multi-faceted and complex nature of the projects, and the impact that this might have. According to one IP: 'among the 18 different interventions we were doing, it was hard to know which one was actually making the difference'. Projects mentioned that this had large limitations; understanding whether literacy and numeracy scores had improved was interesting, but when they were operating holistic programmes that tackled challenges for girls' learning from numerous different angles, they could not dissect the findings further to understand which aspects helped the most, which aspects of the project should be scaled up and continued, and so on. Another IP further noted that with a broad project with many different activity streams, it was difficult to see what the specific impact of the different programme components were, and that quantitative data could only draw correlations.

In some IP interviews, frustration was expressed that the evaluation was not sufficiently designed to answer both ''*what worked'* and '*why*' and some projects followed up independently on this, requiring time and resources to do so. Although the evaluations had the potential to provide lessons about why things work in some contexts but not others: 'you lose the nuance you would need to understand why this specific project working with rural girls in, let's say [country X], has fantastic results while this organisation in [country Y] "didn't hit their targets."' There were concerns that in certain cases the evaluations were not able to get to grips with the specific conditions which enabled or prevented programmes from working well. This could be uncovered by further qualitative research and continuous process-based data. It was also noted that qualitative research was not prioritised as part of GEC-1, which made it more difficult to understand the findings from other areas of the evaluation.

There are instances where qualitative data that were available were seen as beneficial by IPs interviewed. One IP considered the qualitative data to have played a fundamental role in them being

able to understand the factors behind the learning outcomes they had seen. From their focus group discussions, they were able to see that girls were particularly struggling and incorporated monitoring data to design activities to support these girls. Another also found that there was great value in the qualitative data to really understand the lived experience of one girl, or groups of girls. One also found that the data enabled a better understanding of the cases of corporal punishment that were highlighted, by giving insight into the severity of it, along with identifying issues of harassment, and perspectives on solutions to this.

Completing DQA forms using dataset and evaluation report information

Using the BE2 guidelines, the quality of the qualitative data for GEC-1 and GEC-T was assessed. A table was created based on the questions provided by these guidelines and this was used to code the available data as being of good quality (green), of sufficient quality (amber) or poor quality (red) for the purpose of future research (see Table 4).

Based on the assessment of the data we had access to, which provided the ratings above, each project was given an overall rating, based on the ratings across different categories. Overall, no GEC-1 projects assessed were rated green based on their potential for future research from the data that we had access to, although five projects: "Community-Based Education for Marginalised Girls in Afghanistan" (BRAC), "Equal Access to Education for Nomadic Populations in Northern Afghanistan Project" (ChildFund), "Pastoralist Afar Girls' Education Support Project (PAGES) (Health Poverty Action), "Securing Access and Retention into Good Quality Transformative Education" (ChildHope), and "Improving Girls' Access through Transforming Education (IGATE)" (World Vision) were rated green/amber, indicating potential for future research using these data.

The DQA found that the quality of the qualitative data improved between GEC-1 and GEC-T. Five of the 11 GEC-T projects assessed were rated green overall, indicating that their qualitative data was of good quality. However, more generally, there were some concerns with the quality of the data (for the purpose of further research), detailed below.

The DQA demonstrated that generally qualitative data is not underpinned by sufficient consideration of the existing evidence base, and the aims of the research (when detailed) are usually not explicitly based on existing research and evidence gaps.

Reviewing available research prior to the research design is important for determining what is already known and what the gaps in the evidence are. However, we acknowledge that the GEC evaluations were not necessarily designed in a way to contribute to filling the gaps in the evidence base. All of the 11 GEC-1 projects assessed had some evidence of context-setting and review of the literature on girls' education in the context of the project. The majority of GEC-T projects also had some evidence of context-setting and review of the literature prior to their research design, although this was not extensive. One project, "Reussite et Epanouissement Via L'Apprentissage et L'Insertion au Systeme Educatif" (Save the Children/World Vision) had strong evidence of formative research informing their approach to the evaluation. They conducted a 'PESTLE' analysis, which is a tool used by organisations to track the environment they are operating in, to fully understand the barriers for girls' education in the Democratic Republic of Congo.

Table 4: DQA criteria analysis for GEC-1 and GEC-T

Criteria	Sub-theme	GEC-1			GEC-T			All GEC		
		Green total	Amber total	Red total	Green total	Amber total	Red total	Green total	Amber total	Red total
Prior to the Research Design	Evidence of formative research	0	11	0	1	10	0	1	21	0
	Aims of the research based on the existing evidence base	3	8	0	1	10	0	4	18	0
	Justification of research questions	4	7	0	11	0	0	15	7	0
During the planning phase	Justification of methodology to answer research questions	1	9	1	6	5	0	7	14	1
	Sufficient consideration of methods	4	6	1	9	2	0	13	8	1
	Paradigms and considerations articulated and scrutinised	0	0	11	0	11	0	0	11	11
	Protocols following clear description of study purposes and main research questions	5	6	0	9	1	1	14	7	1
	Evidence of piloting	0	2	9	1	3	7	2	4	16
During the data collection phase	Description and justification of sampling	3	8	0	1	10	0	4	18	0
	Data collection protocols – identify participants, and how processes will be undertaken?	6	5	0	10	1	0	16	6	0
	Information about the training of the team who undertook the data collection – any information about positionality and reflexivity?	2	9	0	0	11	0	2	20	0
	Was the timeframe for data collection adequate for the purpose of the research?	2	7	2	4	7	0	6	14	2
	Evidence of a thorough data log	6	1	4	8	3	0	14	4	4
Ethics	Did the research clearly follow ethical guidelines including the 5 Rs?	5	6	0	9	2	0	14	8	0
	Were the harms and benefits of the research adequately considered?	0	8	3	1	6	4	1	14	7
	How did the research team ensure privacy and confidentiality?	7	3	1	7	2	2	14	5	3
	Were ethics committees and institutional review boards consulted prior to research?	0	4	7	0	1	10	1	4	17
	How have dissemination processes been decided upon?	0	1	10	1	0	10	0	2	20
	Have ethical processes and decisions been communicated openly and transparently?	0	5	6	2	0	9	2	5	15
	Is there evidence of member checking and/or peer debriefing?	2	0	9	0	2	9	2	2	18

A common problem in qualitative research is to start with ideas about methods, before discussing and deciding on the overall purpose or goal and related assumptions of the research (DeJaeghere, Morrow & Schowengerdt, 2020). A strong linkage between the existing evidence base can provide justification for research design and establish coherence between elements of the study and its relevance to the purpose and context. Based on the data we had access to, for GEC-1 three project evaluations clearly and explicitly linked the aims of their research to the existing evidence base, but only one GEC-T project evaluation did this: "Educate Girls, End Poverty" (Relief International). This evaluation report detailed the hypotheses of the project, making a clear linkage between the research evidence and the programme design, and subsequently between the programme design and the aims of the research.

The protocols for gathering data generally follow from a clear description of the study purposes and questions guiding the study. However, the DQA has identified that there was not transparent and comprehensive consideration of how the methodology for the research and the methods were chosen.

Even in GEC-1 projects that overall achieved a green/amber rating, the weaknesses in the qualitative data were often related to insufficient transparency of the consideration of how the methodological approach sought to answer the research questions, and thus whether the methodology would be appropriate for further analysis. Only one project in GEC-1 provided a clear justification of the methodology chosen to satisfy the purpose of the research in the documentation we had access to: "Pastoralist Afar Girls' Education Support Project (PAGES)" (Health Poverty Action). However, this improved in GEC-T, when six projects clearly justified the choice of methodology to answer the aims of the research, although none of the evaluations in either GEC-1 or GEC-T articulated and scrutinised the paradigms of the research approach in-depth in the documents we had access to.

Consideration of specific methods that were used as part of the research was given a green rating in four projects in GEC-1 and nine projects in GEC-T. Particularly in GEC-T, baseline reports often included a table that outlined each method choice and their rationale for answering the research questions. However, whilst the rationale for the methods to answer the research question is considered, there were limited justifications for why the specific methods would respond to the research question more effectively than other research methods.

Research protocols generally followed from a clear description of the study purposes and the main questions guiding the study. The majority of GEC-1 and GEC-T projects scored green in this area. Generally, this information was provided within research protocols, where the rationale for the interview/focus group was detailed for the enumerator. For GEC-T projects in particular, often the protocols are structured around the key themes and sub-themes of the research, which were tailored to each participant group.

Whilst the protocols overall appeared to be well-designed to answer the main questions guiding the study, there were very few projects that mentioned whether they conducted pilot-testing to ensure that this was the case. None of the GEC-1 projects we assessed mentioned conducting a pilot, and only one GEC-T project, "Successful Transition and Advancement of Rights for Girls (STAR-G)" (Save the Children) scored 'green' in this area for both detailing the pilot test of the qualitative tools, as well as providing information on how this led to re-ordering or altering of the questions/tools.

The DQA noted that samples were clearly identified but that in many evaluations the reason for the sample was not provided, and in the majority of evaluations the sample size was not justified.

Information on samples was provided for all project evaluations. To facilitate further analysis of the data, it would be important for decisions about the sampling design to be provided. This was justified in some cases, based on overall research aims. For example, the GEC-1 endline report for "Community

Based Education for Marginalised Girls in Afghanistan" (BRAC) noted that they modified the sampling design to ensure that they could undertake the same type of interviews in each sampled community in order to allow for comparative analysis. However, there were a few projects in GEC-1 where the choice or approach to sampling was not justified.

Only one project as part of GEC-T scored green in relation to clarity of sampling decision-making and justification. Additionally, based on the information we had access to, no projects in GEC-1 or GEC-T explained their choice of sample size. This would make it difficult to ensure that re-analysis of this data could provide good quality insights, as sampling decisions will affect how accurate and robust the data collected was.

The DQA has revealed that both GEC-1 and GEC-T projects on the whole provide good information regarding the process of data collection. GEC-T projects also provide good information with regards to data management. However, detail on researcher training is inconsistent.

For GEC-1, six projects scored green in relation to information included in their data collection protocols and six projects had comprehensive and transparent data logs. However, four projects were rated red in relation to data logs, due to insufficient information provided about the date and location where data were collected, format of the data, duration, researcher responsible, original language and critical identifiers from the respondent. This might raise questions as to how data were recorded and stored, but this had improved in GEC-T with ten out of the 11 projects assessed scored green in relation to the information included in their data collection protocols.

All projects in GEC-1 provide some information about the training of enumerators, however there is a lot of inconsistency about the quality and the detail of this training. For example, some projects did not detail enumerator training on the qualitative tools, whilst others such as "Pastoralist Aar Girls' Education Support Project (PAGES)" (Health Poverty Action) provide information on the fact that training was undertaken to familiarise participants with the research instruments and qualitative data collection techniques such as probing. None of the projects detailed training that included reflections on researcher positionality or reflexivity. This level of information and inconsistency is similar for the GEC-T projects assessed.

Related to the data collection phase, one of the main weaknesses was the timeframe of the data. Overall, the data collection period was quite short (for example for GEC-1 this appeared to range from five days to one month), and multiple projects as part of GEC-1 and GEC-T identified that due to timing and budgetary constraints, they conducted qualitative data collection at the same time as quantitative data collection. Given that many of the explicit purposes noted for the qualitative research was to focus in more depth and unpack the findings from the quantitative research, this is a clear weakness as the research design was not based on the findings from the quantitative data collection.

The qualitative data assessment identified a general lack of information on research ethics.

Based on the information that was available to us, across GEC-1 projects, only five scored green on whether their research clearly followed ethical guidelines, including the 5 Rs (relationships, respect, relevant, responsibility and reciprocity) and six scored amber. Even in the projects which scored green ethical considerations were not detailed in depth, and although the notions of informed consent and anonymity were present, how these would be upheld in practice was rarely identified. Additionally, researcher positionality and how this might influence power relations within the research process were not detailed in any of the documentation we had access to. For GEC-T, there is a better consideration of ethics throughout, with nine projects scoring green on whether their research clearly followed ethical guidelines.

Only one project in GEC-1 and none in GEC-T where we undertook a full assessment identified that they obtained institutional review board approval for their research, indicating that there was no external assessment of the ethical implications of their research. Although we recognise that institutional review boards are more common for academic research, the lack of information on an independent review of ethics decisions and processes shared is a cause for concern particularly given the data collection often involved the involvement of potentially vulnerable girls and their households. From the review of some data protocols, the information provided about the aims of the study and the role of the participant were not sufficiently detailed to enable an assessment of whether they could be considered ethical.

There was only one GEC-1 project which transparently considered both the risks and benefits of the research to participants, "A New Equilibrium for Girls" (CAMFED), but this was not one of the projects included for the full assessment, due to a lack of access to transcripts. Therefore, of the projects fully assessed for GEC-1 and GEC-T no project appeared to consider potential benefits to participants from the research. Overall, no project demonstrated wide-reaching and in-depth reflection on ethics, with most focussing solely on child protection and considerations of informed consent and anonymity, and only during the data collection and not throughout the analysis and dissemination processes.

The review of the quality of the data as part of the DQA has identified that there are numerous examples of in-depth and comprehensive transcripts.

For GEC-1 projects, out of the 11 projects assessed, there are a number of projects with well-described data, which makes it clear who is speaking and what their distinct characteristics are. In addition, some case studies provide a detailed and in-depth consideration of marginalisation (e.g. "Community Based Education for Marginalised Girls in Afghanistan" (BRAC)). However, whilst the transcripts we had access to in GEC-1 projects provide interesting insights and information overall, generally more information would be needed to enable a thorough examination of the pathways through which the project supported positive changes in girls' lives. One notable exception is "Equal Access to Education for Nomadic Populations in Northern Afghanistan Project" (ChildFund).

Overall, the review of GEC-T transcripts that were available found that they are of good quality. For example, focus group discussion (FGD) transcripts clearly show differences in answers from different respondents (and provide clear delineation). Enumerators appeared to be well-trained to provide strong probes, although FGDs functioned more as group interviews than as discussion amongst the group, with the facilitator providing framing questions and probes. The strongest examples of transcripts that provided good depth on the key themes of the research were "Girls' Education Finance: Empowerment for Girls' Education" (Opportunity International), "Educate Girls, End Poverty" (Relief International) and "Rwandan Girls' Education and Advancement Programme (REAP)" (Health Poverty Action). Another project with a comprehensive and in-depth set of transcripts was "Successful Transition and Advancement of Rights for Girls (STAR-G)" (Save the Children).

Potential themes for future research using existing qualitative data

Searching through datasets for topics of interest

The review of the qualitative data involved searching for information on the presence of data relevant to the key themes of interest, where possible. This was predominantly identified through a review of the transcripts. However, in some cases where there was only access to a sample of transcripts, but where the quality of these transcripts was good, this review also involved looking at qualitative research tools, and qualitative findings (quotations) within endline (for GEC-1) and Baseline (for GEC-

T) reports. Therefore, although there is an assumption that the theme of interest would be covered by the data, it cannot be confirmed. All projects appear to have useful information on the barriers to girls' education, however much of this has already been captured in existing evidence (see Section I). This is particularly the case for GEC-T, as the data are not explicitly exploring the results of the interventions; these are not provided in detail under the themes below. Rather, the search was related to the themes identified in Table 1 with respect to areas that would warrant further analysis and potentially fill gaps within the existing evidence base.

Findings theme 1: Equity and equality

Qualitative datasets for GEC-1 do not appear to provide a vast amount of information directly related to intersecting forms of disadvantage.

There are some insights from the qualitative GEC-1 data on certain sub-themes on intersecting forms of disadvantage that girls face. For example, "Discovery Project Nigeria" (Discovery Communications) has qualitative data which provides insight into the influence of Ebola and rurality on girls' education. Case studies from "Community Based Education for Marginalised Girls in Afghanistan" (BRAC) and "Innovating in Uganda to Support Educational Continuation by Marginalised Girls (EduFinance)" (Opportunity International) could be further analysed to provide insights into marginalisation, and in particular whether the intervention was able to reach the most marginalised girls. However, overall, the qualitative data from GEC-1 and GEC-T we reviewed do not appear to provide extensive detail on intersecting forms of disadvantage.

Findings theme 2: Access and attendance

Available GEC-1 qualitative datasets provide details on which interventions can improve access and attendance for girls. GEC-T datasets provide further insights into barriers.

Firstly, three GEC-1 projects' qualitative data explore whether financial interventions improve access and attendance for girls through their qualitative data. For example, "Community Based Education for Marginalised Girls in Afghanistan" (BRAC), "Improving Girls' Access through Transforming Education (IGATE)" (World Vision) and "Wasichana Wote Wasome" (Education Development Trust). For GEC-T, "Girls' Education Finance: Empowerment for Girls' Education" (Opportunity International), the qualitative data also has information related to understanding how financial interventions influence access and attendance. Other GEC-1 projects collected data which explored changing parental perceptions on girls' education, and thus further analysis could understand in greater detail whether there are specific contexts in which community-based interventions improve access and attendance for girls, such as "Discovery Project Nigeria" (Discovery Communications), "Empowering Pioneering Inclusive Education Strategies for Disabled Girls" (Leonard Cheshire Disability) and "Improving Girls' Access through Transforming Education (IGATE)" (World Vision).

GEC-1 qualitative data also provides insight into resource provision and infrastructural improvements on girls' access and attendance. For example, "Improving Girls' Access through Transforming Education (IGATE)" (World Vision)'s qualitative research explored their interventions related to sanitary pad provision, transport provision and "Girls' Enrolment, Attendance, Retention and Results (GEARR)" (PEAS) transcripts contains information about infrastructural improvements and girls' attendance.

Findings theme 3 and 4: Learning outcomes and non-cognitive skills

There is GEC-1 qualitative data that shows which interventions have had a positive impact on girls' learning and non-cognitive skills.

One way through which further analysis of GEC-1 and GEC-T data could provide insight is through exploring the different contextual factors that might influence how girls' clubs/peer learning affect learning and non-cognitive outcomes. For example, in GEC-1, data on girls' clubs are included in "Securing Access and Retention into Good Quality Transformative Education" (ChildHope), looking at relationships between participation and non-cognitive skills such as confidence and aspirations; "Improving Girls' Access through Transforming Education (IGATE)" (World Vision), whose data explores girls' confidence as a result of girls' clubs; and "Empowering Pioneering Inclusive Education Strategies for Disabled Girls" (Leonard Cheshire Disability) which explored child-to-child clubs and the influence on girls' self-esteem and confidence. There could be linkages created with data from interventions that explored the results of mentoring on marginalised girls' confidence in debating, e.g. "Community Based Education for Marginalised Girls in Afghanistan" (BRAC). In GEC-T, "Girls' Education Finance: Empowerment for Girls' Education" (Opportunity International)'s qualitative data also explores outcomes resulting from girls' clubs related to girls' self-confidence and "Jielimishe" (I Choose Life) could provide a route through which to deepen understanding about the influence of mentoring on girls' non-cognitive skills, as this focus is included in its qualitative research.

Finding theme 5: Gender-sensitive pedagogy

The implications of teacher training for students' experience was noted in some of the qualitative data. However, this is not necessarily explicitly related to gender-sensitive pedagogy. Therefore, there is limited ability of existing qualitative data from GEC-1 to directly address this theme.

Overall, from the transcripts we had access to, there is minimal qualitative data from GEC-1 that deals explicitly with gender sensitive pedagogy and gender responsiveness of teachers in the classroom as a result of training. Inclusive pedagogy, and of girls' perceptions of this, is referred to but without much information on what this means. Protocols demonstrate that teaching and learning is a focus of numerous evaluations. As such, it is possible that transcripts that we were unable to access might include further information that could be explored. This is particularly the case in GEC-T. For example, "Reussite et Epanouissement Via L'Apprentissage et L'Insertion au Systeme Educatif" (Save the Children/ World Vision)'s qualitative data in their transcripts includes information on whether gender sensitive pedagogy is being carried out. Further research could explore its influence on girls' outcomes more broadly, for example. "Educate Girls, End Poverty" (Relief International)'s transcripts include interesting information about teachers' perceptions on their practice as well as about perceptions from the students. There could be further analysis on whether these are aligned, and what this means for gender-sensitive pedagogy being implemented in practice (together with classroom observations, where these are available).

Findings theme 6: Social norms

Some qualitative data from GEC-1 could help to inform an understanding on interventions tackling social norms. GEC-T transcripts have included initial information about social norm change, which could warrant further data collection.

From GEC-1, qualitative data from "Girls' Enrolment, Attendance, Retention and Results (GEARR)" (PEAS) identified how parent counselling has led parents to support girls' re-enrolment after giving birth. Community engagement workshops and radio programmes as part of "STAGES" (Aga Khan

Foundation) and community conversations as part of "Wasichana Wote Wasome" (Education Development Trust) were shown to have led to changing social norms around girls' education. Finally, "Improving Girls' Access through Transforming Education (IGATE)" (World Vision)'s work with male champions and mothers' group interventions were associated with decreasing incidences of violence. Further analysis of these datasets might be able to draw out similarities and differences in the approaches, and how these have influenced their effectiveness in social norm change. For GEC-T, two projects have included initial information about social norm change in detail, "iMlango Transitions" (Avanti Communications) and "Improving Girls' Access through Transforming Education (IGATE)" (World Vision).

Findings theme 7: Transitions

There is very limited information about transitions in the GEC-1 qualitative data that is available.

"Pastoralist Afar Girls' Education Support Project (PAGES)" (Save the Children) has data which showed that the presence of several of the project's initiatives was seen to contribute to girls' enrolment and transition to the second cycle of primary education. This could be further analysed to consider the intersecting factors that influence transition. A few GEC-T projects could also contribute to this analysis, as their qualitative data provides some insight into relatively successful transitions, particularly "Girls' Education Finance: Empowerment for Girls' Education" (Opportunity International).

Findings theme 8: Boys

There are multiple projects which undertook qualitative data collection with boys, particularly in GEC-T, or included questions about impacts on boys. This could enable learning from the interventions targeting girls for the benefits for boys.

Qualitative data from some projects could identify boys' perspectives on how they have benefited from interventions primarily designed to promote girls' education. For example, from GEC-1 interviews with boys were included in the qualitative data samples for: "Improving Girls' Access through Transforming Education (IGATE)" (World Vision), and "Empowering Pioneering Inclusive Education Strategies for Disabled Girls" (Leonard Cheshire Disability). In GEC-T, there appeared to be more examples of qualitative data that focused on the experiences of boys, with five projects including boys in qualitative data samples.

Findings theme 9-11: Programme change, system change and cost-effectiveness/VfM

Overall, there is limited information on programme or system change and cost-effectiveness as part of the qualitative data.

There are some cases where qualitative data include reflections on system change of the interventions. In GEC-1, qualitative data collected by "Empowering Pioneering Inclusive Education Strategies for Disabled Girls**"** (Leonard Cheshire Disability) there is detail on their work to influence government policy. Annex 3 in every endline report for GEC-1 details the adaptations that programmes have taken throughout the project based on evidence, although it is not clear whether raw qualitative data are available in order to undertake further analysis (although document analysis of the reports would be feasible).

Three GEC-T projects could provide insights into the potential for programme and system change. For example, data from "Girls Learn, Succeed and Lead" (CAMFED) demonstrated how the potentially high

cost of the bursaries meant that a more tailor-made bursary package would be beneficial. In "Educate Girls, End Poverty" (Relief International), interviews with Ministry of Education officials provide some early insights into the long-term influence on national education systems more widely which could be followed up on in further research and analysis, for example. Finally, qualitative data from "iMlango Transitions" (Avanti Communications) provides insight into how the programme relates to ongoing/current government programmes and about the influence on daily practices of Ministry of Education officials. A number of IPs identified how evaluation data had been important for their engagement with government officials. For example, one presented their data at a workshop including state level government officials. Subsequently the state bureau of statistics reached out to ask them to help build a tool to conduct an evaluation to understand the true picture of out of school children. Other IPs gave examples when this engagement has led to tangible impacts. One identified that the Ministry of Education adopted the model they were using on mentorship, which launched last year as a national programme. They used the project's data to understand how to improve teacher skills and quality. They have also worked closely with the Ministry to develop re-entry guidelines for pregnant schoolgirls. To understand this in more detail, research could be undertaken with these officials to consider the influence of the project on their practice, for example.

In some cases, there might be other sources of data from IPs that could be drawn upon to inform programme and systems change, notably from their monitoring data. In some interviews with IPs, they mentioned how reporting requirements as part of the GEC stretched their capacity to utilise this information, and thus made it hard for them to fully analyse and share data beyond their internal project staff, but that more information is available that they considered would be valuable for further analysis. These internal monitoring data have already often helped IPs to understand more about the impact of the project. One IP addressed the issue of the evaluation being very broad through their monitoring, using simple feedback mechanisms to get more information about different aspects of the project. Another IP used their monitoring to identify what schools and communities think of the interventions. Supported by a researcher who conducted focus groups, they were able to unpack and understand community reaction and what girls thought of the programme. It helped the project to really learn what specific aspects of their intervention were useful, down to the minor details (e.g. whether the school having a fence made a difference). They felt that this was really important to helping them to decide what they would continue doing. Another IP similarly reflected on their internal monitoring data being essential for assessing themselves against their own key performance indicators and wanting to learn continuously as a programme in addition to the evaluations. They noted that it was complementary to the evaluator reports, and although wouldn't necessarily score highly on an assessment using the BE2 guidelines, it provided some really useful information.

Further research could also identify how the data from the evaluation has been used to influence programme or system change. During our interviews with IPs, some noted how the evaluations findings fed into a report that helped to improve programme quality and designing a new intervention. In another case, baseline evaluation data identified that some of the girls they had been targeting had some form of disability, leading to re-designing their project to include disability-related interventions. One IP reported, for example, that the evaluation provided them with feedback from the community in terms of challenges to the programme not including boys with disabilities. This has led to them updating their programme to support boys with disabilities as part of GEC-T. The evaluation also highlighted accessibility issues, which has led them to make accessibility improvements in schools. It also showed capacity gaps amongst teachers, which enabled developments in GEC-T to focus on improving instruction and pedagogical practices. Through their evaluations, other IPs also uncovered backlash against the specific targeting of girls from their qualitative data, and this also led to them including boys within the project. For one IP, this meant having clubs for boys and girls together, and clubs for boys as well. Initial data from GEC-T supported another IP to refine their project models to better reach girls out of school by extending their girls' clubs to the community. The evaluation of

another project also found that children with special needs were not being involved sufficiently, and this led to the hiring of a social inclusion officer. One IP noted: 'we have made numerous iterations based on evidence from the field. Following the baseline, we included fun learning activities geared towards key competencies that were lacking at baseline. This particular iteration helps to track learning and also understand how learning happens.'

IPs also reported using evaluation data for the design of new programmes and supporting other organisations working on girls' education. For example, one project noted the evaluation had informed the design of new USAID programme and has contributed to the INEE Guidance Note on Gender, and the 2019 Global Education Monitoring Report. Another has used the data to influence their own programming in other countries.

IPs also mentioned using the data from evaluations to share with girls, communities and schools. One noted it held events with headteachers to discuss the results from evaluations. As there are very limited data of this kind at the school level, they felt that sharing data on learning was very powerful, and helpful for headteachers to reflect on what needed to change. Teams trained facilitators to use that data and facilitate at the district level as well as the national level.

Overarching comments

- For both GEC-1 and GEC-T, overall, generally qualitative data is not underpinned by extensive review of the existing literature, so the aims of the research are not based on the existing evidence base.
- Particularly for GEC-1, there is often inadequate justification of the methodology and the methods of the research and how they were chosen to acquire the knowledge required. This has improved at GEC-T, particularly in relation to demonstrating how the methods chosen seek to answer the research aims.
- Projects in GEC-1 and GEC-T consistently provide good detail about the methods that will be used for research, and the samples. However, the majority of projects do not provide justifications for these methods over other research tools, nor a justification for the chosen sample or sample size.
- Overall, ethical considerations were not well documented throughout GEC-1 qualitative data. This is a large concern, and greatly affects the quality of the data when used for further analysis. There is some indication that this has improved in GEC-T, but there are still large gaps in transparent ethical considerations.
- Of the projects for which transcripts were available for fuller assessment, there are good examples of in-depth and comprehensive transcripts. For further analysis, the thematic review has shown that further analysis of GEC-datasets could provide particular insight into theme 2, 'access and attendance' and theme 3 'learning and non-cognitive skills.' Overall, however, whilst transcripts provide lots of interesting insights, for GEC-1 they are predominantly focused on specific outcomes, such as changes in girls' confidence, but not which specific project interventions contributed to this, or why.
- The possibility of using other sources of existing data, such as from the project's own monitoring data, could be considered in particular to inform programme or systems change. While the raw data are unlikely to have the 'rigour' expected for research purposes, reports could be used for documentary analysis, for example, to complement other data sources.

Recommendations on projects which have qualitative data for further analysis that is of good quality, and covers important thematic areas, can be found in Appendix 4.

SECTION IV: Recommendations

Our recommendations focus on identifying potential priority themes and approaches for future researchers that will promote filling of gaps in the global evidence base on girls' education. This is based both on analysis of existing GEC data as well as in terms of collecting new data related to GEC projects. Overall, we identify that there is already rich GEC project data available, some of which is of sufficient quality for further analysis, particularly with respect to GEC-T. Complementing this with additional data collection and analysis will strengthen the depth with which evidence on particular themes can be gained.

It is important to note that we were mainly only able to access GEC-T data at baseline given the timing of this report. It is possible that further information is available from future rounds of GEC-T quantitative and qualitative data collection. As data were not available for the Leave No Girl Behind programme in Phase 2 of GEC, we also do not consider this explicitly in the recommendations, particularly as the approach to data collection for this programme has differed from GEC-T. However, it is possible that some of the recommendations also apply.

Identifying potential research themes and approaches for GEC to contribute to the global evidence base

Our overall assessment is that existing GEC data provide an important resource for future research. The existing quantitative data provide information on a range of issues related to girls' education. While there are variations in the quality of these data, and some uncertainties of what is covered given the complications we faced in accessing final datasets (particularly for GEC-1), we consider that further analysis of these data would be worthwhile. Ideally, this would be complemented by analysis of the available qualitative data. However, we were able to access these data for fewer projects, and so the possibility of their use for future researchers will depend on their availability (recognising that such data may sometimes be more difficult to anonymise and so can be more challenging to share).

After the topic for data analysis has been identified, and relevant datasets accessed, the first step would be for researchers to undertake further analysis of these data. We anticipate that in most cases, it will be highly beneficial to collect additional quantitative and qualitative data. This will allow for greater depth of information on a particular identified theme covered by GEC projects (given existing data tend to address a very wide range of issues, but not in great detail). It would also facilitate analysis of the effects of the interventions over a longer period of time, including at key transition points in girls' lives – such longitudinal analysis is a particular gap in existing literature.

In this section, we identify key themes and areas of geographical focus to pursue that we consider to be particularly beneficial in terms of filling existing gaps in the global evidence base for which GEC projects have information. We further note potential projects that could be drawn upon for the analysis based on their existing data. We then highlight approaches to further data collection that would be beneficial to complement existing information.

<u>Key themes for GEC projects to contribute to the global evidence-base</u>

In Table 5 we identify thematic areas emerging from our review of the quantitative and qualitative data for which GEC could contribute to the global evidence base. These themes are ones for which there is at least some information in existing datasets (examples of potential datasets to be explored are included in the table - see Appendix 2 for project codes used). It is likely that existing data would benefit from being complemented by further in-depth data collection related to GEC projects. The

selection of themes is also informed by key informant interviews as well as interviews with IPs (for further information about the findings from interviews with IPs, please contact the authors of this report).

Table 5: Key themes for further exploration in GEC projects, with examples of selected datasets

Theme	Possible projects with existing quantitative data[12]	Possible projects with existing qualitative data
1. Social norm change: including gender-based violence, early marriage, child work	2 GEC-1 projects: REP, PAS 16 GEC-T projects: EGE, ENG, GEA, IGA, MGA, REP, SOM, STA, NEQ, RRL, DPG, DPK, DPN, GQE, WWW, GDK	3 GEC-1 projects: REP, GQE, IGA 4 GEC-T projects: NEQ, EGE, IMP, REP
2. Health-related: water, sanitation and hygiene, sanitary wear,[13] nutrition, mental health (including anxiety and depression)[14]	5 GEC-1 projects: MGA, PAS, PIE, STA, GDK 25 GEC-T projects: EGE, ENG, GEA, IGA, IMP, KEE, MGA, VAS, MGC, MGS, PAG, REP, RWZ, SFS, SOM, STA, STE, VKG, NEQ, RRL, DPG, DPK, DPN, WWW, EDF	1 GEC-1 project: REP 1 GEC-T project: REP
3. Child protection/safeguarding (including safety)[15]	6 GEC-1 projects: NEQ, MGA, PAS, PIE, STA, GDK 25 GEC-T projects: EGE, ENG, GEA, IGA, KEE, MGA, VAS, MGC, MGS, PAG, REP, RWZ, SFS, SOM, STA, STE, VKG, NEQ, RRL, DPG, DPK, DPN, GQE, WWW, EDF	Not clear from data we had access to.
4. Gender-sensitive pedagogy (including teacher quality)[16]	4 GEC-1 projects: PAS, PIE, MGC, GDK 25 GEC-T projects: EGE, ENG, GEA, IGA, KEE, MGA, VAS, MGC, MGS, PAG, REP, RWZ, SFS, SOM, STA, STE, VKG, NEQ, RRL, DPG, DPK, DPN, GQE, WWW, EDF	No GEC-1 projects 2 GEC-T projects: VAS, EGE
5. Transitions within and beyond education[17]	0 GEC-1 projects 22 GEC-T projects: EGE, ENG, GEA, IGA, MGA, VAS, MGC, MGS, PAG, REP, RWZ, SFS, SOM, STA, STE, VKG, NEQ, RRL, DPG, DPK, DPN, EDF	1 GEC-1 project: REP 4 GEC-T projects: NEQ, VAS, EDF, REP
6. The role of interventions on girls' education for promoting national system change	1 GEC-1 project: STA 2 GEC-T projects: STA, SOM	1 GEC-1 project: IGA 2 GEC-T projects: IMP, REP

For research on all of these themes, we would recommend that:

- the starting point is to frame research questions based on knowledge in the existing evidence base
- focus in particular on the effects of GEC interventions, with a particular focus on how they are addressing barriers in the specific contexts in which projects are being undertaken
- consider both the impact (from quantitative data) and process (from qualitative data) of changes in outcomes due to the intervention
- the focus is on marginalised girls in particular, with analysis potentially having a specific focus on particular groups, for example girls with disabilities or those from mobile populations such as pastoralists, where sample sizes allow
- across all themes, the effects of interventions on both boys and girls should be considered, where sufficient data on boys are included
- the analysis should focus on outcomes related to both access/retention as well as learning, and where possible also non-cognitive skills and gender empowerment.

Geographical focus

Given there is a concentration of some projects in particular locations, these could help to inform programmes and strategies on girls' education in these contexts. Examples of this are provided in Table 6. The same recommendations as indicated for thematic areas mentioned above also apply to any research adopting a geographical focus.

Table 6: Key areas of geographical focus for further exploration in GEC projects, with examples of selected datasets

Geographical context	Possible projects with existing quantitative data	Possible projects with existing qualitative data
East Africa[18]	14 GEC-1 Projects: DPK (Nairobi), DPK (Wajir), EEP, EGE, GDK, IMP, NEQ, PAS, PIE, REP, RWZ, SOM, VKG, WWW 16 GEC-T projects: DPK (Nairobi), DPK, EDF, EGE, GDK, GEA, GQE, IMP, KEE, NEQ, REP, RRL, RWZ, SOM, VKG, WWW	2 GEC-1 projects: REP, GQE 6 GEC-T projects: NEQ, SOM, EDF, EGE, IMP, REP
Conflict-affected – e.g. Afghanistan, Somalia	4 GEC-1 projects: EGE, MGA, SOM, STA 4 GEC-T projects: EGE, MGA, SOM, STA	2 GEC-1 projects: MGA, NPA 2 GEC-T projects: SOM, EGE
Natural disasters e.g., Ebola, drought[19]	4 GEC-1 Ebola-context projects: DPN, EEP, GDK, VAS 7 GEC-T Ebola-context projects: DPN, EDF, ENG, GDK, GEA, MGS, VAS 0 GEC-1 drought-context project 1 GEC-T drought-context project: PAG	1 GEC-1 project: IGA 1 GEC-T project: EGE
Rural/urban slums/other urban[20]	4 GEC-1 projects: DPG, DPN, DPK_N, EGE 11 GEC-T projects: EGE, KEE, VAS, MGS, REP, STE, VKG, DPN, GQE, WWW, EDF	This was not explicitly searched for in qualitative datasets, but for projects assessed in rural/urban locations, there is likely to be information.

Criteria for selection of existing data to include in identified thematic/geographical research studies

1. Use GEC-T data as a starting point

In both the quantitative and qualitative datasets, there is a step-change in the accessibility and potential quality of the GEC-T data. We therefore propose that, for any research study, this is used at the starting point.

However, once particular projects have been selected, we would recommend that the GEC-1 data is also reviewed for possible inclusion in the analysis. This will give a longer timeframe for addressing the issue under review, which is rarely available. Once it is clear what specific issue the GEC-1 data for a particular project are seeking to identify, it could be easier to locate the required information (provided final datasets can be identified). There could be a small number of cases where GEC-1 data can be used as the starting point.

2. Prioritise projects identified as having higher quality data

For quantitative data, those that are identified as amber or green for the majority of indicators of research design in Table 2 should be used as the starting point, it will be important even so for researchers to look carefully at the data and accompanying documentation to make final judgements on relevant data to use.

For qualitative data, those identified as green or green/amber overall could be used as the starting point. This includes five projects for GEC-1 and seven projects for GEC-T for which we had access to transcripts at the baseline. If transcripts are made available for additional projects, these could also be considered using the criteria (or other relevant criteria) for assessing the quality of the data outlined in Section II of this Report.

3. Ensure that the IP is engaged in supporting the research

Even if data are accessible, the IP needs to engage with researchers to ensure the correct datasets to be used for the analysis are identified, data are being properly understood, research questions being framed appropriately, and interpretations linked with the analysis are relevant.

As such, it will be important to establish upfront that the IPs are interested in engaging in the research process and are willing and have time and capacity to support it. From our engagement with IPs to date, some are extremely engaged and would be keen to be involved in further developing the global evidence base, while others are less engaged or indicate they have no time for this.

In addition, given the multiple challenges faced during the data collection for each evaluation (e.g. security), and on the complexities on collecting certain types of data (e.g. ethnicity) it will be essential to learn from the IPs, and those involved in the data collection to understand how they navigated some of these challenges.

Using project's internal monitoring documentation

The possibility of using other sources of existing data, such as from the project's own monitoring data, could be considered in particular to inform programme or systems change. While the raw data collected for monitoring purposes are unlikely to have the rigour expected for research given this was not the intention, reports could be used for documentary analysis, for example, to complement other data sources.

Methodological approaches for additional data collection and research

The GEC programme potentially provides a unique opportunity for tracking progress longitudinally over time, given the length of the programme. Unfortunately, it does not seem currently possible to track girls who participated in GEC-1 through to GEC-T, as the sampled girls differed across the two phases. This limits the possibility for longitudinal research, which is crucial for analysis of gender and social norm change over time. If at all possible, collecting additional data that tracks girls in GEC-1 to identify what they are now doing would be extremely beneficial. This could be extended to tracking girls who are now out of school, which would provide important information on the particular barriers they have faced. There will be limitations for researchers in this approach. For example, tracking the control group retrospectively will be more difficult, and there are ethical concerns.

Research questions need to be informed by the existing evidence base and the particular features of the GEC projects. In most cases, research questions associated with the key themes will benefit from mixed methods approaches – to understand what is working, as well as why and how. Existing GEC data can be useful for this to some extent. However, it is likely that further data collection will be beneficial to probe into specific themes in greater depth. Where new data collection is involved, it will be important to consider the purpose of the qualitative data – is this to inform the patterns in the quantitative data? If so, data collection needs to be sequential – starting with the quantitative analysis, followed by initial analysis to identify patterns which are then used to inform the questions to be asked in the qualitative phase, as well as the sampling etc.

There then needs to be better consideration of the choice of qualitative methods for answering the identified research questions. For the qualitative approaches, this could include a wider consideration of approaches, such as participatory and visual methods, among others, to engage children and young people.

Ethics

Any research needs to be explicit about how it has addressed ethical issues prior to, during and after the research. This is vital in any study, but even more so with research such as GEC that involves children who are marginalised. There are standard processes that universities adopt to ensure ethical clearance is approved, which can be drawn upon – notably drawing on the British Educational Research Association (BERA) guidelines. These processes need to ensure independent review to ensure ethical standards are met. Beyond these formal guidelines, there is a need for explicit write up on the process of designing and conducting the research with potentially vulnerable populations.

References

Alcott, B., Rose, P., & Sabates, R. (2016) *Targeted multidimensional approaches to overcome inequalities in secondary education: Case study of Camfed in Tanzania*. *Background Paper, The Learning Generation*. [Online] The Education Commission. Available from: https://www.repository.cam.ac.uk/bitstream/handle/1810/262420/Targeted-Multidimensional-Approaches-%20%283%29.pdf?sequence=1&isAllowed=y. [Accessed 20 August 2020].

Asadullah, M.N., Alim, M.A., & Hossain, M.A. (2017) Enrolling girls without learning: Evidence from public schools in Afghanistan. *Development Policy Review,* 37(4), p.486-503.

Bartlett, L., Dowd, A.J., & Jonason, C. (2015) Problematising early grade reading: Should the post-2015 agenda treasure what is measured? *International Journal of Educational Development,* 40(C): 308-315.

Building Evidence in Education (2015) *Assessing the strength of evidence in the education sector.* [Online]. Available from: https://www.usaid.gov/sites/default/files/documents/1865/BE2_Guidance_Note_ASE.pdf. [Accessed 20 August 2020].

Carr-Hill, R. (2018) Analysing girls' educational performance in developing countries. *Compare,* 48(6): p.1010-1014.

Carter, E., Rose, P., Sabates, R., & Akyeampong, K. (2020) Trapped in low performance? Tracking the learning trajectory of disadvantaged girls and boys in the Complementary Basic Education programme in Ghana. *International Journal of Educational Research* Vol 100.

Coffey (2015) *Baseline Report – Step Change Window*. [Online] Available from: iati.dfid.gov.uk/iati_documents/5782584.pdf. [Accessed 27 August 2020].

DeJaeghere, J., Morrow, V., Richardson, D., Schowengerdt, B., Hinton, R., Muñoz Boudet, A. 2020. *Guidance Note on Qualitative Research in Education: Considerations for Best Practice*. [Online] DFID. Available from: https://www.youthpower.org/resources/be2s-guidance-note-qualitative-research-education-considerations-best-practice. [Accessed 27 August 2020].

Eron, L.J., & Emong, P. (2017) Tackling education of girl child with disability in urban settings of Uganda. In Pink, W.T., & Noblit, G.W. (Eds.) *Second International Handbook of Urban Education.* Switzerland: Springer International Publishing.

Evans, D., & Yuan, F. (2019) *We can learn a lot about improving girls' education from interventions that don't target girls.* [Online] Centre for Global Development. Available from: https://www.cgdev.org/blog/we-can-learn-lot-about-improving-girls-education-interventions-dont-target-girls. [Accessed 20 August 2020].

Gordon, R., & Rose, P. (*forthcoming*) Education and Livelihoods for Marginalised Adolescent Girls, University of Cambridge.

Gordon, R., Marston, L., Rose, P., & Zubairi, A. (2019) *12 Years of quality education for all girls: A Commonwealth perspective.* [Online] University of Cambridge: REAL Centre. Available from: https://www.educ.cam.ac.uk/centres/real/downloads/Platform%20for%20Girls/REAL%2012%20Years%20of%20Quality%20Education%20for%20All%20Girls%20FULL%2084pp.pdf. [Accessed 20 August 2020].

Griffiths, S., & Batran, M. (2017) Endline evaluation report, strategic partnerships window. [Online] Coffey International Development. Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/676562/GEC-EM-SPW-Endline-Evaluation-Report-Final.pdf. [Accessed 20 August 2020].

Griffiths, S., Bisiaux, R., & Di Paolo, E. (2017). *Endline evaluation report innovation window.* [Online] Coffey International Development. Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_d

ata/file/676557/GEC-EM-IW-Endline-Evaluation-Report-final-report.pdf. [Accesses 20 August 2020].
Griffiths, S., Poli, F., & Amili, S. (2017) *Endline evaluation report step change window.* [Online] Coffey International Development. Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/676532/GEC-EM-SCW-Endline-Evaluation-Report-final.pdf. [Accessed 20 August 2020].
Hills, L. (2017) Doing the basics better in Africa: How school support, autonomy and accountability improved outcomes for girls in PEAS schools. *Childhood Education,* 93(5), p.382-387.
Hoffman, J.V. (2012) Why EGRA – a clone of DIBELS – will fail to improve literacy in Africa. *Research in the Teaching of English,* 46(4), p.340-357.
Holden, J., & Patch, J. (2017) *Does skin in the game improve the level of play? The experience of Payment by Results (PbR) on the Girls' Education Challenge (GEC) programme.* [Online] Girls' Education Challenge. Available from: https://www.pwc.com/gx/en/government-public-sector-research/assets/skin-in-the-game-pbr-on-the-gec-final.pdf. [Accessed 20 august 2020].
Lloyd, C. (2011) *Evidence Paper for the Girls' Education Challenge Fund.* Unpublished Consultancy Report for DFID
Miske, S.J., & Joglekar, A.B. (2017) Using or misusing the early grade reading assessment? Examining a measure of payment by results in the Girls' Education Challenge. In Wiseman, A.W. *Annual Review of Comparative and International Education.* Bingley: Emerald Publishing Limited.
Outhred et al. (2019), *GEC-T girls' baseline learning levels and predictors of learning.* Unpublished report. OPM & Oxford MeasureEd.
Psaki, S. (2019) Delivering results in girls' education: How to evaluate what works, what doesn't and what we don't know. [Online] GIRL Center Research. Brief No. 4, Population Council. Available from: https://www.popcouncil.org/uploads/pdfs/Delivering_Results.pdf. [Accessed 27 August 2020].
Rose, P., & Yorke, L. (2019) *General versus girl-targeted Interventions: A response to Evans and Yuan*. [Online] University of Cambridge: REAL Centre. Available from: https://www.educ.cam.ac.uk/centres/real/downloads/Policy%20papers/REAL%20Girl-targeted_interventions_Response%20to%20Evans%20&%20Yuan.pdf. [Accessed 20 August 2020].
Sabates, R., Rose, P., Delprato, M., & Alcott, B. (2018) *Cost-effectiveness with equity: Raising learning for marginalised girls through Camfed's programme in Tanzania. Research and Policy Paper 18/2*. [Online] University of Cambridge: REAL Centre. Available from: https://www.educ.cam.ac.uk/centres/real/downloads/REAL%20Policy%20Brief%20Cost-effectiveness%20Camfed%20A4_FINAL.pdf. [Accessed 20 August 2020].
Sperling, G.B., & Winthrop, R. (2016) What Works in Girls' Education: Evidence for the world's best investment. Washington DC: Brookings Institution Press.
Unterhalter, E. (2015) Global inequalities, multipolarity, and supranational organisations engagements with gender and education. *Journal of Supranational Policies of Education,* 3, p.10-28.
Unterhalter, E. (2017) A review of public private partnerships around girls' education in developing countries: Flicking gender equality on and off. *Journal of International and Comparative Social Policy,* 33(2), p.181-199.
Unterhalter, E., North, A., Arnot, M., Lloyd, C., Moletsane, L., Murphy-Graham, E., Parkes, J., & Saito, M. (2014) *Girls' education and gender equality. Education Rigorous Literature Review*. [Online] Available from: https://www.edu-links.org/sites/default/files/media/file/Interventions_to_enhance_girls____education_and_gender_equality._Education_Rigorous_Literature_Review.pdf. [Accessed 20 August 2020].

Van Egmond, M.C., Omarshah, T. Berges, A.N., Benton, J,. Zalira, U., & Morrell, F. (2019) The relationship between caregivers' gender equality norms and girls' need satisfaction and self-esteem under conditions of resource scarcity. *Motivation and Emotion,* 44, p.257-269.

World Bank Development Data Group & UNESCO Institute for Statistics (2003). *A framework for assessing the quality of education statistics.* [Online] Available from: https://unstats.un.org/unsd/dnss/docs-nqaf/WB-UNESCO-DQAF%20for%20education%20statistics.pdf. [Accessed 20 August 2020].

Appendices

Appendix 1: Scoping of the global research literature that has used/addressed GEC data

In order to explore to what extent data from the GEC has been used thus far to contribute to the global evidence base, we undertook a specific key search for literature that has emerged based on GEC data. We identified four published studies that made use of GEC evaluation data. Eron and Emong (2017) made use of data from *"Cheshire Services'"* project in Uganda. They refer to qualitative data, including focus group discussions with School Management Committee members to explain the barriers to education for children with disabilities. Asadullah, Alim and Hossain (2017) evaluated baseline data from BRAC's project in Afghanistan. Their analysis shows that public schools are failing to ensure minimal learning, through demonstrating that higher grade progression lead to almost no gain in numeracy, implying that simply enrolling girls in school without improving the relationship between grade completion and learning is unlikely to transform the lives of women in Afghanistan. This evidence made a significant contribution to understanding learning outcomes in Afghanistan.

Van Egmond et al (2019) analysed data from the endline evaluation for "Theatre for a Change's" GEC project in Malawi. They found that need satisfaction matters for adolescent girls' self-esteem in this context, regardless of the level of resource scarcity they are exposed to. However, parental gender equality beliefs are negatively related to the satisfaction of the needs for relatedness and competence and show a stronger relationship with autonomy under conditions of high resource scarcity, than under low levels of scarcity. A multiple mediation analysis shows that need satisfaction partially mediates the relationship between parental gender equality beliefs and self-esteem. This study was unique in the field of a focus on basic psychological need theory in relation to marginalised girls; there is a lack of research on the influence of scarcity on the relationship between need satisfaction and self-esteem in the sub-Saharan African context, and so the use of GEC data here contributed to a gap in the evidence base. Hills (2017) also referred to improvements in education at PEAS schools, although the article did not specify which data were used to generate these statistics.

There were also two studies reviewing the GEC programme, in particular the approach to evaluation. Carr-Hill (2018) described the 'typical design for evaluation of GEC programme' (p.1011) and is critical of the evaluation approach for not taking sufficient account of the heterogeneity of schools. Even with sophisticated statistical adjustments, this paper argues that there will still be substantial residual heterogeneities, given the unobserved variables that will directly determine learning outcomes. The author suggests that a potential solution would be to compare the difference between the change in performance between baseline and follow-up scores of girls relative to the change in performance of boys' scores in the intervention group compared to the corresponding differences in the change in performance between baseline and follow-up of girls' scores relative to boys' scores in the control group. This would reduce the potential that inside classroom differences affected the outcome. However, it was acknowledged that this would be more costly and time consuming.

Miske and Joglekar (2017) looked at the evaluations that GEC has produced, particularly examining the use and adaptation of Early Grade Reading Assessments and of payment by results. They conclude that the way in which EGRA oral reading fluency was used for payment by results was a misuse of the tool, and ultimately did not serve project beneficiaries. This paper refers to GEC findings from the outset, citing the positive impact on girls' self-efficacy, transformed relationships between community stakeholders and school authorities, increased voice and decision-making by and for girls and women, and evidence of increased gender equality in schools, households and religious groups, as well as increased reading comprehension. However, they state that these dramatic changes were irrelevant

to Payment by Results, which was based solely on oral reading fluency and maths scores. This paper criticised the appropriateness and adequacy of the literacy outcome measure used to determine payment by results based on the growing evidence base challenging EGRA's conceptual and theoretical underpinnings and its expanded use (Barlett, Dowd & Jonason, 2015; Hoffman, 2012); in brief, it is criticised for isolating core components of reading and their development in stages, rather than emphasising how the various components of learning to read develop together. The oral reading fluency test has come under particular scrutiny. The FM's instructions were to use the oral reading fluency sub-test score for the Payment by Results calculation, even when many have powerfully argued that oral reading fluency cannot be used as a proxy for comprehension. The article did, however, acknowledge that the EM highlighted the flaws of the Payment by Results structure.

Two papers were specifically focused on the involvement of the private sector, or for-profit organisations in international development. Unterhalter (2015) uses the GEC as an example of how features of dispersal and multipolarity are evident in policy declarations, programme descriptions and framing discourses. She argues that there were assumptions that the state was by necessity a junior or minor player, and that step changes or innovations were to be achieved by the NGO sector and private companies. Another notion that echoed features of dispersal is that technologies and partnerships are magic bullets which could fix problems of girls out of school; a final feature of the dispersal discourse was that girls out of school could be brought in quickly, and that this problem of marginality could be detached from wider issues of poverty and inequality. The author criticises the GEC for not working to change the wider institutional context that represents barriers to girls' education, nor for working with state institutions which is still the largest provider of education in all the countries. She also notes that from the outset (baseline of GEC-1), it was not clear what additional insight or information were provided as a result of 'the huge investment in the baseline studies.'

In a later paper, Unterhalter (2017) more explicitly examined public private partnerships (PPPs) as part of the GEC. She discusses whether or not the GEC's partnership framework sets up a space for in depth engagements with gender equality and women's rights. This paper draws on policy and evaluation documents from the GEC. Unterhalter argues that the baseline evaluation was useful in developing insight into particular intervention combinations that may work to improve girls' literacy, numeracy and attendance in particular contexts, but it does not directly help to gain insight into whether the PPP form of governance is more or less efficient than one with a different structure. She also critiques the 'one size fits all' evaluation of the GEC, arguing that it misses opportunities to learn in depth about work on girls' education, gender and women's rights. She references the evaluations of GEC which note that PPP project teams have struggled to understand and respond to the complexity of marginalisation (Coffey, 2015).

It is also noted that there is some unpublished grey literature publicly available which draws on research and experience from the GEC (Holden & Patch, 2017), and some that could not be publicly accessed.[21] Such evidence could also be useful to inform a global evidence base, but has not necessarily been quality assured (as with the peer review process for published journal articles, for example), and cannot always be easily identified. Thus, we recommend using a more streamlined and holistic approach on the GEC website, to ensure that researchers can learn 'what works' from the GEC more easily.

Appendix 2: Project codes

Project DQA Code	**Project name** (GEC-1)	**Project name** (GEC-T)	**Implementing partner**	**Country**
GI_CCS	Child-Centered Schooling: Innovation for the Improvement of Learning Outcomes for Marginalised Girls in Zambia	N/A	CAMFED	Zambia
GI_CLC	Creative Learning Centres in Kampala	Building Girls to Live, Learn, Laugh and 'SCHIP' in Strong, Creative, Holistic, Inclusive, Protective, Quality Education	Viva	Uganda
GI_DPG	Discovery Project (Ghana)	Discovery Project (Ghana)	Discovery Communications	Ghana
GI_DPK	Discovery Project (Kenya)	Discovery Project (Kenya)	Discovery Communications	Kenya
GI_DPN	Discovery Project (Nigeria)	Discovery Project (Nigeria)	Discovery Communications	Nigeria
GI_EDF	Innovating in Uganda to Support Educational Continuation by Marginalised Girls (EduFinance)	Girls' Education Finance: Empowerment for Girls' Education	Opportunity International	Uganda
GI_EEP	Keeping Marginalised Girls in School by Economically Empowering their Parents	N/A	Eco Fuel	Uganda
GI_EGE	Educate Girls, End Poverty	Educate Girls, End Poverty	Relief International	Somaliland
GI_ENG	Educating Nigerian Girls in New Enterprises (ENGINE)	Educating Nigerian Girls in New Enterprises (ENGINE) II	Coca Cola/Mercy Corps at GEC-T	Nigeria
GI_FAR	Increasing the Access and Quality of Basic Education for Marginalised Girls in Faryab (ACTED)	N/A	ACTED	Afghanistan
GI_GDK	Supporting Slum and Homeless Street Girls with Disabilities in Kampala City to Access Quality Primary Education	Empowering Girls with Disabilities in Uganda through Education	Cheshire Services	Uganda
GI_GEA	Girls' Enrolment, Attendance, Retention and Results (GEARR)	GEARR-ing Up for Success After School	Promoting Equality in African Schools (PEAS)	Uganda
GI_GQE	Securing Access and Retention into Good Quality Transformative Education	Excelling Against the Odds	ChildHope	Ethiopia

GI_IGA	Improving Girls' Access through Transforming Education (IGATE)	Improving Girls' Access through Transforming Education (IGATE)	World Vision	Zimbabwe
GI_IMP	The iMlango Project	The iMlango Project	Avanti Communications Ltd	Kenya
GI_KEE	Kenya Equity in Education Project (KEEP)	Kenya Equity in Education Project	World University Service of Canada (WUSC)	Kenya
GI_MBE	Mobile Broadband and Education	N/A	Ericsson	Burma
GI_MGA	Community Based Education for Marginalised Girls in Afghanistan	Community Based Education for Marginalised Girls in Afghanistan	BRAC	Afghanistan
GI_MGC	MGCubed (Making Ghana Girls Great!)	Making Ghanaian Girls Great!	Varkey Foundation	Ghana
GI_MGS	Supporting Marginalised Girls in Sierra Leone to Complete Basic Education with Improved Learning Outcomes	Girls' Access to Education (GATE)	Plan International	Sierra Leone
GI_NEQ	A New 'Equilibrium' for Girls	Girls' Learn, Succeed and Lead	CAMFED	Tanzania and Zimbabwe
GI_NPA	Equal Access to Education for Nomadic Populations in Northern Afghanistan Project	N/A	ChildFund	Afghanistan
GI_PAG	Promoting Advancement of Girls' Education in Mozambique (PAGE-M)	Successful Transition and Advancement of Rights for Girls (STAR-G)	Save the Children	Mozambique
GI_PAS	Pastoralist Afar Girls' Education Support Projects (PAGES)	N/A	Save the Children	Ethiopia
GI_PIE	Empowering Pioneering Inclusive Education Strategies for Disabled Girls	Expanding Inclusive Education Strategies for Girls with Disabilities in Kenya	Leonard Cheshire Disability	Kenya
GI_REP	Rwandan Girls' Education and Advancement Programme (REAP)	Rwandan Girls' Education and Advancement Programme (REAP)	Health Poverty Action	Rwanda
GI_RRL	A Community Based Approach: Supporting Retention, Re-entry and Improving Learning	A Community Based Approach: Supporting	BRAC/CAMFED at GEC-T	Tanzania

		Retention, Re-entry and Improving Learning		
GI_RWZ	Life Skills and Literacy for Improved Girls Learning in Rural Wolaita Zone	Supporting Transition of Adolescent Girls through Enhanced Systems (STAGES)	Link Community Development	Ethiopia
GI_SFS	Sisters for Sisters' Education	Sisters for Sisters' Education	Voluntary Service Overseas (VSO)	Nepal
GI_SOM	Girls' Education Promotion Project (SOMGEP)	Somali Girls' Education Project - Transition	CARE	Somalia
GI_STA	Step Towards Afghan Girls' Education Stages (STAGES)	Step Towards Afghan Girls' Education Stages (STAGES)	Aga Khan Foundation	Afghanistan
GI_STE	Supporting the Education of Marginalised Girls in Kailali District (STEM)	Supporting the Education of Marginalised Girls in Kailali District (STEM) II	Mercy Corps	Nepal
GI_TIP	Tiphunzire! (Let's Learn!)	N/A	Theatre for a Change	Malawi
GI_VAS	Valorisation de la Scholarisation de la Fille (VAS-Y)	Réussite et Épanouissement via l'Apprentissage et L'Insertion au Système Éducatif (REALISE)	International Rescue Committee (IRC)/Save the Children at GEC-T	Democratic Republic of Congo
GI_VKG	Improved School Attendance and Learning for Vulnerable Kenyan Girls through an Integrated Intervention	Jielimishe (Educate Yourself)	I Choose Life	Kenya
GI_WUP	What's Up Girls?! Project	N/A	Red Een Kind	Uganda
GI_WWW	Wasichana Wote Wasome (WWW - Let All Girls Read)	Let our Girls Succeed (Wasichana Wetu Wafaulu	Education Development Trust (formerly CfBT)	Kenya

Appendix 3: Projects recommended for future quantitative analysis

This section provides information on the five GEC-1 and five GEC-T projects that have data that provide examples of ones that could form the basis of further research and analysis. These projects in Table 8 and 9 are considered promising according to the indicative findings of the DQA. It should be reiterated that no GEC-1 or GEC-T project provides complete information with regards to all DQA criteria or themes. As such, future researchers are encouraged to consider the DQA overview document before undertaking initial investigation into any GEC dataset(s) (Please contact authors for more information).

Five examples of GEC-1 projects for potential future analysis

Project DQA Code	Project name	Implementing partner	Country	Features beneficial to future researchers	Features creating potential challenges for future researchers
GI_NEQ	A New 'Equilibrium' for Girls (CAMFED)	CAMFED	Tanzania and Zimbabwe	• The selected final datasets provide information of direct relevance to 13 (of 37 potential) theme indicators • Positive information on 8 of the 15 DQA criteria were identified • The data is readily available to future researchers, as this was submitted by CAMFED to the UK Data Service	• Information on some DQA indicators was not identified
GI_NPA	Equal Access to Education for Nomadic Populations in Northern Afghanistan Project	ChildFund	Afghanistan	• The final data gave information of direct relevance to 11 theme indicators • Useful information on 6 of the 15 DQA criteria were identified	• There are challenges in telling whether a number of variables in this dataset refer to midline or endline. (This point is supported by the FM's Quantitative Review Template.) • There are some variables which contain only blank values (e.g., 'q27na0' which is labelled as 'ASER Score in Numeracy')
GI_DPN	Discovery Project (Nigeria)	Discovery Communications	Nigeria	• Final datasets provide information of direct relevance to 14 theme indicators • 8 DQA criteria were found to be positive • Datasets are largely mergeable and are in good order (with full codebooks) • Key datasets permit longitudinal investigation	• The format of the data conflicts slightly with the tools found within the project files • The data provides no information on boys' attainment • Limited information on quality control is available in the endline report

				• Equivalent data is available for the Discovery Project Ghana and Kenya (Nairobi) programmes	
GI_PIE	Empowering Pioneering Inclusive Education Strategies for Disabled Girls in Kenya	Leonard Cheshire Disability	Kenya	• Useful information on 10 DQA criteria were identified • Information of direct relevance to 13 theme indicators was found • Data is available at two time points in the single dataset considered to be final, which permits some longitudinal research	• The data provides no information on boys' learning • While variables are generally well labelled, more detailed labelling (or an overview codebook) could facilitate researcher understanding of variable prefixes that indicate, for example, the type of respondent (including caregivers)
GI_STA	Step Towards Afghan Girls' Education Stages (STAGES)	Aga Khan	Afghanistan	• Useful information on 10 DQA criteria were found in the baseline report, including reassuring information on data quality control • Information on 14 theme indicators was identified • The data gives baseline, midline and endline information	• The single final dataset selected has a very large number of variables (2004), which could potentially discourage future researchers • While labelling is generally adequate, investigation into this dataset might be enhanced by a codebook • The format of some variables appears to be unclear. For example, there are 2 (essentially non-labelled) 'survey number' variables, which hold ID information. One of these variables has 682 unique values, while the other has 1817. It is not readily apparent how these variables function.

Five examples of GEC-T projects for potential further analysis

Project DQA Code	Project name	Implementing partner	Country	Features beneficial to future researchers	Features creating potential challenges for future researchers
GI_ENG	Educating Nigerian Girls in New Enterprises (ENGINE)	Mercy Corps	Nigeria	• The selected final datasets provide information of direct relevance to 18 (of 37 potential) theme indicators • Useful information on 10 of the 15 DQA criteria were identified • Datasets are well labelled and appear clean upon initial inspection	• There are a large number of final datasets (18), which could deter potential future researchers • Information is lacking on some theme indicators, with key omissions including the lack of geographical data, non-cognitive skills information and data on boys (in learning-focused datasets)
GI_EGE	Educate Girls, End Poverty (Relief International)	Relief International	Somalia	• Information of direct relevance to 26 (of 37) theme indicators was found • Useful information on 12 of the 15 DQA criteria were identified • Datasets appear well labelled (with key variables also explained in Annex 8 of the baseline report)	• A relatively large number of final datasets (8) were identified, which might slow secondary analysis
GI_VAS	Valorisation de la Scholarisation de la Fille (VAS-Y)	Save the Children	DRC	• Datasets provide information of direct relevance to 19 theme indicators • Useful information on 10 of the 15 DQA criteria were identified • All three final datasets can be merged (to some degree)	• Boys are not included in the datasets reviewed
GI_MGA	Community Based Education for Marginalised Girls in Afghanistan (BRAC)	BRAC	Afghanistan	• Information of direct relevance to 19 theme indicators was found • Useful information on 11 DQA criteria were identified	• A relatively large number of dataset files are available for this project, meaning caution is needed in which dataset is chosen for analysis
GI_IGA	Improving Girls' Access through Transforming	World Vision	Zimbabwe	• Information of direct relevance to 19 theme indicators was identified • Useful information on 12 DQA criteria were found in the baseline report, including	• A relatively large number of final datasets (9) were identified, which could potentially deter future researchers from engaging with this project

	Education (IGATE)			reassuring information on data quality control	• Researchers would have to read dataset names carefully, as these provide information on those that are/are not final • The transition-focused dataset (amongst the selected final datasets) requires additional labelling (or a codebook) to facilitate meaningful interpretation in future analysis

Appendix 4: Projects recommended for future qualitative analysis

Four examples of GEC-1 projects for potential further analysis

Project DQA Code	Project name	Implementing partner	Country	Features beneficial to future researchers	Features creating potential challenges for future researchers
GI_IGA	Improving Girls' Access through Transforming Education (IGATE)	World Vision	Zimbabwe	• Key informant interview and FGD transcripts are provided in full. There is good detail about follow up questions and in FGDs, there is clear delineation between participant responses • Overall, the transcripts appear to enable an examination of the pathways through which the project supported changing outcomes	• There appear to be some gaps in certain transcripts, where not all questions are answered • The researcher would have to engage with the external evaluator, as there are some slight inconsistencies noted in the review of the transcripts in the way in which original and follow up questions are denoted
GI_MGA	Community Based Education for Marginalised Girls in Afghanistan	BRAC	Afghanistan	• In depth interviews and FGD transcripts have been well transcribed. In particular, the FGD transcripts make it very clear who is speaking and what their distinct characteristics are • Case studies provide an in-depth consideration of certain aspects of the project	• There are sometimes confusing parts to the transcripts, which would require engagement with the external evaluator, e.g. the in-depth interview with a government teacher appears to include multiple participants, so may be wrongly labelled • There may be some challenges in thoroughly examining the pathways through which the project supported changing outcomes, due to the scope of the evaluation
GI_NPA	Equal Access to Education for Nomadic Populations in Northern Afghanistan Project	ChildFund	Afghanistan	• Key informant interview transcripts are clear, and answers are extensive- they are clearly structured around key themes, and probes are used well to gain insights into the impact of the project • FGD transcripts are really well transcribed, making it clear who is speaking and what their distinct characteristics are • Overall, the transcripts should enable a thorough examination of the pathways	• There are some details missing from final transcripts, e.g. there are 'potential activities' detailed in the protocol, but it is not always clear whether these were undertaken - liaison with the external evaluator would therefore be necessary

				through which the project supported changing outcomes	
GI_GQE	Securing Access and Retention into Good Quality Transformative Education	ChildHope	Ethiopia	• KIIs are fully transcribed and clear. • FGD transcripts are thorough, and reflections on group dynamics would help a future researcher to understand contextual factors of relevance	• The answers in the KIIs are sometimes quite short, so in some cases there may be insufficient detail for analysis • As the purpose of each method is not fully detailed, there may be challenges in examining the pathways through which the project supported changing outcomes

Five examples of GEC-T projects for potential further analysis

Project DQA Code	Project name	Implementation partner	Country	Features beneficial to future researchers	Features creating potential challenges for future researchers
GI_NEQ	Girls Learn, Succeed and Lead	CAMFED	Tanzania	• KII and FGD transcripts are thoroughly transcribed. There is good detail about follow up questions and in FGDs, there is clear delineation between participant responses • Overall, the transcripts provided informative details on barriers to girls' education	• There is only access to sample transcripts, so in order to do a full re-analysis, engagement with the external evaluator would be necessary. • The FGDs appear to operate more like a group interview, with some respondents that are more vocal than others; future researchers would need to consider this dynamic
GI_VAS	Reussite et Epanouissement Via L'Apprentissage et L'Insertion au Systeme Educatif (REALISE)	Save the Children/ World Vision	Democratic Republic of Congo	• The sample KII and FGD transcripts have been well transcribed • Transcripts make it clear that questions are based on the research aims, and the purpose of the interviews clearly underpins the research	• There is only access to sample transcripts, so in order to do a full re-analysis, engagement with the external evaluator would be necessary • There is an indication that some of the transcripts have been written up based on field notes, rather than recordings, and whilst answers are detailed and reflective, this could limit the ability to do a full secondary analysis
GI_SOM	Somali Girls Education Project – Transition	CARE	Somaliland	• There is really good consideration of ethics in this project • There is evidence that enumerators were using probes and showed strong interviewing skills, particularly in probing for engagement from the least vocal members of the FGDs • FGD transcripts are really well transcribed, making it clear who is speaking and what their distinct characteristics are	• There is only access to sample transcripts, so in order to do a full re-analysis, engagement with the external evaluator would be necessary • The FGDs appear to operate more like a group interview, with some respondents that are more vocal than others; future researchers would need to consider this dynamic in depth

GI_EGE	Educate Girls, End Poverty	Relief International	Somaliland	• All transcripts are clearly presented in Excel – it is clear who is speaking, and the FGD transcripts show strong moderation, with good probes for answers • Generally, answers are comprehensive and answer the questions • The KII transcripts are also very clear, with linkages to the protocol • All transcripts are included which would allow for a thorough analysis	• In some cases, there is ambiguity about who is conducting interviews, e.g. in one interview with a Ministry of Education official, it seems like the person interviewing is also part of the Ministry of Education • In some cases, e.g. the FGDs with girls, it would have been useful to have some further follow up questions, such as why some girls did not get a solar lamp and others did

Endnotes

[1] This fund was distributed between three windows: the Strategic Partnerships Window, the Innovation Window and the Step Change Window.
[2] This report was prepared before COVID-19. Some of the GEC-T endline plans are likely to change in the light of the current crisis.
[3] https://dsbb.imf.org/content/pdfs/dqrs_Genframework.pdf.
[4] We also draw on the 2003 Framework for Assessing the Quality of Education Statistics developed by the UNESCO Institute for Statistics and World Bank Development Data Group (World Band Development Data Group & UNESCO Institute for Statistics, 2003)). The influence of this 2003 framework is shown by, for example, the examination of indicators relevant to participation in education . Further, the Building Evidence in Education note on assessing the strength of evidence has been influential (Building Evidence in Education, 2015) The manner in which project data sampling is investigated corresponds closely with Building Evidence in Education's consideration of the research methods used to collect data (ibid, pp. 9-11).
[5] Identification of the most recent dataset when multiple datasets on the same measures are available is aided by the data source. The source of datasets that are likely to produce the most recent data are listed in decreasing order: The FM, The EM, independent evaluators and individual implementing partners.
[6] This guidance note is under final review by the BE2 donor group, so not yet publicly available.
[7] For another related approach for assessment of qualitative data, see
[8] Discovery Project (GEC-1 and GEC-T) are treated as separate projects in the quantitative assessment.
[9] While Data Quality Assessment was carried out on fewer GEC-T projects than GEC-1 projects, more final datasets were selected for examination from GEC-T. The selection of more final datasets from GEC-T reflected the superior organisation of project data at this GEC stage.
[10] Email: REALCentre@edu.cam.ac.uk
[11] An example of how this search process could be applied to publicly available data is available at: https://github.com/JoeMarkWatson/dataset_searching.
[12] Project codes are used above to save space. Project codes and names are provided in Appendix 2.
[13] Searched for amongst other variables indicative of supply-side information. The list of projects reflects those for which supply-side information is available.
[14] Not specifically searched for in datasets.
[15] Safety was searched for amongst other words indicative of violence. The list of projects reflects those for which violence information is available.
[16] Focusing on the first sub-theme within gender-sensitive pedagogy: teacher number and quality.
[17] Focusing on any project with direct information relevant to the first or second sub-theme of transitions: transition rates or girls beginning work.
[18] East Africa = Kenya, Uganda, Tanzania, Rwanda, Burundi, Ethiopia, South Sudan, Djibouti, Eritrea, and Somalia
[19] National context considered only.
African countries that have experienced Ebola = The Democratic Republic of the Congo, Gabon, Guinea, Ivory Coast, Liberia, Mali, Nigeria, Senegal, Sierra Leone, South Africa, South Sudan, Uganda (https://www.mercycorps.org/blog/ebola-outbreaks-africa-guide/chapter-3). African countries affected by drought in recent years: Chad, Mauritania, Burkina Faso, Senegal, Mozambique
[20] Focusing on any project for which information of direct relevance to 'geography' is available. The term 'slum' was not specifically searched for in the data, so some relevant information might have been missed. However, rural, urban, latitude, longitude and geography and location were searched for.
[21] https://www.open.edu/openlearncreate/course/view.php?id=2579.

UNIVERSITY OF CAMBRIDGE
Faculty of Education

REAL
Research for Equitable Access and Learning

REAL Centre
Faculty of Education
University of Cambridge
184 Hills Road, Cambridge
CB2 8PQ, UK

@REAL_Centre

REAL Centre Director: Professor Pauline Rose
Email: pmr43@cam.ac.uk
Telephone: +44 (0) 1223 767511

REAL Centre Administrator
Email: REALCentre@educ.cam.ac.uk
Telephone: +44 (0) 1223 767693

www.educ.cam.ac.uk/centres/real

All details correct at the time of going to print, August 2020.