

## **GEC-T MEL Guidance Part 2**

**May 2017**

# Contents

<b>Part 1: Introduction to the document .....</b>	<b>1</b>
1. Introduction .....	1
<b>Part 2: The GEC project-level MEL .....</b>	<b>2</b>
2. The Monitoring Process .....	3
3. The Evaluation Process .....	5
3.1 Defining the objectives and intended outcomes .....	6
3.2 Defining the audience for the evaluation .....	6
3.3 Identifying the evaluation objectives and research questions .....	7
3.4 Selecting the evaluation approach .....	10
3.5 Identifying the data requirements .....	13
3.6 Establishing governance arrangements .....	14
3.7 Commissioning and conducting the evaluation .....	15
3.8 Using and disseminating the evaluation findings .....	16
4. The Learning Process .....	16
4.1 Communities of practice – learning clusters .....	17
4.2 Timeframe for learning activities .....	18
4.3 Dissemination and influencing .....	19
<b>Part 3: Outcomes and intermediate outcomes in project-level evaluations .....</b>	<b>20</b>
5. Outcome 1 – Learning .....	20
5.1 Measuring learning .....	20
5.2 Literacy assessments .....	21
5.3 Numeracy assessments .....	22
5.4 Third learning outcome .....	23
5.5 Designing learning outcome assessment tools .....	23
5.6 Evaluating learning outcome achievements .....	32
5.7 Communicating and presenting learning result findings .....	39
6. Outcome 2 – Transition .....	41
6.1 What is transition? .....	41
6.2 Ambition for transition on GEC-T .....	43
6.3 How transition should be measured on GEC-T .....	43
6.4 Establishing the transition outcome target .....	45
6.5 Indicators for transition .....	45
7. Outcome 3 – Sustainability .....	46
7.1 Planning for sustainability .....	46
7.2 Defining sustainability indicators and applying the Sustainability Scorecard .....	48
7.3 Financial sustainability .....	50
8. Intermediate Outcomes .....	53
8.1 Intermediate Outcome 1: Attendance .....	53
8.2 Other intermediate outcomes .....	56
8.3 Measurement of Intermediate Outcomes .....	57
<b>Part 4: Methodology for outcome and intermediate outcome evaluation .....</b>	<b>59</b>
9. Identifying target beneficiary groups .....	60
9.1 Defining target groups .....	60
9.2 Geographical mapping of intervention schools and communities .....	62
9.3 Mapping for transition .....	63

10. Evaluation approach to additionality of outcomes.....	65
10.1 The Difference-in-Difference approach .....	66
10.2 Valid approaches to apply Difference-in-Difference.....	67
10.3 Expected challenges to selecting comparison schools and communities .....	68
11. Quantitative and Qualitative methods.....	69
11.1 Why mix methods? .....	69
11.2 Quantitative methods .....	70
11.3 Qualitative methods .....	71
12. Evaluation ethics and child protection .....	79
12.1 Child protection policy requirements for external evaluators .....	79
12.2 Ethical standards in the evaluation process .....	80
13. Sampling for outcomes in GEC-T .....	86
13.1 GEC-T sampling principles and responsibilities .....	86
13.2 Representative sampling.....	88
13.3 Sampling error, compliance, and contamination.....	91
13.4 Sampling points .....	92
13.5 Suggesting a comparison group.....	93
13.6 Power calculations and sample size determination .....	94
13.7 Methods to calculate sample sizes.....	96
14. Sampling for Learning .....	97
14.1 Baseline learning treatment sample selection .....	98
14.2 Baseline benchmarking for learning .....	99
14.3 Baseline learning comparison sample selection.....	99
14.4 Sampling and cohort tracking for learning beyond baseline .....	100
15. Sampling for Transition .....	101
15.1 Baseline transition treatment sample selection .....	102
15.2 Baseline transition comparison sample selection .....	103
15.3 Baseline benchmarking for transition .....	104
15.4 Tracking for transition beyond baseline.....	104
16. Principles for linking learning and transition samples .....	105
17. Sampling for Intermediate Outcomes .....	108
17.1 Sampling points for each intermediate outcome.....	108
17.2 Baseline sampling for intermediate outcomes.....	109
17.3 Sample sizes for intermediate outcome samples .....	109
17.4 Sampling for intermediate outcomes beyond baseline .....	110
18. Summary of Sampling Approaches .....	110
<b>Appendix A: Useful resources .....</b>	<b>111</b>
<b>Appendix B: MEL Framework template.....</b>	<b>112</b>
<b>Appendix C: Terms of reference for external evaluator.....</b>	<b>120</b>
<b>Appendix D: Sampling framework template .....</b>	<b>130</b>
<b>Appendix E: Payment by results (PbR).....</b>	<b>131</b>
<b>Appendix F: Gender Analysis Guidance and Framework.....</b>	<b>135</b>
<b>Gender Equality and Social Inclusion: The GEC Minimum Standards .....</b>	<b>135</b>

## List of acronyms and abbreviations

Learning Category	EGRA/SeGRA Equivalent
GEC AGA	Girls Education Challenge
GEC-T	Girls Education Challenge
GESI	Gender equality and social inclusion
HH-based	Household based
HHS	Household survey
ICC	Intra-cluster correlation
IO	Intermediate outcomes
KII	Key informant interview
M&E	Monitoring & evaluation
MDE	Minimum detectable effect
MEL	Monitoring, evaluation and learning
PbR	Payment by result
PIRLS	Progress in International Reading Literacy Study
PISA	Programme for international student assessment
QGIS	Quantum geographic information system
RCT	Randomised control trial
SACMEQ	The Southern and Eastern Africa consortium for monitoring educational quality
SD	Standard deviation
SDG	Sustainable development goals
SEGMA	Secondary grade mathematic assessment
SeGRA	Secondary grade reading assessment
SRHR	Sexual and reproductive health rights
TIMSS	Trends in International Mathematics and Science Study
ToC	Theory of change
ToR	Terms of Reference
TVET	Technical vocational education and training
VfM	Value for money
WPM	Word per minute

# Part 1: Introduction to the document

## 1. Introduction

Following the successful application of your project into the GEC Transition Window (GEC-T), we are launching the guidance you will need in order to engage in Monitoring, Evaluation and Learning as part of the programme. This is the second part of the guidance, and should be read in conjunction with the first part which was previously issued. The enclosed second part of the guidance will provide you with everything you need for the deliverables relating to Evaluation – specifically the **MEL Framework**, for which a first draft is due on **17<sup>th</sup> July 2017** alongside an updated **Logframe** including high-level outcomes. Final versions of these documents are required by **31<sup>st</sup> August 2017**.

In developing this guidance, the Fund Manager (FM) has tried to be as comprehensive as possible to ensure that the requirements are clear. In doing so however, the document has naturally become quite large.

Please note that the information in these packs is designed to be guidance only and should be read in conjunction with the GEC-T Handbook you have already received as part of your Accountable Grant Arrangement.

To assist the reader, we have structured the document into three parts. Depending on an individual's role in the project, they may not necessarily need to read all parts. Each subsequent part increases in depth and complexity of the concepts covered. We outline these below in order to signpost which sections will be most relevant to you as a reader.

### Part 2 – The GEC Project Level MEL

This part covers the overall Monitoring, Evaluation and Learning (MEL) framework on GEC-T – with most emphasis in this particular document on the evaluation component. This part covers the principles at a high level, and therefore should be required reading for most project staff. The project manager and contract signatory should have a good understanding of the concepts covered in this part as a bare minimum.

It is recommended that all project staff are familiar with the contents of Part 2 to promote common understanding of the GEC-T evaluation requirements.

### Part 3 – Outcomes and Intermediate Outcomes

This part goes into more specific detail on the measurement of outcomes and intermediate outcomes, and would be required reading for anyone responsible for project-level monitoring or indicator development. It is highly recommended that the project manager is familiar with the concepts covered in this part. It would be required reading for staff with an evaluation function on the project team.

### Part 4 – Methodology for Outcome and Intermediate Outcome Evaluation

This part is the most technical part of the document, and covers the specific methodological requirements relating to the evaluation design and implementation. It is crucial that External Evaluators are well versed in the requirements covered in this section, and it is highly recommended that project M&E officers also have a good understanding of the concepts covered in this section. You should have an internal team member with the skills and expertise to understand the content of Part 4, and therefore be well positioned to manage the External Evaluators who will have to deliver against this guidance.

The points listed above are a suggested minimum. We would recommend that all staff read and understand as much of this guidance as possible in order to promote a common level of understanding of the Evaluation function on GEC.

Should you have any questions about the information in this pack, or the deliverables, you should contact your Portfolio Manager in the first instance, who can advise you or refer you to other Fund Manager team members.

## **Part 2: The GEC project-level MEL**

## 2. The Monitoring Process

Monitoring of GEC-T projects is undertaken by implementing organisations and by the Fund Manager, on behalf of DFID. Both levels of monitoring are important in keeping track of progress and in order to learn about which aspects of project implementation and management are working well and which aspects would benefit from adaptation or adjustment. Set out below are the requirements for projects' own monitoring and reporting, the plans and processes for which should be clearly set out in project MEL framework. In the GEC-T handbook, you will find details of the FM monitoring which will take place in parallel to projects' own monitoring. The approach includes regular monitoring (such as review of documents, attendance and observation of selected project activities, interviews with project staff and beneficiaries) and the facilitation of six monthly review and adaptation meetings with projects, to support adaptive project management.

As part of the GEC-T MEL framework it is a requirement that projects set out the detail of their approach to monitoring. As communicated in the Handbook and in first batch of the MEL guidance (issued in early May 2017), projects are responsible and accountable for regularly monitoring activities and overall project progress and reporting on progress quarterly and annually. Programmatic monitoring should include aspects related to the pace of implementation as well as the **quality of implementation**. Financial monitoring includes the monitoring of downstream partners including the appropriateness of spend and the controls around releasing payments down the supply chain. This is in addition to the information provided at application stage about your own financial processes and procedures as lead partner. In GEC-T the FM will be working with projects through its own monitoring approach to facilitate adaptive project management and a focus on what is working well in order for projects to achieve the best outcomes possible.

**Monitoring in GEC-T should be gender-sensitive**, should cover the Activity and Output levels of your logframe and should relate closely to the project Workplan. The Intermediate Outcomes and Outcomes levels of the logframe will be assessed through the independent evaluations which each project will commission.

In the following table are the key elements you should include in the Monitoring section of your MEL framework, describing your approach and how you will implement it. Please use this as a guide for drafting your monitoring strategy as part of your MEL framework.

The Monitoring section of your MEL framework will be reviewed by the relevant in-country FM team and discussed with projects at their first Review and Adaptation meeting, to ensure the FM fully understands the project's monitoring approach and what can be expected in the quarterly and annual reporting resulting from projects' own monitoring. Monitoring is an integral part of project delivery, reflected in the quarterly workplan tracker where projects will report on the monitoring of activities as well as their delivery.

**Table 1: Defining your monitoring strategy**

Steps involved	Questions to consider
1. For each activity and output, decide WHAT needs to be monitored.	<ul style="list-style-type: none"> <li>• What are the aspects of each activity which are important and appropriate to monitor in order to know about project progress and the quality of implementation?</li> <li>• What information do you expect to gather from the monitoring?</li> <li>• What is the sampling measure that you will apply to know about progress of each activity?</li> </ul>
2. For each activity and output, define the appropriate FREQUENCY of monitoring	<ul style="list-style-type: none"> <li>• How often do you need to monitor this activity?</li> <li>• Do different aspects of the activity need to be monitored more/less often?</li> </ul>
3. For each activity and output, describe HOW they will be monitored	<ul style="list-style-type: none"> <li>• Will the monitoring take place in person, or remotely (if so how, e.g. by mail, by phone?)</li> <li>• What tools do you already have which you will use to monitor?</li> <li>• What new tools might be needed?</li> <li>• Are tools gender-sensitive, disability-friendly and, where relevant, child-friendly?</li> </ul>
4. For each activity and output, decide WHO will do the monitoring	<ul style="list-style-type: none"> <li>• Which staff will be involved? (This may be different for different activities and may involve a number of different people)</li> <li>• Who will write up the monitoring findings?</li> <li>• Who will be accountable for acting on the results?</li> </ul>
5. Consider how you will monitor downstream partner expenditure and financial processes	<ul style="list-style-type: none"> <li>• What are your processes and controls in relation to the financial monitoring of your downstream partners to ensure that the contractual terms of the GEC AGA / contract (including appropriateness of spend) are met by all downstream partners?</li> <li>• What are your processes and controls in relation to releasing payment to downstream partners?</li> <li>• Who is responsible for implementing these controls?</li> <li>• Who provides oversight that these controls are in operation?</li> <li>• Confirm that the record of these controls being adhered to will be made available to the Country Finance Monitor at each visit.</li> </ul>
6. Identify any capacity gaps with regards to monitoring needs	<ul style="list-style-type: none"> <li>• Do you have the right staff with the right capacity, for example is there GESI expertise in the team?</li> <li>• Are the responsibilities for monitoring clearly set out in relevant job descriptions?</li> <li>• Is any training needed in order for project staff to be able to monitor confidently and effectively?</li> </ul>
7. Plan your internal processes and procedures for reviewing and acting on monitoring information?	<ul style="list-style-type: none"> <li>• Who will review monitoring findings in the first instance? (this may be one person or a group of people at a regular meeting)</li> </ul>



Steps involved	Questions to consider
	<ul style="list-style-type: none"> <li>What is the suite of actions that could follow? E.g. further monitoring or data collection; escalation; sharing with FM etc.</li> <li>How will learning from the monitoring process be taken forward in your project and in your organisation?</li> </ul>

### 3. The Evaluation Process

The GEC takes a rigorous approach to evaluation. Each project-level evaluation should be shaped to robustly assess the impact of the project on GEC-T outcomes. It should also be shaped to the project including the specific Theory of Change of the project, the barriers to education that are overcome, and the intermediate outcomes that express changes to barriers over time. This Chapter provides an overview of the key steps that each GEC-T project needs to follow to plan, design, conduct, and manage the evaluation so that it achieves this objective. For each key step, this chapter outlines the minimum standards and requirements to which GEC-T evaluations must adhere to. This should include and differentiate girls from a variety of sub groups, including those with disabilities, from the start of the project. This data should track girls' experiences and whether interventions are responding to their needs.

Table 2 below summarises the key steps involved in the evaluation process which are explained in more detail in the subsequent paragraphs.

**Table 2: Steps involved in designing and commissioning GEC project-level evaluation**

Steps involved	Questions to consider
1. Defining the objectives and intended outcomes	<ul style="list-style-type: none"> <li>What is the project logic or theory of change about how inputs lead to outputs, intermediate outcomes, outcomes and impacts, in the particular policy context?</li> </ul>
2. Defining the audience for the evaluation	<ul style="list-style-type: none"> <li>Who will be the main users of the findings and how will they be engaged?</li> </ul>
3. Identifying the evaluation objectives and research questions	<ul style="list-style-type: none"> <li>What does the audience need to know about what difference the project made, and/or how it was delivered?</li> <li>How broad is the scope of the evaluation?</li> </ul>
4. Selecting the evaluation approach	<ul style="list-style-type: none"> <li>What type of evaluation is required?</li> <li>How extensive is the evaluation likely to be?</li> <li>What level of robustness is required?</li> </ul>
5. Identifying the data requirements	<ul style="list-style-type: none"> <li>At what point in time should the impact be measured?</li> <li>What data is required?</li> <li>What is already being collected / available?</li> <li>What additional data needs to be collected?</li> <li>Who will be responsible for data collection and what processes need to be set up?</li> </ul>
6. Establishing governance arrangements	<ul style="list-style-type: none"> <li>What quality assurance processes will be put in place?</li> </ul>
7. Commissioning and conducting the evaluation	<ul style="list-style-type: none"> <li>Who will be responsible for specification development, tendering, project management and quality assurance?</li> </ul>

Steps involved	Questions to consider
	<ul style="list-style-type: none"> <li>• When does any primary data collection need to take place?</li> <li>• When will the evaluation start and end?</li> </ul>
8. Using and disseminating the evaluation findings	<ul style="list-style-type: none"> <li>• What will the findings be used for, and what decisions will they feed into?</li> <li>• How will the findings be shared and disseminated?</li> <li>• How will findings feed back into the MEL cycle?</li> </ul>

Source: Adapted from *The Magenta Book*

### 3.1 Defining the objectives and intended outcomes



- *What is the project logic or Theory of Change about how inputs lead to outputs, intermediate outcomes, outcomes and impacts, in the particular policy context?*

At the earliest stage of the project and evaluation lifecycles, it is important to develop a clear understanding of the rationale for intervention and the contextual conditions that potentially drive or constrain the potential effects and success of an intervention. By developing an understanding of the nature of the problem the project will be better able to understand the type of changes that are anticipated and as a result the most appropriate way of measuring those changes over time.

GEC-T projects have been responsible for the development of their Theories of Change and logframe indicators at the output and intermediate outcome level during the proposal stage and, upon contracting, during the first phase of the changeover period. It is important that during the MEL Framework development, projects are guided by their Theories of Change and logframes. As stated in the Grant Recipient Handbook, projects are required to finalise their logframes in parallel with the development of their MEL Frameworks.

### 3.2 Defining the audience for the evaluation



- *Who will be the main users of the findings and how will they be engaged?*

Defining the audience for the evaluation is important if projects want to ensure that their evaluations are targeted to subjects who can learn from and act on the findings, thereby making the greatest impact in terms of dissemination of results and lessons learning.

A tentative list of potential audiences for the evaluations is the following:

- **The recipient organisation and its implementation partners.** The first audience of an evaluation should always be the personnel of the organisation leading the project, at all its levels. This is to ensure that the results are understood across the organisation and used to learn for future programming.
- **UK Department for International Development (DFID London HQ and country offices)** is another obvious audience who, like the recipient organisation, wants to learn the results of the money being spent and to inform further programming.
- **The GEC Fund Manager** uses results from the project-level evaluations to take management decisions on the continuation of the projects and to inform DFID on the overall progress made by beneficiaries in the GEC portfolio.

- **The GEC Evaluation Manager** uses results from the project-level evaluations to re-analyse its results and produce an independent view on the impacts of the GEC portfolio, for DFID and the general public.
- **The national and local governments.** The national and local governments will be interested to learn about the progress of the project. In some cases the results are critical to build confidence in the government that aspects of the project are worth adoption on a larger scale, directly under the initiative of the government.
- **Local stakeholders, other donors, communities participating in the evaluation, academia and research institutes, general public**

Each project should ensure that their approach to identifying the evaluation audiences aligns with their Learning strategy.

### 3.3 Identifying the evaluation objectives and research questions



- *What does the audience need to know about what difference the project made, and/or how it was delivered?*
- *How broad is the scope of the evaluation?*

Defining the scope of a project evaluation amounts to asking the question: *What is going to be evaluated?* During the evaluation process stakeholders frequently express a desire to examine many potential research questions. However, in order to reach sound evidence-based conclusions, the project evaluation should be confined to an examination of project performance against a set of clearly defined research questions and judgement criteria. Projects are to define their research questions as part of the development of their MEL Frameworks.

Very broadly, there are two types of research questions: one asks about results (i.e. what impact did the programme make?), the other asks how these were achieved (i.e. how did the project achieve or not achieve its impact?). Table 3 below provides a checklist of questions that projects need to consider to define its specific research questions.

**Table 3: Questions to consider when developing research questions**

For the <i>what</i> -type questions, i.e. what impact did the project make?	For the <i>how</i> -type questions, i.e. how did the project achieve its impact?
How will you know if the programme is a success?	Is it important to understand why the programme does or does not achieve planned outcomes?
Do you need to quantify impacts, as well as describe them?	Which aspects of the delivery process are innovative?
What were the impacts for the target group?	Is it important to learn about uptake, drop-out, attitudes etc.?
Who has the project delivered for? And, were there different impacts for different groups?	What contextual factors might affect delivery?

Setting the key research questions should be one of the first steps in developing the project MEL strategy. These questions are of central importance to designing a project evaluation and research methods (both quantitative and qualitative) that are fit for purpose. Crucially, they set out what the project expects to learn from through the evaluation process, which means that they must be accurately and clearly defined. The research questions make the theoretical assumptions set in the logical framework more explicit. They provide a research framework for testing the assumptions that underpin the impact logic that has been proposed – for this reason it is important that the key research questions focus on the objectives, purpose and aim of the project to ensure that the findings from the evaluation are as relevant, meaningful and credible as possible.

In GEC-T, it is also important that the project's research questions include the programme's research questions. This is to ensure that project-level evaluations are, as far as possible, able to complement and feed into the evidence required for responding to higher-level programme and policy questions.

The programme-level research questions in GEC-T are anchored to examining the effectiveness, impact and value for money of the GEC-T projects and are focused on the GEC-T outcomes, i.e. learning, transition and sustainability. These evaluation questions are outlined below.

**GEC programme-level evaluation questions:**

1. Was the GEC successfully designed and implemented? Was the GEC good Value for Money?
2. What impact did the GEC Funding have on the transition of marginalised girls through education stages and their learning?
3. What works to facilitate transition of marginalised girls through education stages and increase their learning?
4. How sustainable were the activities funded by the GEC and was the program successful in leveraging additional interest and investment?

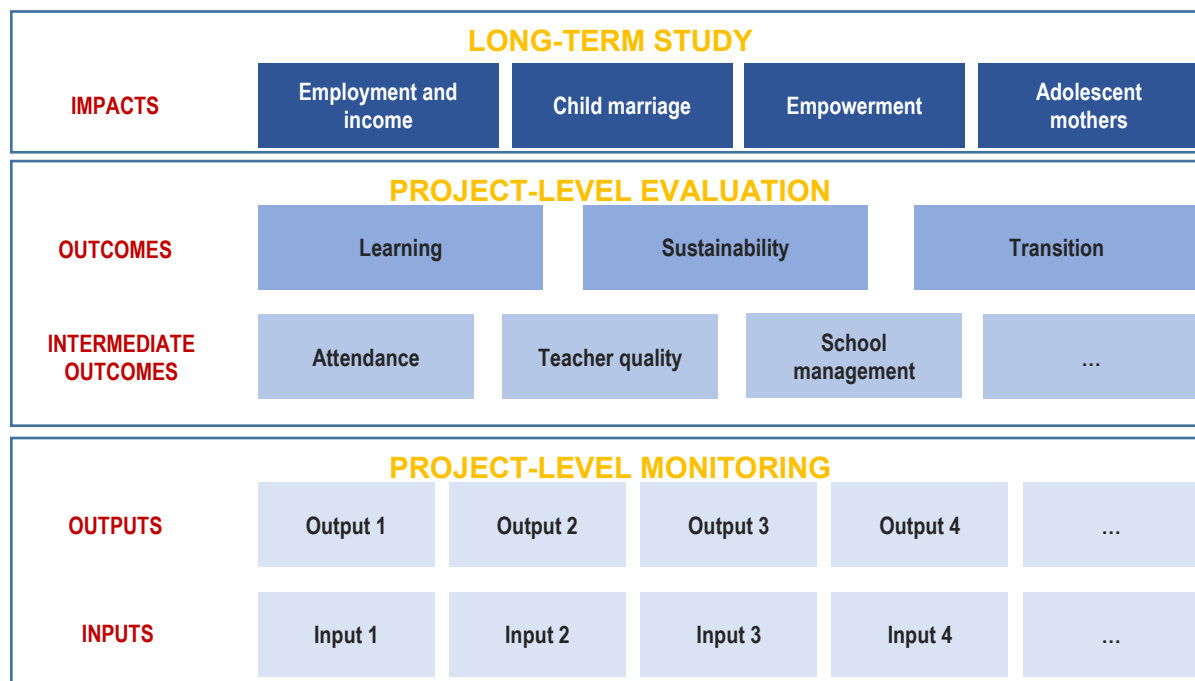
Further to the programme-level evaluation questions, each project should specifically develop complementary evaluation questions which are tailored to the interventions and to the context where they operate. In particular, each project needs to develop questions about effectiveness and impact of the project on the intermediate outcomes that they have selected. When defining these questions, the key audiences of the evaluation need to be consulted to ensure that limited MEL funds are allocated to responding the questions that are most relevant to the final users and consumers of the evaluation. The first part of GEC-T MEL guidance, issued separately in early May 2017, provides additional guidelines for the selection of Learning Themes and Clusters that will allow the GEC to leverage a large body of evidence and inform public debates on education.

Chapters 5 – 8 provide guidance to projects on how to respond to the impact questions on learning, transition, sustainability, and the intermediate outcomes.

It is important to note that, differently from the first phase of the GEC, in GEC-T, the logframe output indicators are not explicitly part of the evaluation function of projects. The evaluation process and, in particular, the work of the External Evaluator focus on impact at the outcome and intermediate outcome levels, while progress against outputs is reported and assessed through the project's monitoring function. Nonetheless, in order to trace the impact pathway from outputs to intermediate outcomes and outcomes, External Evaluators need to

comprehend what the outputs measure and what progress the project is making. Finally, any long-term impact of the project will be measured and assessed outside of the project-level evaluations and any assessments for this are not within the remit of the projects. Figure 1 below divides the logframe indicators across the respective functions and processes.

**Figure 1: Scope of the project-level evaluation**



### 3.3.1 Gender Equality and Social Inclusion

DFID is committed to ensuring that highly marginalised girls benefit from the GEC and that Recipients carefully consider the design of interventions and research questions to ensure their inclusion in education. As well as the gendered factors that may lead to educational marginalisation, Recipients should consider and address other social and economic factors that can contribute to different groups of girls being excluded from accessing learning opportunities.

Recipients must ensure:

- That their **evaluation is gender-sensitive** with regards to process, design, objectives, methodology and deliverables
- That projects' gender analysis informs the articulation of barriers and enablers to transition for each beneficiary and each potential step of their pathway. Clear activities should also be linked to each barrier and enabler
- A gender lens is applied to the analysis of GEC outcomes and intermediate outcomes

Recipients should refer to the Gender Equality and Social Inclusion (GESI) Minimum Standards that can be found in the handbook. Where applicable, the GESI minimum standards have been added to this document.

### 3.4 Selecting the evaluation approach



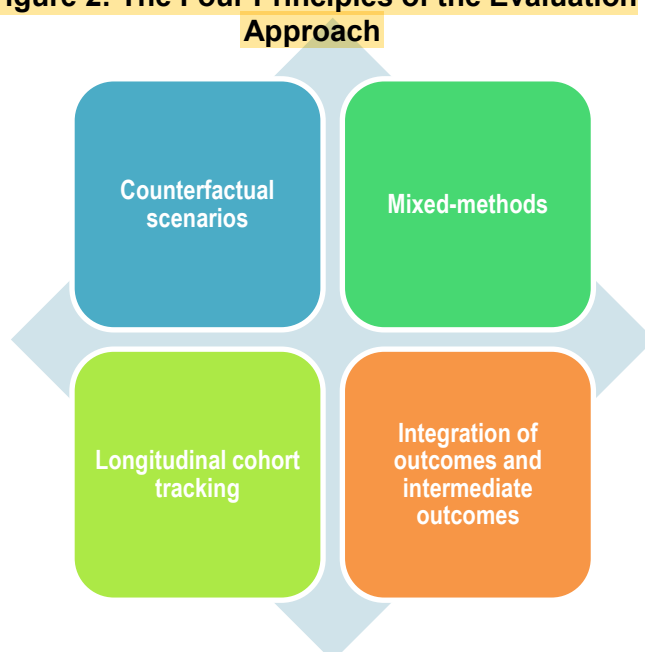
- What type of evaluation is required?
- How extensive is the evaluation likely to be?
- What level of robustness is required?

The evaluation approach is at the heart of designing and conducting the evaluation study, and is crucial for ensuring its success. It includes the choice of rigorous methods that will allow the projects to claim the impact they make on their beneficiaries. In selecting their evaluation approaches, it is recommended that the GEC-T projects follow the four key principles outlined in this section.

#### *First principle: counterfactual scenarios to demonstrate causality of impact*

The key principle in the selection of the evaluation approach is for it to be based on measuring change that can be rigorously attributed to the interventions. At the outcome level, this means that research needs to demonstrate that the outcome can be attributed to the intervention and not to other contextual factors, such as natural progression of girls through school or individuals' self-selection into the project. Measuring additional impact requires selection of a counterfactual scenario to the intervention – i.e. a state of the world that is identical to the intervention scenario except for the intervention – against which the performance of the intervention can be benchmarked.<sup>1</sup>

**Figure 2: The Four Principles of the Evaluation**



As further detailed in Chapter 10 on Evaluation approach and Sampling, in the GEC, the key approach to demonstrating causality is the 'Difference-in-Differences' approach. This approach measures the effect of the intervention as the change in the outcome observed for a group of beneficiaries before and after the intervention against the change observed for a comparison group of comparable non-beneficiaries. The change happening in the control group provides the counterfactual scenario to the interventions. The control groups can be selected through experimental or quasi-experimental designs (please see Chapter 10 for further details).<sup>2</sup>

Further to establishing a counterfactual for the package of project's interventions, projects can set up different treatment arms to isolate impact of specific activities. Beneficiaries in each arm would be exposed to different packages of interventions. When beneficiaries' selection into different arms is randomised, impacts of specific activities can be determined and

<sup>1</sup> Intermediate outcomes may also be measured against a counterfactual scenario, but this is not a mandatory requirement in the GEC.

<sup>2</sup> Pre-post experimental designs which do not involve primary data collected from a control group can be selected in extreme circumstances, and only after discussion and agreement with the Fund Manager.



compared rigorously. However this approach depends on the suitability of the project's set-up and entails greater complexity in running the project and the evaluation. Projects interested in setting up different treatment arms should work in collaboration with the External Evaluator and the FM to determine the suitability of this approach for the project.

### *Second principle: mixed-methods evaluations*

The GEC project-level evaluations will use a mix of quantitative and qualitative research methods. Quantitative and qualitative research methods may be used to research the same topic or indicator. However, they differ in terms of the type of data required for collection and analysis. In a nutshell, quantitative data can be understood as closed-ended, i.e. data that is codified and limited by the researcher in definite numerical values or definite strings of text, while qualitative data can be understood as open-ended data, i.e. data that is not limited or pre-defined.

The role of quantitative research methods is to provide a numerical measurement of the change that may be caused by the project. The advantage of these research methods is that, if conducted properly, they permit to generalise findings from a small representative sample of subjects to a larger population. The use of quantitative data also enables the researcher to define the kind and level of change expected to be observed from a certain intervention or project in advance of the data collection and analysis. Although the interpretation of quantitative data can sometimes vary, quantitative results are usually verifiable and difficult to manipulate ex-post. Hence, these methods can provide results that are at once empirically rigorous, impartial, and objective ([Rao and Woolcock, 2003](#)).

Conversely, as further detailed in Chapter 11.3, the role of qualitative research is to explore personal and social experiences, meanings, and practices as well as the role of context in shaping these. Qualitative research therefore takes as a starting point the belief that there are benefits to exploring, unpacking, and describing social meanings and perceptions of a phenomenon or programme ([Flick, 2002](#)). Not only can qualitative research give voice to people who were intended to engage with and/or benefit from an intervention, it can also help explain 'how', 'why', 'for whom', and 'under what circumstances' intended or unintended changes are taking place.

The value of mixed-methods research lies on the triangulation of findings from across distinct data sources and methods and on the capacity to build a clear and nuanced picture of what change is or is not taking place and why. Chapter 11 on the quantitative and qualitative methods further describes the benefits of this approach.

### *Third principle: longitudinal cohort tracking with individual-level outcomes*

Longitudinal evaluations involve repeated observations of the same individuals across the lifetime of a project. The minimum criteria for all GEC project evaluations is to longitudinally track a cohort of beneficiary girls for both the learning and transition outcomes. This implies that External Evaluators will need to conduct research at several evaluation points.<sup>3</sup> With outcome measurement conducted at the individual level, longitudinal evaluations allow to track progress in learning and successful transition through different stages of education. They also allow to measure the performance of the project. In order to produce generalizable findings about the population of beneficiaries, the cohorts of girls will need to be drawn by random sampling, as further detailed in Chapter 13.

---

<sup>3</sup> The number of evaluation points depends on the length of the project. Please see logframe template and logframe guidance.

To achieve a high level of rigour and establish a strong link between learning and transition outcomes, it is recommended to **employ only one cohort of girls tracked for both learning and transition**. In this approach, the girls of the cohort would be both tested for learning and assessed for their progression in school and through different transition points. The one-cohort approach is recommendable but projects must be aware that it requires a high degree of technical capacity from the External Evaluator in conducting cohort tracking, as girls who are originally sampled at the schools need to be reached at their households so that their parents and caregivers can be interviewed as part of a household survey.

Where the one-cohort approach is not deemed feasible or suitable, as a minimum criteria, each project will need **two distinct cohorts of girls** – one for the learning and one for the transition outcome. It is recommended that the sampling and tracking for learning takes place in the schools while the sampling and tracking for transition takes place at the households via a random household survey. Conducting a household survey is critical to measuring the transition outcome. This is due to difficulties in tracking girls who move from one school to another or possibly drop out at transition points at the school. As such, having two distinct cohorts is seen as a simple and reliable approach to measuring both outcomes.

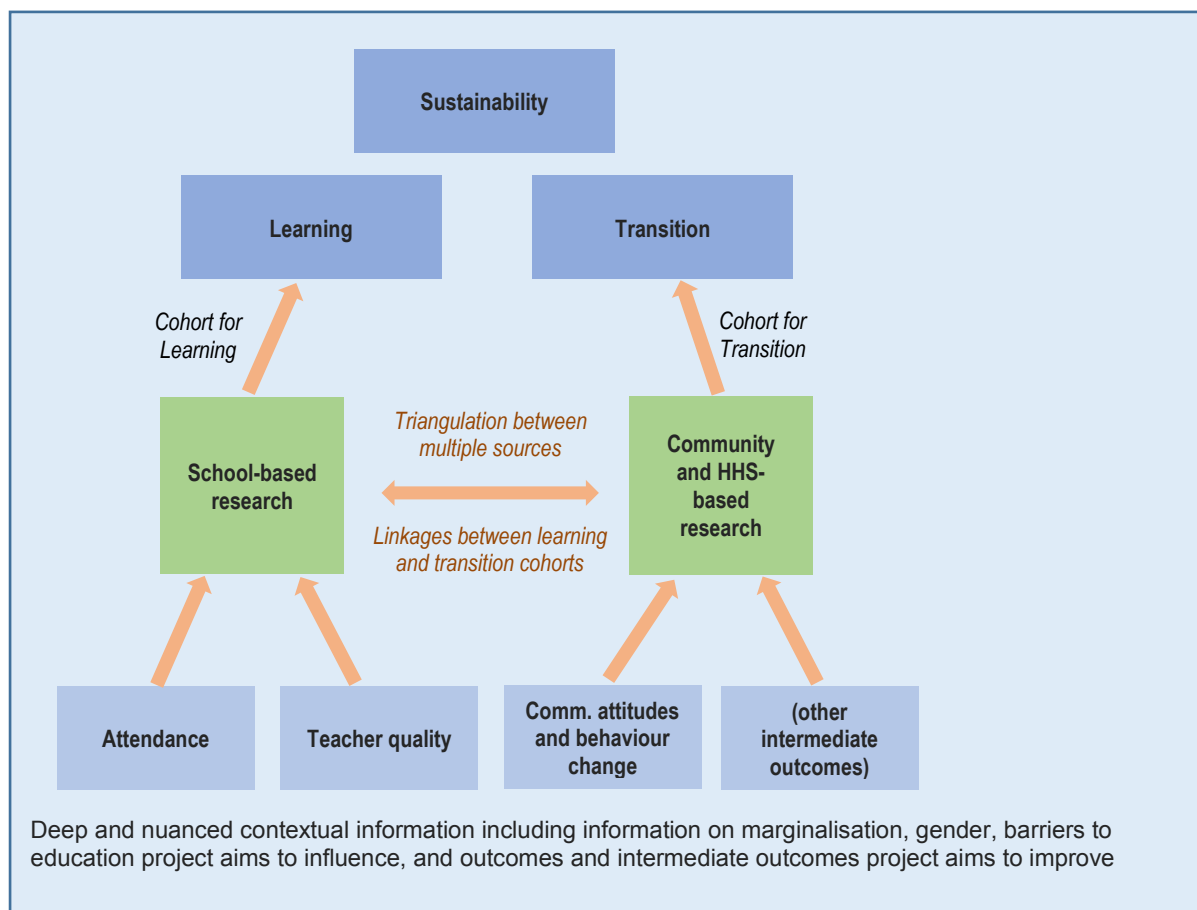
#### *Fourth principle: integration of research for outcomes and intermediate outcomes*

Projects and their External Evaluators should not only be able to demonstrate what impact the project made separately on each outcome and intermediate outcome, but also **explicitly link performance from across outcomes and intermediate outcomes**. Establishing these links is crucial for the projects to be able to demonstrate how the projects' intervention mechanisms worked to deliver across distinct outcomes as well as to understand what contextual driving and causal factors facilitated or hindered such mechanisms.

For this reason, **research on outcomes and intermediate outcomes will need to be integrated in a consistent and structured manner to enable cross-referencing of results**. A successful integration of outcomes and intermediate outcomes would be based on the combination of data sources and methodologies providing complementary information on performance as well as a wide range of contextual information. To this aim, projects and their External Evaluators need to plan for appropriate school-, community- and household-based research. The integration also implies managing data in a way that results can be cross-referenced, most commonly by using unique community and school identifiers. In other words, **when impact on one of the outcomes is observed for girls in a given community, the project and its External Evaluator would be able to link this with impact on the other outcomes and intermediate outcomes**. This implies that sampling strategies are integrated across the outcomes and intermediate outcomes. The figure 3 below illustrates how the relationship between school- and community- and household-based research delivers evidence across the outcomes and intermediate outcomes.



**Figure 3: Integrating outcomes and intermediate outcomes through school-based and community and HH-based research**



### 3.5 Identifying the data requirements



- *At what point in time should the impact be measured?*
- *What data is required?*
- *What is already being collected / available?*
- *What additional data needs to be collected?*
- *Who will be responsible for data collection and what processes need to be set up?*

Evaluations require collection and analysis of multiple kinds of data. As described in the previous section, in the GEC, both quantitative and qualitative data are necessary for providing nuanced evidence of a project's impact. To ensure that necessary evidence is captured for demonstrating this impact, it is important to identify data requirements for the evaluation in advance. Questions that should be considered in this identification process have been outlined above.

In their MEL Frameworks, the projects are required to **assess their specific data needs for the evaluation**. This includes type and availability of data (primary and/or secondary), optimal timing of data collection at baseline and subsequent evaluation points, tools and their availability for data collection and the need for developing data collection tools.

Depending of the type of data, different data sources can be used. Table 4 sets out some examples of data sources for both primary and secondary data.

**Table 4: Examples of data sources**

Primary information	Secondary information
<ul style="list-style-type: none"> <li>• Key information interviews</li> <li>• Survey of beneficiaries</li> <li>• Survey of households of the beneficiaries</li> <li>• Learning tests</li> <li>• Focus group discussions</li> <li>• Teacher observations</li> <li>• Case studies</li> </ul>	<ul style="list-style-type: none"> <li>• Monitoring data (outputs)</li> <li>• School and club records</li> <li>• Government and administrative data</li> <li>• Statistical agencies data</li> <li>• International organisations data</li> <li>• Previous evaluations and studies</li> </ul>

### 3.6 Establishing governance arrangements



- *What quality assurance processes will be put in place?*

Establishing governance arrangements that will cover the full length of the evaluations is crucial for ensuring that the evaluation reports are of the required quality and that the results feed into project and GEC-wide lessons learning. In particular, projects are expected to:

- Manage their External Evaluators;
- Monitor their operations;
- Ensure they receive and understand the GEC FM guidance; and
- Quality-assure their deliverables in terms of completeness, clarity, and adherence to best practice.

The deployment of structured review processes, performed with the required capacities, needs to be planned and put in place from the onset of the project. Importantly, projects are responsible for ensuring that the evaluations are well-resourced to deliver the required outputs. In order to achieve this, it is essential that projects are equipped with adequate internal capacity.

It is also expected that the External Evaluators will conduct professional internal quality assurance of their tools and deliverables. This would include, for example:

- Piloting of all research activities;
- Training of enumerators and researchers conducting the mixed-methods primary research, including in research ethics;
- Logistical and management planning;
- Field work protocols and data verification including back-checking and quality control by supervisors; and
- Data cleaning and editing before any analysis.

In an effort to support high-quality evaluation products, the GEC FM's role is to assist the projects and their External Evaluators through provision of guidance, structured reviews and feedback on the evaluation reports and data.

### 3.7 Commissioning and conducting the evaluation



- *Who will be responsible for specification development, tendering, project management and quality assurance?*
- *When does any primary data collection need to take place?*
- *Is a piloting or cognitive testing of research instruments required?*
- *When will the evaluation start and end?*

The commissioning and subsequent management of the evaluation is a vital aspect of the overall GEC-T programme. This requires a rigorous approach in fully understanding the evaluation objectives, parameters, key objectives and specific scope of the GEC-T. This also is predicated upon the appointment of suitably qualified External Evaluators with the necessary expertise, particularly from a mixed-methods research perspective, contextual knowledge and project management.

In a similar way to that adopted during GEC1, It is the recipient's responsibility to manage the full end-to-end process of commissioning and managing the evaluation.

Recipients should refer to and fully understand the requirements and specification detailed within both the recipient handbook and terms of reference documents, as well as the full contents of this Evaluation Guidance document.

Recipients should adhere to the following steps in the context of commissioning an External Evaluator:

- Review and sufficiently tailor the terms of reference provided in Annex C for the appointment of an External Evaluator;
- Reinforce the importance of the need for high quality, specialist evaluation, research expertise and experience – including experience in undertaking mixed-methods evaluation and reporting;
- Manage a competitive procurement process before the ultimate selection of a preferred provider. This is a vital requirement in the interests of value for money, an examination of best-fit against the key evaluation requirements and project management processes thereafter;
- Recipients may wish to re-contract the same provider to that from GEC1. However, a clear assessment is required in order to demonstrate their ability to meet the broadened parameters of GEC-T evaluations. For example, the qualitative requirements within the intermediate outcomes; and,
- Recipients should refer to the key milestones detailed within the recipient handbook in terms of the required timescales for this process.

In order to assist this process, and in the interest of deriving as much learning from GEC1 as possible, the FM will engage with recipients proactively in the final stages of commissioning. This specific process may vary from one project to another, but as minimum this will include:

- An FM review of the finalised project-level terms of reference and tender documentation;
- A review, by the relevant FM Evaluation adviser, of up to 3 shortlisted proposals once received; and
- A discussion, following the review, to share (i) reflections of proposals made based on wider programme experience (ii) advisory inputs in terms of strengths, weaknesses and risks and (iii) a recommendation – if required – concerning on a best-fit provider against the tender criteria and weighting.

It remains the recipient responsibility to contract with their preferred supplier and to then manage this contract and relationship thereafter. Recipients should also note the importance of ensuring dedicated evaluation resource – within respective project teams – to aid the management of this contract.

The role of the FM will be in these final stages of selection and will be based on advisory inputs as opposed to active decision-making.

### 3.8 Using and disseminating the evaluation findings



- *What will the findings be used for, and what decisions will they feed into?*
- *How will the findings be shared and disseminated?*
- *How will findings feed back into the MEL cycle?*

Finally, to maximise the impact of project evaluations, all projects are required to develop and adopt strategies for the use and dissemination of their findings. In the GEC, the use and dissemination of evaluation findings have a range of objectives:

- To provide evidence of project's impact for GEC (FM and DFID) decision making;
- To demonstrate accountability for the funding received;
- To inform the project management team, project partners and stakeholders in order to make improvements in the delivery of the project;
- To transfer ideas and knowledge around what works in education to implementing partners, government, donors, academia and research institutes.

In the GEC, projects will be expected to have a dissemination strategy and to regularly and timely update the GEC FM on any dissemination activity. The first batch of the MEL Guidance (issued to projects in early May 2017) provides further requirements and recommendations on this aspect.

## 4. The Learning Process

As part of the GEC MEL framework it is important to understand the 'process of learning' and why things are changing. In GEC-T we will be working with projects through a series of activities to build and share our understanding on what is working for girls' education and sharing our learning with partners within the GEC but also with broader audiences.

Below is a seven-step process for defining your learning strategy. Projects should use this as a guide for drafting your learning and influencing strategy as part of your MEL Framework.

**Table 5: Steps involved in defining your learning strategy**

Steps involved	Questions to consider
1. Define what areas of change you would like to learn more about	<ul style="list-style-type: none"> <li>• As part of your theory of change which activities do you see as critical to its success?</li> <li>• What are the principle factors that contribute to its impact?</li> <li>• What assumptions need to be tested as part of this?</li> </ul>
2. Identify where this change is happening and who is involved in making the change happen	<ul style="list-style-type: none"> <li>• Who are the main male and female agents of change? E.g. teachers, trainers, mentors?</li> <li>• How can they contribute to learning about that change through interviews, insights or data collection?</li> </ul>
3. Understand what data you are already planning to collect that could contribute to learning	<ul style="list-style-type: none"> <li>• What information are you already collecting and how could it contribute to your understanding of change?</li> <li>• How will this be integrated into ongoing monitoring and evaluation processes?</li> </ul>
4. Design any new activities you need to put in place	<ul style="list-style-type: none"> <li>• What activities do you need to put in place to understand why things are changing?</li> <li>• What new tools might be needed?</li> <li>• Are tools gender-sensitive, disability-friendly and, where relevant, child-friendly?</li> </ul>
5. Discuss who will be collecting and analysing this data and information in your team?	<ul style="list-style-type: none"> <li>• Who will be responsible for driving project learning about your intervention?</li> <li>• Who will be responsible for designing collecting data and its analysis?</li> <li>• Will those responsible have confidence to apply a GESI lens to their analysis?</li> </ul>
6. Plan what products you want to produce?	<ul style="list-style-type: none"> <li>• What products do you plan to produce over the next year and longer term and who will be involved in their delivery?</li> <li>• What are your planned dates for delivery?</li> </ul>
7. Outline targets for the dissemination of learning and how you hope to influence others and use the learning	<ul style="list-style-type: none"> <li>• Who are your target audiences (national and international) for your learning?</li> <li>• How will your project learning be shared outside the GEC learning clusters?</li> <li>• What change would you like to influence with your results?</li> <li>• How will you track and measure this?</li> </ul>

#### 4.1 Communities of practice – learning clusters

The GEC will coordinate and collate project level learning through a series of learning clusters. These will be groups of projects that will share their results and experience on specific themes and topics, to learn from each other what is working. Further to the previous guidance for learning, details of the learning clusters are now set out below. Each project can join up to 3 clusters, but you can also receive learning from other clusters. There are a small number of initial core learning clusters – these will be increased in 2018. In your MEL framework you will need to indicate which clusters you would:

- a) like to join and contribute learning, and
- b) like to hear from with learning updates.

### Core initial learning clusters:

Theme	Contribute to learning (join up to 3)	Receive learning (not limited)
School-based Learning		
Learning Cluster 1	Teaching, learning & assessment - Literacy	
Learning Cluster 2	Teaching, learning and assessment - Numeracy	
Learning Cluster 3	School governance & management	
Learning Cluster 4	Non-cognitive skills	
Community/HH Learning		
Learning Cluster 5	Social Norms around Education – community based attitudes	
Learning Cluster 6	Gender Equality & Inclusive Education (includes boys, girls with disabilities etc.)	

As part of your membership of a learning cluster, you will need to **identify a focal point and key contact for cluster activities** whilst recognising that learning activities will involve other team members and key informants. Projects will report on their learning within these clusters as part of their quarterly reporting.

The FM will coordinate the individual learning clusters and support the sharing of learning through these clusters with a series of focused activities e.g. webinars, workshops etc beginning in September 2017. DFID advisors and external academics may also participate in cluster activities.

Individual learning clusters may also identify sub areas within each learning cluster, sharing innovative tools and agreeing any joint methodologies. Clusters could also investigate the feasibility of joint dissemination and influencing strategies.

### Depth and breadth of learning



- *Are there any specific areas of learning within these themes that you would like to focus on?*
- *Would you like to propose additional themes for inclusion - Please elaborate in your MEL framework.*

### 4.2 Timeframe for learning activities

A plan of activities and learning deliverables will be required as part of your MEL Framework. This will detail what knowledge products you intend to produce with a brief description and when. Projects will report against this plan and it will also be monitored through regular GEC processes.

### **4.3 Dissemination and influencing**

Knowledge sharing both amongst GEC partners and with stakeholders is an important part of the GEC and understanding what works and why is an important deliverable of the GEC programme.

There will be opportunities to share and build learning amongst other GEC partners and also to also share results and findings with schools, local governments, policy makers and practitioners. Audiences for learning will build on those already identified in Section 2.2 Defining the audience for the evaluation above. Explaining how results were achieved and effectiveness of different approaches will add contextual and rich learning to results data.

Audiences for different data and results should be elaborated and where learning may be shared on an ongoing basis in addition to period dissemination at specific data points e.g. evaluation points.

## Part 3: Outcomes and intermediate outcomes in project-level evaluations

### 5. Outcome 1 – Learning

The GEC learning indicator is Outcome 1 for the GEC-T, and therefore should be central to the theory of change of the programme and all projects within it.

This chapter is structured with the following topics: Section 5.1 to 5.4 provides an overview to Measuring Learning on the GEC for literacy, numeracy, and the optional 3<sup>rd</sup> learning outcome; Section 5.5 provides frameworks on designing learning outcome assessment tools; Section 5.6 outlines topics relating to evaluating learning outcome achievements; Lastly, Section 5.7 provides guidance on how to communicate learning outcome results and findings to external stakeholders.

#### Learning outcome indicator: requirements

1. Learning will be assessed via the project-level evaluation, with external evaluators testing a cohort of girls on literacy, numeracy and an optional third learning outcome.
2. For projects including the optional third learning outcome, the outcome should be developed by projects. The project and external evaluator will need to develop tests for the third, optional learning outcome, which will need to be agreed with the FM alongside the MEL Framework.
3. Literacy and numeracy will be measured by tests to be developed by projects and agreed with the FM during the MEL Framework stage. These tests must follow the frameworks outlined by the FM in this guidance document and subsequent guidance provided.
4. Projects and evaluators are responsible for design and adaptations of the learning assessment tools. However the FM will support projects with links to open source references and provide examples items for sub-tasks.
5. Projects must design at least three versions of the reading and numeracy tests prior to baseline. These tests must be calibrated and piloted on a range of grades that reflects where the project's beneficiaries will start at baseline and end at endline.
6. The learning target will be quantified and proportional to existing distributions. The learning target of 0.25 standard deviations will apply to literacy, numeracy and the optional third learning outcome per year of implementation, using a difference-in-difference method.
7. Projects must agree with the FM on benchmark sampling approaches to be implemented by evaluators at baseline.
8. Projects will need to design evaluations that can measure targeted learning improvements in a way that is statistically significant, and that also provides an interpretation of the results relevant to the context. (Further detail in Section 5.7)

#### 5.1 Measuring learning

For GEC-T, the learning of marginalised girls is the central goal. The wording on this logframe outcome indicator is currently:



### ***Number of marginalised girls supported by GEC with improved learning outcomes.***

The learning outcome for GEC-T will have two core, and one optional, measurement indicators. Both **literacy** and **numeracy** will continue to be the core indicative measures of learning. They will also remain the **only basis for making bonus payment for those projects on PbR**. For this reason, **in GEC-T it is expected that projects follow as closely as possible the guidance set out in Chapter 5.**

Given that the sample of girls tested for learning needs to be representative of the broader beneficiary population, **the learning test results in terms of percentage performance against target will be used to scale the overall project beneficiary number in order to inform the logframe indicator – with literacy and numeracy weighted equally.** In this manner, **learning results for the sample can be converted into an estimate of marginalised girls with improved learning outcomes.**

There will also be an **optional third learning measure**. This should be in cases where a project has a particular curriculum focus that is central to its Theory of Change (ToC), but distinct from literacy and numeracy. Therefore the results from the third learning measure will not be included as part of the calculation used to scale the overall project beneficiary number in order to inform the logframe indicator.

The key point to note is that literacy and numeracy will continue to be a focus of learning assessment, and will need to be rigorously measured, with tests piloted to ensure they are appropriate to the project context.

## **5.2 Literacy assessments**

The ability to read is acknowledged as a core requirement for school success and international research is finding that a lack of basic reading ability contributes to drop out and an overall lack of success in schools<sup>4</sup>.

In GEC-T, literacy assessments are expected to be more standardized across the portfolio than they were in GEC1 in terms of structure and overarching framework. Nonetheless, these structures and frameworks will remain flexible to project contexts.

All projects will need to follow a similar structure to an **Early Grade Reading Assessment (EGRA)** and will need to adopt a new framework suggested in this note, henceforth referred to as the **GEC Secondary Grade Reading Assessment (GEC SeGRA)**. This full range of testing will be required in GEC-T starting at baseline. This will allow projects to track learning from early grades to higher levels over the lifetime of GEC-T. In addition, more bespoke tests may need to be developed to adjust for disability and social inclusion. **Note that other forms of testing (e.g. tests developed by government bodies) will not be allowed for the GEC-T Evaluation.** However it is encouraged that projects consider other learning assessment tools for internal monitoring and progress tracking purposes.

The GEC's project activities intend to target and reach populations of marginalised girls that may vary substantially in their abilities to read. In order to effectively measure this range, the below guidelines should be considered when designing reading assessments. These guidelines are further detailed in later sections of this note:

---

<sup>4</sup> Reading is just one subject for assessment. Others include: mathematics, science, social studies, etc.

- **Progression from Letters, to Words, to Comprehension:** The ability to read fluently and with comprehension relies upon the mastery and coordination of a number of underlying reading skills, including ability to detect and manipulate the sounds of a language (phonemic awareness); ability to automatically recognise the letters and corresponding sounds of a language's alphabet (phonics); ability to automatically decode unfamiliar words and recognise familiar words; ability to quickly retrieve the meaning of vocabulary words, whether presented orally or in written text; and the ability to make sense of connected text. Therefore, to be most effective, an assessment of reading ability should test this entire spectrum of reading-specific skills, even in girls in higher grades. Naturally, the focus particularly at higher primary and secondary levels should be on more complicated reading comprehension.
- **Timed reading, and more complex reading to accommodate fluency:** When measuring reading comprehension at the primary level, it is useful to measure fluency by timing sub-task and generating a score of correct responses within the time allowed. Measuring fluency allows the assessment to measure progress not only in terms of numbers of words that were correct but also in terms of the extent to which the student *was able to automate their response*, thereby providing more information and a greater ability to detect progress in their capacity to respond automatically over time. This timed fluency measure thus has the capacity to address potential ceiling effects in primary grades, as the words per minute count accommodates students with greater fluency. At all evaluation points, the timed fluency sub-task will be tested to allow for progress to be measured over the lifetime of the programme. Therefore, for secondary grades, while more complicated passages with slightly more challenging questions are introduced, the timed fluency sub-task will be required to be tested. This is further detailed in Section 5.5.2.
- **Design/Calibration of Multiple Tests at Baseline:** Reading assessments calibrated at the same level of difficulty should be designed at baseline for use at baseline, midline, and endline stages in order to track learning progress over time. It is important to note that while similarity in difficulty and scoring methodology are key, projects must NOT use the exact same test. Rather, items (questions) must be changed in each test to ensure that we can disentangle ability from recall. In order to enhance the quality of calibration, all projects must design and pilot at least three versions of the reading tests prior to the time of baseline. Projects that may require more versions to accommodate, for instance, language, disability, and/or duration of project should discuss these requirements with the FM proactively.

### 5.3 Numeracy assessments

In GEC-T, like the literacy assessments, numeracy assessments are also expected to be more standardized across the portfolio than they were in GEC1 in terms of structure and overarching framework. As above, these structures and frameworks will remain flexible to project contexts.

All projects will need to follow a similar structure to an **Early Grade Mathematics Assessment (EGMA)** and will need to adopt a new framework suggested in this note, henceforth referred to as the GEC **Secondary Grade Mathematics Assessment (GEC SeGMA)**. This full range of testing will be required in GEC-T starting at baseline. This will allow projects to track learning from early grades to higher levels over the lifetime of GEC-T. In addition, more bespoke tests may need to be developed to adjust for disability and social inclusion

In order to effectively measure numeracy, the below guidelines should be considered when designing assessments. These guidelines are further detailed in later sections of this note:

- **Progression from Numbers, to Addition/Subtraction and Multiplication/Division:** As with reading, the ability to effectively perform higher-order mathematics tasks relies upon the ability to fluently identify and manipulate numbers. For this reason, assessments of mathematics ability should include at a minimum measures of number identification, number manipulation, addition and subtraction and finally multiplication and division. In order to ensure that students are able to apply these operations, a small number of word problems must also form part of the standard test.
- **Linkages to curriculum to accommodate mastery:** Unlike the reading assessment which will rely on timed tasks, and more complicated reading passages to address potential ceiling effects, in mathematics this accommodation will mean that projects will approach country-specific curriculum benchmarks instead. In most cases, this will imply that projects will design additional sub-tasks for measurement and space data, algebra, and a larger number of word problems employing more sophisticated mathematical operations.
- **Design/Calibration of Multiple Tests at Baseline:** Like reading assessments, mathematics assessments calibrated at the same level of difficulty should be designed at baseline for use at baseline, midline, and endline stages in order to track learning progress over time. It is important to note that while similarity in difficulty and scoring methodology are key, projects must NOT use the exact same test. Rather, items (questions) must be changed in each test to ensure that we can disentangle ability from recall. In order to enhance the quality of calibration, all projects must design and pilot at least three versions of the mathematics tests prior to the time of baseline. Projects that may require more versions to accommodate, for instance, language, disability, and/or duration of project should discuss these requirements with the FM proactively.

## 5.4 Third learning outcome

GEC-T includes the addition of a third, *optional* domain of learning – beyond literacy and numeracy - that is context and/or project-specific. This is based on a review of GEC1 which highlighted that many projects are delivering interventions that support valuable other aspects of learning, particularly those working at secondary school and beyond.

The third learning outcome is optional and will depend on a project's objectives. Examples include financial literacy, ICT skills, science, foreign languages, and vocational training. The inclusion of life skills as a third learning outcome will need to be fully justified including a clear definition of which skills are specified. The FM encourages recipients to propose ideas/thoughts on which 'other' domains of learning are relevant to each project and context for consideration. These proposals would need to be set out clearly at MEL Framework stage and would be agreed with the FM at contracting. Subsequently, during the development of the MEL Framework for GEC-T, the FM would ultimately have to sign-off on an agreed measurement tool following a pilot.

As this area would be at the outcome level of the logframe, there will be a requirement for this outcome to be measured against a comparison group by the external evaluator.

## 5.5 Designing learning outcome assessment tools

### 5.5.1 Framework for the development of EGRA and EGMA style learning assessment tools

This section outlines the key sub-tasks that make up the EGRA and EGMA style tests. For reading, the skills tested include Letter Sound Identification, Familiar Word, Invented Word, Short Paragraph – Word per Minute (WPM) and Comprehension. The math skills tested include Number Identification, Quantity Discrimination, Missing Numbers, Addition,

Subtraction and Word Problems, including multiplication and division. Projects are expected to closely follow this framework and are responsible for designing tests. However, for further support, the FM will also provide sample tests in English and reference links for existing open source items (questions) in July 2017. Alterations to the framework presented here is possible if the context dictates so, and projects should feel free to proactively discuss any anticipated challenges with this framework with the FM.

### Developing EGRA/EGMA Style Tools

Projects will be responsible for developing assessment tools in conjunction with their evaluators. To facilitate the process, the FM will provide:

- Overarching framework for the EGRA/EGMA style test – in this note
- Detailed blueprints for EGRA/EGMA style tests. These will include types of items (questions) that should be included in each sub-task by content and cognitive domain, number of questions, timing, scoring etc. – in this note

Note that the FM will provide further guidance, support and advice on design and calibration during the test design phase. This may also include the provision of access to additional test items, sample tests, and references for locating open source test items for further adaptation. However, it will be the responsibility of the recipient and their external evaluator to finalise tests that align to the given framework and that are signed off and approved by the FM.

### EGRA style framework<sup>5</sup>

Five components are generally accepted as necessary to master the process of reading: **phonological awareness, phonics** (method of instruction that helps teach sound–symbol relationships), **vocabulary, fluency, and comprehension**<sup>6</sup>. The skills within each component are not sufficient on their own to produce successful reading, but they build on one another and work together to reach the ultimate goal of reading comprehension. The table below illustrates how the different components relate to these skills.

Early Grade Reading Assessment (EGRA)		
Component	Early reading skill	Skill demonstrated by students' ability to:
1. Listening comprehension	Listening comprehension; oral language	Respond correctly to different types of questions, including literal and inferential questions about the text the assessor reads to them
2. Letter identification: Letter names and/or letters sounds	Alphabet knowledge	Provide the name and/or sound of letters presented in both upper case and lower case in a random order
3. Nonword reading	Decoding	Make letter-sound (grapheme-phoneme correspondences, or GPCs) through the reading of simple nonsense words
4. Oral reading fluency with comprehension	Oral reading fluency	Read a text with accuracy, with little effort, and at a sufficient rate
	Reading comprehension	Respond correctly to different types of questions, including literal and inferential questions about the text they have read
5. Initial or final sound identification, or letter sound discrimination, or phoneme segmentation, identification of onset/ rime sounds	Phonological awareness	Identify/differentiate the onset/rime sounds of words or the initial or final sounds of words, or segment words into phonemes by having the assessor and then the student read the phonemes aloud
6. Familiar word reading	Word recognition	Read words which are randomly ordered and drawn from a list of frequent words

<sup>5</sup> References to EGRA throughout this section are an adaptation or reproduction of an original work published by RTI International and licensed under the Creative Commons Attribution 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

<sup>6</sup> (Armbruster, Lehr, & Osborn, 2003; Vaughn & Linan-Thompson, 2004)

## *GEC-TEGRA style standard subtasks*

### **Subtask 1: Letter sound identification (phonological awareness. i.e. mapping sounds onto letters)**

This subtask consists of a hundred total items. Letters of the alphabet, plus any digraphs and diphthongs if appropriate, are distributed randomly, with ten letters to a line in horizontal rows, and evenly distributed among upper and lowercase letters.

Most of the characters will be presented multiple times. The percentages calculated in the table to the side act as a guide for the frequency with which the letters, diphthongs, and/or digraphs appear in the task sheet.

Letter frequency table									
E	12.02%	G	2.70%	Y	2.11%				
A	8.12%	U	2.88%	W	2.09%				
R	6.02%	D	4.32%	K	0.69%				
I	7.31%	P	1.82%	V	1.11%				
O	7.66%	M	2.61%	X	0.17%				
T	9.10%	H	5.92%	Z	0.07%				
N	6.95%	Q	2.03%	J	0.10%				
S	6.28%	B	1.49%	0	0.11%				
L	3.98%	F	2.30%						

Source: English letter frequency (based on a sample of 40,000 words). Itasca, New York: Department of Mathematics, Cornell University. Retrieved September 2013 from <http://www.math.cornell.edu/~mec/2003-2004/cryptography/subs/frequencies.html>

To create calibrated versions of this sub-task for future evaluation points, it is recommended to reorder the letters within the individual rows (in order to retain relative subtask difficulty). This allows the frequency percentages to remain the same.

### **Subtask 2: Familiar word (phonics. i.e. recognition of words)**

Familiar words are high-frequency words selected from first-, second-, and third-grade reading materials and storybooks in the language and context. Word lists for this task are created from national reading textbooks from the grade levels that will be included in the study. 50 common, familiar, and simple words representing different parts of speech (e.g. nouns, verbs, adjectives, if applicable). The items within rows of the grid can be reordered (re-randomized) for calibrating an equivalent test for future evaluation points.

#### **Sub-task 2: Key rules**

- Pronunciation of the words is unambiguous and familiar in the relevant language or dialect
- Balance between decodable familiar words (e.g., “cat”) and common sight words (e.g., “the”)
- Word length and spelling patterns are representative of those found in early grade readers and words are composed of a variety of letters, with none repeated disproportionately
- No one letter words or words from other languages

### **Subtask 3: Invented word (phonics. i.e. decoding new words using knowledge of sounds)**

This portion of the assessment includes a list of 50 one- and two syllable non-words, five per row, with the patterns of letters within the words adjusted as appropriate by language. Non-words follow the rules of the language, using letters in legitimate positions (e.g., in English, not “wuj” because “j” is not used as a final letter in English). Also, they are restricted to consonant-vowel combinations that are typical of the language and are not homophones of real words (e.g., in English, not “kat,” homophone of “cat”).

The items within rows of the grid can be reordered (re-randomized) for calibrating an equivalent test for future evaluation points.



#### **Subtask 4: Short paragraph - Words per Minute (WPM) (fluency. i.e. reading text quickly, accurately, proper expression)**

To create the oral reading fluency with comprehension subtask, it is recommended to use narratives from children's reading materials. A narrative story has a beginning section where the characters are introduced, a middle section containing some dilemma, and an ending section with an action resolving the dilemma. It is not a list of loosely connected sentences. The length of the story should be around a minimum of 100 words, with longer passages (e.g. up to around 240 words) being preferable.

##### **Sub-task 4: Key rules**

- Character names frequently used in the school textbook are to be avoided, as students may give automated responses based on the stories with which they are familiar. However, character names must be typical of the language and context
- Only one to two characters, to avoid the task becoming about memory recall;
- Names and places reflect the local culture
- Story text contains some complex vocabulary (e.g., inflected forms, derivations) and sentence structures
- No pictures are included.

#### **Subtask 5: Comprehension**

The associated list of five comprehension questions to the Short Paragraph - WPM (Subtask 4) includes ones that can be answered directly from the text as well as at least one inferential question requiring students to combine knowledge and experience from outside the text to respond correctly. These inferential questions will have more than one right answer, but the answers must be logical based on the text and the context. Literal questions that are linked directly to the oral reading passage are the easiest type of comprehension measure.

When calibrating this subtask for future evaluations, it is recommended to make simple changes to the story. For example, names of story subjects, actions, and adjectives can be replaced with similar grade-level alternatives.

#### ***EGMA style framework<sup>7</sup>***

The Core EGMA measures foundational mathematical skills. Unlike literacy, mathematical skills continue to build upon each other throughout a lifetime. However, there is not necessarily linear developmental progress between components (i.e. addition is not a pre-requisite for division). Therefore, the following components chosen in EGMA are chosen due to the fact they reflect the curriculum, are predictive, and are teachable. The table below illustrates how these components relate to the skills tested.

---

<sup>7</sup> References to EGMA throughout this section are an adaptation or reproduction of an original work published by RTI International and licensed under the Creative Commons Attribution 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

## Early Grade Math Assessment (EGMA)

Component	Early mathematics skill	Skill demonstrated by the student's ability to...
1. Addition and Subtraction Level 1*	Knowledge of and confidence with basic addition and subtraction facts	Mentally solve addition and subtraction problems, with sums and differences less than 20
2. Quantity Discrimination	Number comparison	Identify the larger of a pair of numbers
3. Missing Number	Number patterns discernment and completion	Determine the missing number in a pattern of four numbers, one of which is missing
4. Addition and Subtraction Level 2*	Application of procedural addition and subtraction knowledge.	Solve addition and subtraction problems that involve the knowledge and application of the basic addition and subtraction facts assessed in the Level 1 subtask
5. Word Problems	Interpretation of a situation, planning, and problem solving.	Solve problems presented orally using any strategy that they wanted, including the use of paper and pencil and/or counters supplied by the assessor

### *GEC-T EGMA style standard subtasks*

It should be noted that the Core EGMA is also an oral assessment, so no student writing is required. Answers to all items are orally transmitted from a student to an assessor.

#### **Subtask 1: Number identification**

Number Identification subtask consists of 20 items that increase in difficulty. Students are asked to say each number aloud.

##### **Sub-task 1: Key rules**

- The first three items of the subtask include the numerals 0, 9, and one other single-digit number.
- The next 12 items consist of two-digit numbers from 10 to 99
- Last five items are three-digit numbers from 100 to 999

#### **Subtask 2: Quantity Discrimination**

The Quantity Discrimination subtask consists of 10 items. Each item consists of a set of two numbers, one of which is greater than the other. Students state the higher of each set of two numbers (pointing at the correct number is insufficient evidence for scoring).

##### **Sub-task 2: Key Rules**

- The first item is a set of one-digit numbers;
- Next five items are sets of two-digit numbers;
- Last four items are three-digit numbers.
- Include a mix of numbers that are further apart (e.g., 29 and 83) and those closer together (e.g., 32 and 29) to assess a broader range of abilities.

### Subtask 3: Missing Numbers

The Missing Number subtask consists of 10 items. The items are presented as four horizontally aligned boxes, three of which contain numbers and one of which is empty (the target missing number). Eight of the items increase in number from left to right; two of the items decrease in number from left to right. Students are asked to state the number that belongs in the empty box.

#### Sub-task 3: Key rules

- Items 1, 2, and 6 increase by one (in a set of one, two, and three-digit numbers), respectively;
- Items 3, 4, 5, and 8 increase by tens, hundreds, twos, and fives, respectively;
- Items 7 and 9 decrease by twos and tens, respectively.
- Last item with numerals within the range of 1–20 increases by fives, but does not begin with a multiple of five

### Subtasks 4 and 5: Addition and Subtraction

The Addition and Subtraction Level 1 subtasks consist of 20 items each that increase in difficulty. No addends are greater than 10, and no sums are greater than 19. The Addition and Subtraction Level 2 sub-tasks consist of five items each that increase in difficulty. No sums are greater than 70.

#### Sub-task 4 and 5: Key Rules

- The subtraction problems are the inverse of the addition problems
- Three items of Addition and Subtraction Level 1 mirror three of the Word Problems items

### Subtask 6: Word Problems, including multiplication and division

The Word Problems subtask consists of six items each that increase in difficulty. It should be noted that the word problems are presented to students orally to prevent the possible confound of literacy.

#### Sub-task 6: Key rules

Problem type	Example
Change: Result Unknown	Two children are on the bus, three more children get on. How many children are on the bus altogether?
Combine: Result Unknown	There are six children on the bus, two are boys. The rest are girls. How many girls are there on the bus?
Compare: Change Unknown	There are two children on John's bus and seven children on Mary's bus. How many children must join John's bus so that it has the same number of children as Mary's bus?
Change: Start Unknown	Five children get on the bus. Now there are 12 children on the bus. How many children were on the bus to begin with?
Sharing	Four children share twelve candies equally between themselves. How many candies does each child get?
Multiplicative	There are five seats on the bus. There are two children on each seat. How many children are on the bus altogether?



### 5.5.2 Framework for test adaptations for higher levels

As girls transition to higher grades in the next phase of GEC, it is particularly important to ensure learning assessments are fit for purpose, and can accommodate these transitions.

The following section provides preliminary guidance on when and how projects will be expected to make adjustments to their assessment system for higher levels of learning. Additional guidance with even more detailed assessment design principles will be made available during MEL Framework design phase. This should allow projects enough time to develop, calibrate, and pilot their assessment tools prior to the GEC-T baseline.

#### *Introducing GEC SeGRA and GEC SeGMA*

The GEC Secondary Grade Reading Assessment (GEC SeGRA) and the GEC Secondary Grade Mathematics Assessment (GEC SeGMA) are custom-built assessment frameworks designed for the GEC in order to accommodate for potential ceiling effects in GEC-T.

Like the EGRA/EGMA style tests, the GEC SeGRA/SeGMA will be structured around subtasks that become more difficult and test higher orders of learning. The specific subtasks are designed in line with the content and cognitive domains used in international assessment systems such as the SACMEQ, PIRLS, PISA, and TIMSS. They have also been adjusted for developing country contexts through a sample review of curricula in GEC countries.

In addition, these new subtasks follow two key guiding principles. First, the additional subtasks attempt to act as a bridge between early grade testing and grade-appropriate curriculum-referenced testing. In the first phase of GEC, projects that tried to use grade-appropriate curricula to develop assessment tools often faced serious challenges with floor effects. This is not surprising, as such projects often went from testing students on basic number identification to testing them on the curriculum prescribed calculus. The GEC SeGRA/SeGMA subtasks are instead designed so that they test reading and mathematics skills in a progressive fashion, approaching fully functional levels of literacy and numeracy in the final subtasks. For most countries, however, it should be noted that this does not imply that students will be tested at exact grade appropriate curricular standards.

Second, the GEC SeGRA/SeGMA shifts from the mixed oral tradition in early grades testing to a completely written tradition in the secondary levels. This shift is justified on the basis of a couple of assumptions. For one, students shifting to the secondary level are expected to have already mastered phonics at the primary level, and thus should be ready to handle the proposed written content. For another, the written tradition is likely to be in line with the norms of assessment at secondary level within countries, and thus it is important that projects prepare and test more advanced students in this manner. Finally, because all projects are expected to continue to administer the basic EGRA/EGMA style tests to students, GEC SeGRA/SeGMA subtasks will pick up whether students have successfully moved from one learning benchmark to another.

Please note that these forms of testing may nonetheless disadvantage certain students. Guidance on further adaptations for disability and inclusion are provided below in this guidance document.

#### *GEC SeGRA/SeGMA framework*

To capture progression of learning skills, the GEC SeGRA and the GEC SeGMA will contain a total of three subtasks each. These subtasks will roughly correspond to the three key transition points identified for GEC-T (1) primary to lower secondary, (2) lower secondary to upper secondary, and (3) upper secondary to beyond.

**ALL projects will be expected to design and implement GEC SeGRA and SeGMA subtask 1 to ALL beneficiaries.** For subtasks 2 and 3, projects will be given the ability to opt-in if they find these subtasks appropriate for some or all of their beneficiaries. Appropriateness will depend on factors such as the relevant transition points for the project, ensuring all ranges of performance will be picked up (i.e. a normal distribution of scores), providing enough difficulty to allow proper measurement of progression over time to avoid potential ceiling effects, and project aspirations for beneficiaries. All projects focusing on life skills, vocational skills, accelerated learning programmes, and other interventions outside of the standard schooling system should individually discuss with the FM whether subtasks 2 and 3 will be useful for their contexts.

The below table presents the basic framework for the GEC SeGRA and SeGMA, alongside the basic framework for the EGRA/EGMA-style tests for ease of comparison. These frameworks will be further defined through additional learning test support and guidance to be provided.

	<b>Early Grades</b>	<b>Subtask 1</b>	<b>Subtask 2</b>	<b>Subtask 3</b>
Applicable for	<b>ALL PROJECTS</b>	<b>ALL PROJECTS</b> Transition of primary to lower secondary	OPT-IN Transition of lower secondary to upper	OPT-IN Transition of upper secondary to beyond
Literacy	Letter Sound Identification Familiar Word Invented Word Short Paragraph – WPM Comprehension	Longer, more complicated comprehension paragraph, with more analytical questions	Longer, more complicated comprehension paragraph, with more inferential questions	Short essay construction
Numeracy	Number Identification Quantity Discrimination Missing Numbers Addition Subtraction Word Problems, including multiplication and division	Advanced multiplication and division, proportions (fractions, percentages), space and shape (geometry), and measurement (distance, length, area, capacity, money) presentation questions	Algebra questions	Data interpretation and sophisticated word problems, solved using complex, multiple operations including algebra

### *Developing GEC SeGRA/SeGMA Tools*

Design and implementation of the GEC SeGRA/SeGMA will follow similar principles to the ones presented for the EGRA/EGMA style tests. This means that projects will be responsible for developing these assessment tools in conjunction with their evaluators.

The overarching framework for the GEC SeGRA/SeGMA is provided in this guidance document.

Note that the FM will provide further guidance, support and advice on design and calibration during the test design phase. This may also include the provision of access to additional test items, sample tests, and references for locating open source test items for further adaptation. However, it will be the responsibility of the recipient and their external evaluator to finalise tests that align to the given framework and that are signed off and approved by the FM.

### 5.5.3 Adaptation for languages

In first instances, it is recommended that projects requiring versions of their tests in a language other than English search for versions of that test using reference links for existing open source items (questions). The FM will provide support where possible, however the responsibility rests with recipients and their external evaluator to locate or develop appropriate test items in local languages other than English.

### 5.5.4 Adaptations for disability

Careful consideration needs to be paid when administering assessments to girls and/or boys who have impairments to ensure that the process is both accessible and represents a fair reflection of their abilities. This section summarizes some of the ways in which projects can make adaptations to ensure their assessments are inclusive. At a broader level, however, projects should be cognisant of the fact that it may not just be their assessment approaches, but also their wider approaches to teaching and learning that require adjustments for inclusiveness. In instances where this might be the case, projects are encouraged to approach the FM to provide support for bespoke solutions to the overall teaching, learning, and assessment strategies proactively.

The best preparation for making adaptations for disability will always be knowing as early as possible what barriers some children could face in accessing the assessments. Ensuring students with disabilities have been identified will help the preparations but it will be important to ensure that all assessment processes can respond to the accommodation needs of those expected to take them.

#### *Cases where design adjustments are possible*

On a general level, there are some barriers which are relatively simple to remove. For example, the assessment itself needs to be carried out in an environment which is physically suitable for everyone (can those with mobility impairments access the space/are the seating arrangements suitable?); is well lit (for those with more limited vision/or who rely on visual communication); and is relatively free from distraction (either auditory or visual for those who may have more difficulty with concentration)? All of these factors will improve the experience for everyone, but it may have a more positive impact on the comfort levels of those with some forms of disability.

Sometimes the assessment itself may need to be modified so the beneficiary can take it themselves – for example, for those with low vision or less severe cognitive or hearing impairments the assessment would benefit from being reproduced in large print format using *sans serif* fonts like Arial or Verdana with paragraphs that are not justified. Others may just need a little extra time to take the assessment, especially if they have mobility or communication impairments that impact on their speech.

#### *Cases where redesign may be required*

In some cases however, modifications to the test will not be sufficient to overcome the specific barriers faced and alternative assessment processes will need to be devised. Girls who are deaf, who have more severe cognitive impairments or have specific language impairments learn to read differently to their peers and therefore administering a standardised assessment process to these girls, even with modifications, will not provide them with a fair opportunity to

demonstrate their capacity. In cases where relatively high numbers of girls with more severe impairments are to be included, work should begin early on the development of literacy and numeracy assessment processes which are a more accurate and fair representation of progress.

Please be aware that transliterating text from the written language to Braille will provide girls who use this system with the opportunity to take the assessment independently but literacy fluency rates will be slower so they will require more time to complete the assignments. This will also impact on Word Per Minute rates which will need to be calibrated against local Braille literacy fluency standards. Note, this does not imply that the expected standards of achievement should be lower – only that the assessment process itself is a more accurate representation of how the girls learn.

### *Delivering the assessment system fairly*

Finally, all enumerators need to undergo some form of disability awareness training/briefing so they are prepared to engage with girls who have a range of different impairments regardless of whether there are known to be disabled girls in the test sample.

## **5.6 Evaluating learning outcome achievements**

### **5.6.1 Piloting and calibration of tests and assessing ceiling and floor effects**

For the purposes of the GEC-T Evaluation, learning assessment tools must be calibrated to the same level of difficulty. All test versions should be designed prior to baseline for use of baseline, and all subsequent evaluation points in order to track learning progress over time. It is important to note that while similarity in difficulty and scoring methodology are key, projects must NOT use the exact same test. Rather, items within each sub-task must be changed in each test for each evaluation point to ensure that we can disentangle ability from recall. In order to enhance the quality of calibration, all projects must design and pilot at least three versions of the reading and numeracy tests prior to baseline on a range of grades that reflects where the project's beneficiaries will start at baseline and end at endline. Projects that may require more versions to accommodate for instance language, disability, and/or duration of project should discuss these requirements with the FM proactively.

### *Calibration*

Best practice for subtask modification recommends limiting the need for post-administration statistical equating. Techniques for preparing equivalent forms may include:

- Making simple changes in the names of story subjects, actions, and adjectives, replacing them with grade-level equivalents
- For subtasks that are presented to learners on stimuli sheets that are in a grid format, shuffling items within the grid rows, so that no matter how far a student gets in the assessment before the time is up, his or her experience with that assessment will be the same as with a previous test administration.

While these techniques are intended to limit the need for equating, they do not guarantee equivalent forms, nor do they remove the need to test for equivalence after piloting. For situations in which these techniques are used but still result in non-equivalent test forms, statistical equating methods may be required.

### *Piloting*

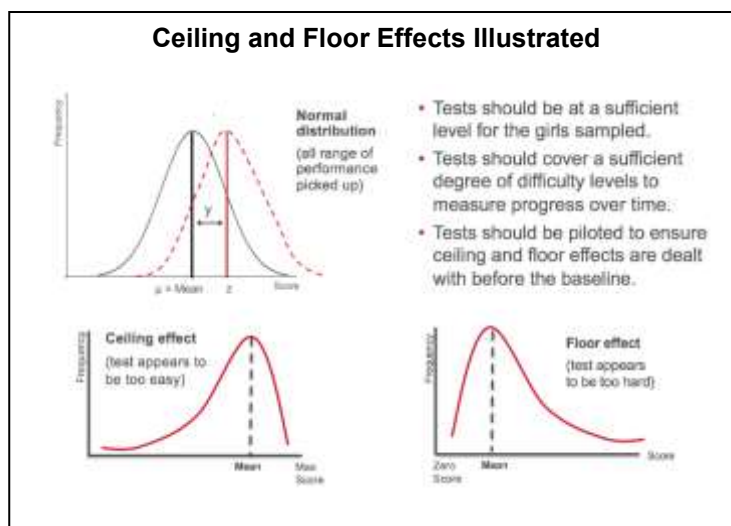
It is therefore important that learning tests are developed in good time, and this requires planning before starting fieldwork. Every project in the GEC-T will need to follow a clear

process to ensure learning tests are ready and 'signed-off' prior to the evaluations being undertaken. The key steps in this process are as follows:

1. Develop multiple versions of the test: projects will need to develop enough versions of both literacy and numeracy tests for every evaluation point. This includes EGRA subtask 4 (Short Paragraph – WPM), SeGRA subtask 1, SeGMA subtask 1 and any additional SeGRA/SeGMA subtasks agreed with the FM.
2. Develop enough versions (for all required evaluation points and an extra version) of the agreed upon subtasks. Tests should be developed in the same languages as the languages used at the project baseline.
3. Conduct a pilot on a small sample (75-150) of students:
  - a. The pilot should include enough versions (all required evaluation points, including an extra version) of the agreed upon subtasks. This will allow the project to analyse which versions are most closely matched and to be used for each GEC-T evaluation point.
  - b. The FM would suggest 75 to 150 students to be a sufficient sample. Each student must take all versions of the agreed upon subtasks to pilot.
  - c. Students should be at appropriate grades as the cohort in order to avoid floor effects on the test (for example lots of children scoring zero words per minute).
  - d. The students should not be selected if they will potentially make up part of the project's GEC-T intervention or control sample at any point of the programme
  - e. The pilot can be conducted in any number of schools and does not need to be geographically representative.
4. Analyse the scores for the pilot assessments of the different versions and adjust versions as required. The results of the pilot should be assessed to see how closely all of the versions of the test align. Following these results, if any of the versions have results that indicate they are significantly of an easier or higher level of difficulty than the other versions piloted, that version should be adapted accordingly. The tests should be re-piloted if significant changes are required.
5. Tests should then be submitted to the FM and 'signed-off' in order to become the final tests to be used for GEC-T.
6. Ensure that the tests developed involve the external evaluator and that the tests are not distributed prior to baseline being undertaken. Projects should be clear who has access to the tests, and if the project has access, it will need to provide an assurance that midline and endline tests will not be distributed to project schools prior to the evaluations taking place.

### *Ceiling and floor effects*

The purpose of introducing higher level tests (SeGRA and SeGMA) and piloting tests is to ensure the avoidance of ceiling and floor effects at evaluation points. The agreed upon framework of the tests and subtasks to opt-in to should allow for all ranges of performance to be picked up (i.e. normal distribution). It should also provide enough difficulty to allow proper measurement of progression over time. This means the particular subtasks selected and aggregation of these subtasks should be consistent at each evaluation point. If a ceiling or floor effect is experienced at any evaluation point, bespoke solutions can be discussed with the FM. (See illustration above)



### 5.6.2 Rigorous review of learning outcome results

As with GEC-1, the FM will conduct a rigorous review of learning outcome results, particularly as they will now be linked to payment by results (PbR) payments for all projects (excluding those without comparison groups). This replication process will rely on the submission of raw data used for filling out the outcomes spreadsheet and any regressions undertaken, so external evaluators will also be expected to retain and submit their underlying data at each evaluation point.

The following sections provide guidance and principles that should determine learning targets for the GEC-T logframe outcome indicator on learning at subsequent evaluation points. This will also be linked to PbR as outlined in the Handbook. Please note that the Outcomes Spreadsheet is used to automatically calculate learning targets based on your project data – the Outcomes Spreadsheet is compulsory GEC-T evaluation tool for all projects, including those with a specific exemption from PbR.

### 5.6.3 Establishing the learning outcome target – 0.25 standard deviations

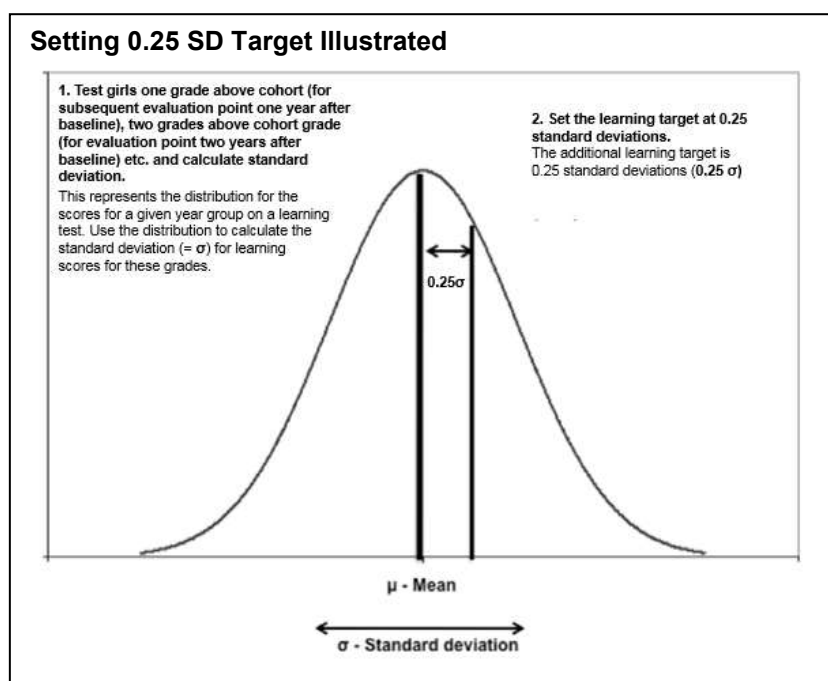
It is recognised that targets for GEC-T should reflect the programme ambition, as well as being achievable. The targets for GEC1 were based on standard deviations of improvement over and above a comparison group, or over a benchmark for projects without comparison groups.

As with GEC1, projects will have to measure the **additional outcomes for a cohort of marginalised girls that the project is generating over and above what would have happened in the counterfactual of no intervention.** The learning tests on the cohort will be applied at baseline, and subsequent evaluation points – to both intervention and comparison groups. External evaluators will undertake learning assessments either within the household or in the school or both and these tests will form the basis of learning targets and achievements on the learning outcome.

The learning target for GEC-T projects will be **0.25 standard deviations per year of implementation** – applying to literacy, numeracy and the third learning outcome. (See illustration)



Performance will be measured in the form of an additional achievement for girls in intervention schools on their literacy and numeracy learning assessments over and above the increase achieved by girls in comparison schools at subsequent evaluation points – a difference-in-difference methodology.



### Target setting and benchmarks for learning

Target setting for learning will follow a similar methodology to GEC1, where benchmark results will be taken from older grades that learning beneficiaries are expected to move into. Based on the standard deviations of the collected results, a target will be calculated. Progress against this target is measured using a difference in difference approach.

Of importance to the sampling approach is that this target setting approach is based on a benchmark sample. This may include individuals to be tracked in the longitudinal sample, as well as 'one-off' sampled individuals where their results from learning tests are solely taken for the purposes of establishing a benchmark.

This benchmark sample should be identified and sampled using the same framework as for the individuals the project and evaluator decide to track longitudinally, i.e., there should be limited bias on the individuals chosen as a benchmark. Suitable demographic information should be collected to check that the bias in the sampling technique has been minimised. Benchmark learning tests need only be administered to older grades of girls in beneficiary schools.

*Illustrative example: Benchmark sampling approach (refer to the table below and corresponding explanation below):*

Baseline	Midline (1 year later)	Endline (2 years later)
<b>Project grades</b>		
<b>S2</b>	<b>S3</b>	<b>S4</b>
<b>S3</b>	<b>S4</b>	<b>S5</b>
<b>S4</b>	<b>S5</b>	<b>S6</b>
<b>Benchmark grades</b>		
<b>S5</b>	n/a	n/a
<b>S6</b>	n/a	n/a

Identify the upper boundary of your learning beneficiaries: assuming that they will stay enrolled in the interventions through the life of the project, what is the last stage that the project could actively work with these individuals to improve their learning? Evaluators should collect baseline learning information for all grades between the lowest baseline grade and this upper limit benchmark grade.

*Illustrative example:*

- A three-year project is targeting their learning interventions initially at individuals about to move into grades S2 to S4 in 200 schools.
- Over the course of the lifetime of the project, the learning beneficiaries would be expected to move into higher grades, and by the time of the end of the project, the learning beneficiaries would have just completed grades S4 to S6.

Identify the necessary sample size needed for treatment and comparison areas (see relevant section in this guidance document). From this total, decide how many individuals will need to be sampled from each grade, including benchmark grades. Benchmark grades should need fewer individuals than the grades intended to be tracked.

- A sampling framework is developed (see relevant section in this guidance document) to choose individuals to administer learning tests to. The framework identifies 100 randomly selected intervention schools, and 100 comparison schools from an appropriate comparison district.

Apply the project's sampling framework approach to all treatment and comparison areas, identifying schools, and individuals within the appropriate grades to be selected to sit learning tests

- Individuals fulfilling marginalisation criteria are sampled randomly from the identified schools in grades relevant to the evaluation: in this case grades S2 to S6.
  - As the principal beneficiary group, larger samples are taken from grades S2 to S4. Appropriate household information is collected from the individuals in these grades to facilitate re-contacting them at their households in this and subsequent evaluation points.
  - As benchmarking grades only (blue and green), those in grades S5 and S6 are not expected to be tracked over the course of the evaluation, and so a smaller number of responses should be collected. The same criteria and randomisation approach should be made, however, to ensure that the benchmark individuals collected are representative of the beneficiaries they are expected to act as a benchmark for. No extra household information need be collected – purely the learning test data.
  - Note, the benchmark individuals for those moving into grade S2 at baseline are those moving into S3 and S4, who are identified and sampled as beneficiaries, not benchmark grades (red and orange). Where older beneficiaries can act as benchmarks for younger beneficiaries, no additional benchmark sampling needs to be done.

Once appropriate individuals for beneficiary and benchmark learning samples have been identified, learning tests should be administered, with the results stored and recorded in the project outcomes spreadsheet. This spreadsheet will automatically calculate learning targets based on the results of the samples from treatment and comparison groups, including the benchmark individuals



In summary, the benchmarking process for learning requires a separate benchmark sample to be established, and this sample will not be tracked.

### *Learning targets for out-of-school girls*

For projects with out-of-school girls in their cohort, the learning target for them should also be an additional learning target over and above changes in control areas. For out-of-school girls, the target for the next subsequent evaluation point after baseline in most cases will be 0.25 standard deviations of the average of all in-school girls' scores on learning assessments. However, if the out-of-school girl cohort includes drop-outs, projects will need to further discuss and agree benchmarking sampling approaches with the FM.

## 5.6.4 Calculating learning achievement

### *The Difference in Difference Approach and calculating PbR on the outcome*

The following section provides examples of calculating learning achievement using the Difference in Difference Approach for EGRA and EMGA style assessments. Further guidance on calculating learning achievement for SeGRA and SeGMA sub-tasks will be provided following the design of the tools. Learning outcome achievement for the GEC-T evaluation is always reported in percentage terms against the 0.25 SD target over and above the control group (difference in difference). This percentage informs the project's eligibility to receive a PbR bonus payment.

It should be noted that the aggregation of sub-tasks (i.e. EGRA/EGMA and SeGRA/SeGMA) which inform the learning achievement calculations will be agreed on a case by case basis by all projects with the FM.

#### *Example 1: EGRA style test: Short paragraph – WPM sub-task*

This example illustrates how the learning achievement is calculated using the Difference in Difference Approach, reported against the 0.25 SD target, and finally applied to PbR. For simplicity, only the Short Paragraph – WPM sub-task from the EGRA style test is used. The table below demonstrates this example in action.

1. The baseline assessment shows the mean score in EGRA is 20 words per minute (wpm) in intervention schools and 21 wpm in control schools for Grade 2 girls.
2. The learning outcome target for literacy has been determined at midline (i.e. one year after baseline) as an additional 3.75 wpm as measured by EGRA.
3. The midline assessment is undertaken on the cohort of girls, these are the same girls assessed who were at Grade 2 at baseline. This shows mean scores of 39 wpm in intervention schools and 35 wpm in control schools.
4. This means that the mean improvement for girls on EGRA compared to baseline is 19 wpm in intervention schools and 14 wpm in control schools.
5. Assuming the control group are similar in relevant respects, this means that the intervention has secured an additional 5 wpm for marginalised girls on EGRA.
6. The learning target for literacy is therefore achieved in full with some over-achievement.

For projects, the external evaluator should collect, analyse and submit these calculations to the FM. The FM will also require access to the raw data showing the learning scores achieved by treatment and control cohorts such that we can observe any changes in the distribution of attainment within projects' cohorts.

	Learning mean		Standard deviation ( $\sigma$ )	Target score ( $T = 0.25\sigma$ )
-	Intervention group	Control group		
<b>Baseline</b>				
Grade 2 cohort of girls (A)	20 wpm (1)	21 wpm (2)	-	-
Grade 3 girls (the grade above the cohort at baseline)	30 wpm (3)	-	15 wpm (4)	+ 3.75 wpm $T = (4) \times 0.25$
<b>Midline (one year after baseline)</b>				
Cohort of girls (A)	39 wpm (5)	35 wpm (6)	-	-
Change on baseline	19 wpm (5) – (1)	14 wpm (6) – (2)	-	-
Additional midline score (Difference in Difference)	+ 5 wpm $\Omega = ((5) - (1)) - ((6) - (2))$		-	-
Learning achievement	= +5 / +3.75 = 1.33 (133% achievement of target) $\Omega / T$		-	-
PbR	Eligible for 33% bonus payment			

### Example 2: EGMA

The following table is for illustrative purposes only if a project were to equally weight sub-tasks from only the EGMA style assessment for a final numeracy score out of 100%. The same process is followed as above in Example 1: the difference in difference approach is applied, the learning achievement is determined against the 0.25 SD target, this is then applied to PbR bonus payment eligibility. Again, aggregation of sub-tasks (including SeGMA subtasks) will be agreed on case by case basis for each project with the FM.

Sub-task	Test level	Numeric score	Example Weighting
Sub-task 1	Number identification	/ 20	10%
Sub-task 2	Quantity discrimination	/ 10	10%
Sub-task 3	Missing number	/ 10	20%
Sub-task 4	Addition and Subtraction Level 1 and 2	/ 25	20%
Sub-task 5	Subtraction Level 1 and 2	/ 25	20%
Sub-task 6	Word Problems	/6	20%

## 5.7 Communicating and presenting learning result findings

Detailed guidance on how to report learning data will be communicated closer to each evaluation point when templates for evaluation reports are provided. This section nonetheless highlights three basic principles that projects should consider whenever they present learning results to internal and external audiences.

### *Reporting against benchmarks*

It is critical that all learning results are presented against the key benchmarks of (1) set targets, and (2) previous evaluation points i.e. against baseline and/or midline, whichever is relevant. When presenting results against either benchmark, projects should clearly also show how the performance of treatment groups compares to that of the relevant comparison (i.e. over and above). In addition, projects are encouraged to report results against benchmarks for key sub-groups. These could for instance be by geography, grade, and/or marginalization level.

### *Reporting in a meaningful way*

When reporting in GEC-T, projects will be expected to convert the standard EGRA/EGMA subtask results into more meaningful learning categories when providing narrative on results within evaluation reports. From a reporting perspective, it is expected that projects will report on the *meaningful, functional progression* of girls' learning across all sub-tasks tested (e.g. and not just focus solely on WPM scores or similar).

These categories will follow progressions normally found in the assessment industry such as for example "Recognizes Letters", "Recognizes Words", and "Can Read Connected Text at Grade 3 level", etc. for Literacy. For Numeracy, these categories will for instance include "Recognizes Numbers", "Can do Simple Addition", etc. In other words, rather than reporting on mean score and standard deviation for the EGRA sub-task one on letter recognition, projects will instead report on what percentage of students achieved the learning category of "Recognizes letters". Note that subtask reporting in the current form will however need to continue within the data that is submitted to the FM (i.e. stated as achievement of the intervention group over and above the comparison group against a target). Moreover, WPM reporting will also continue as per the current requirement.

This conversion to learning categories is expected to allow more meaningful discussion on what beneficiaries are actually learning, and should allow learning results to be interpreted and compared within the wider learning context. This requirement should also put GEC-T learning results reporting in line with global trends on reporting assessment findings.

The below tables summarize the relevant learning categories for GEC-T, which are derived from EPDC categories.

### *Literacy*

Learning category	EGRA/SeGRA equivalent
Cannot read	% of students who have zero scores on the letter sub-task
Recognizes letters	Inverse of % of students who have zero scores on the letter sub-task
Can read words	Inverse of % of students with zero scores on familiar words sub-task
% of students who can read a grade 3 story	Inverse of % of students with zero scores on the reading subtask

## Numeracy

Learning category	EGMA/SeGMA equivalent
Is not numerate	% of students who have zero scores on the letter sub-task
Recognizes numbers	Inverse of % with zero scores on number identification
Can do simple addition	Inverse of % with zero scores on addition
Can do simple subtraction	Inverse of % with zero scores on subtraction

In presenting learning results against the previous evaluation point, projects should also consider the use of a flow diagram which illustrates the proportion of students moving from one learning category to another. The below table illustrates this principle for numeracy.

	Midline Learning Category			
		Recognizes numbers	Can do simple addition	Can do simple subtraction
	Recognizes numbers	55% of students who achieved highest category of number recognition at baseline still can only recognize numbers	35% of students who achieved highest category of number recognition at baseline can now do simple addition	10% of students who achieved highest category of number recognition at baseline can now do simple subtraction
	Can do simple addition	5% of students who achieved highest category of simple addition at baseline can now only recognize numbers	40% of students who achieved highest category of simple addition at baseline can still only do simple addition	55% of students who achieved highest category of simple addition at baseline can now do simple subtraction
	Can do simple subtraction		5% of students who achieved highest category of simple subtraction at baseline can now only do simple addition	95% of students who achieved highest category of simple subtraction at baseline can now still do simple subtraction

## Reporting within the context

In the analysis of learning results, it will be important for projects to further contextualise their results, commenting on the relevance of the level of skills achieved in relation to expected learning levels for their grade, and what the improvements in learning mean for girls' daily lives including potential contributions to addressing their practical and strategic gender needs.

Projects can do this in a variety of ways and should adopt whatever way they feel is most appropriate. For instance, they could compare what percentage of students can do addition in Grade 2 to the country's curriculum, which expects mastery of this topic by this grade. Alternatively, they could comment on other learning results from within the country or region, comparing their own findings and explaining why they may be better or worse. Finally, projects could also comment on for instance the performance of a more marginalized group against say that of a less marginalized one, and explain what this might mean for their transition to a further grade or to employment.

## 6. Outcome 2 – Transition

Transition forms the second outcome for the GEC Transition Window. The transition outcome constitutes a new outcome on the programme logframe for GEC-T projects. The wording of this logframe outcome indicator is:

**Number of marginalised girls who have transitioned through key stages of education, training or employment (primary to lower secondary, lower secondary to upper secondary, training, employment or other)**

As a logframe outcome, it is expected that all projects will collect and report data on transition. Additionally, each project will have targets set for each evaluation point against which performance will be measured. This section of the guidance covers:

- The GEC-T definition of transition and the expected ambition of projects in addressing barriers to successful transition
- Measurement of the transition outcome both quantitatively and qualitatively

### 6.1 What is transition?

Transition in educational contexts can be described and understood in various competing ways. This section defines the understanding of transition being adopted for GEC-T.

#### *Definition 1: Transition*



**Transition** in GEC-T is understood as: progression into and through successive grades of formal and non-formal education, vocational training, or into safe, fairly-paid employment or self-employment.

By formal education, the definition is referring to educational institutions such as primary, secondary, tertiary and vocational training institutions<sup>8</sup>. Non-formal education refers to classes and interventions outside this hierarchical system that serve particular groups, but that have identifiable learning objectives linked to the formal system<sup>9</sup>. Non-formal education is often designed to open up formal structures to excluded groups e.g. child labourers, those who have dropped out. Informal education may refer to, for example, ALPs or transitional courses.

Vocational training can be understood as courses designed to equip individuals with applied and practical skills that aim to prepare individuals for success in employment or other aspects of economic life. In GEC-T these aspects will widely be referred to as Technical Vocational Education and Training, or TVET<sup>10</sup>.

For some projects, transition into employment will be an appropriate transition pathway for their beneficiaries to follow, but this will largely depend on the age and demographic of their beneficiaries. For those projects where employment is defined as an appropriate transition pathway, as a minimum standard the FM expects this employment to be physically safe, non-

<sup>8</sup> UNESCO defines formal education as, “the hierarchically structured, chronologically graded ‘education system’, running from primary school through the university and including, in addition to general academic studies, a variety of specialised programmes and institutions for full-time technical and professional training.”

<sup>9</sup> UNESCO defines non-formal education as, “any organised educational activity outside the established formal system - whether operating separately or as an important feature of some broader activity - that is intended to serve identifiable learning clienteles and learning objectives.”

<sup>10</sup> UNESCO defines TVET as, “those aspects of the educational process involving, in addition to general education, the study of technologies and related sciences and the acquisition of practical skills, attitudes, understanding and knowledge relating to occupation in various sectors of economic life.”

exploitative, and fairly-paid to provide an appropriate income for the individuals' situations. Project evaluations will be expected to measure the extent to which the employment of their beneficiaries fulfils these criteria, and evaluators should structure household surveys to gauge, at a high level, if an employed individual who is interviewed is working in a safe, and fairly paid environment.

While projects will have less influence on working conditions per se, they should be empowering individuals with the self-efficacy and awareness to demand these basic rights. Projects considering employment as an appropriate pathway for their beneficiaries should consider how to best encourage a local employment market that is both safe and fairly-paid.

The table below lays out some further distinctions to the above definition, on what counts and does not count as a 'successful' transition under the above definition. More detail on each debate is given beneath the table.

**Table 6: Further defining transition**

<b>Issue</b>	<b>What may be considered as transition</b>	<b>What is not considered as transition</b>
Between-school movements	Change in enrolment of children from informal to formal institutions and vice versa, i.e. successfully progressing from one stage of education to the next	A 'transition to life' where projects equip girls with crucial life skills for a post-education life. In most cases, 'transition to life' will be understood as early drop out from school, and will not be measured as a successful transition <sup>11</sup>
Within-school progression	Within schools, the progression of children from current grade to the next grade	Within schools, the repetition of grades due to poor performance or attendance in current grade <sup>12</sup>
Transition to employment	If age and context appropriate, the movement of young people into paid employment <sup>13</sup> (formal or self-employed)	The transition from education into unemployment or unpaid work

Transition on the GEC is best understood in terms of the pathways that girls follow. These pathways map the enrolment trajectories that girls could move to over time in the project. While girls who have previously dropped out of school may remain that way, project activities may actively encourage these girls to re-enrol into an accelerated learning programme (ALP). These girls have, therefore, followed a successful transition pathway as a result of the project activity.

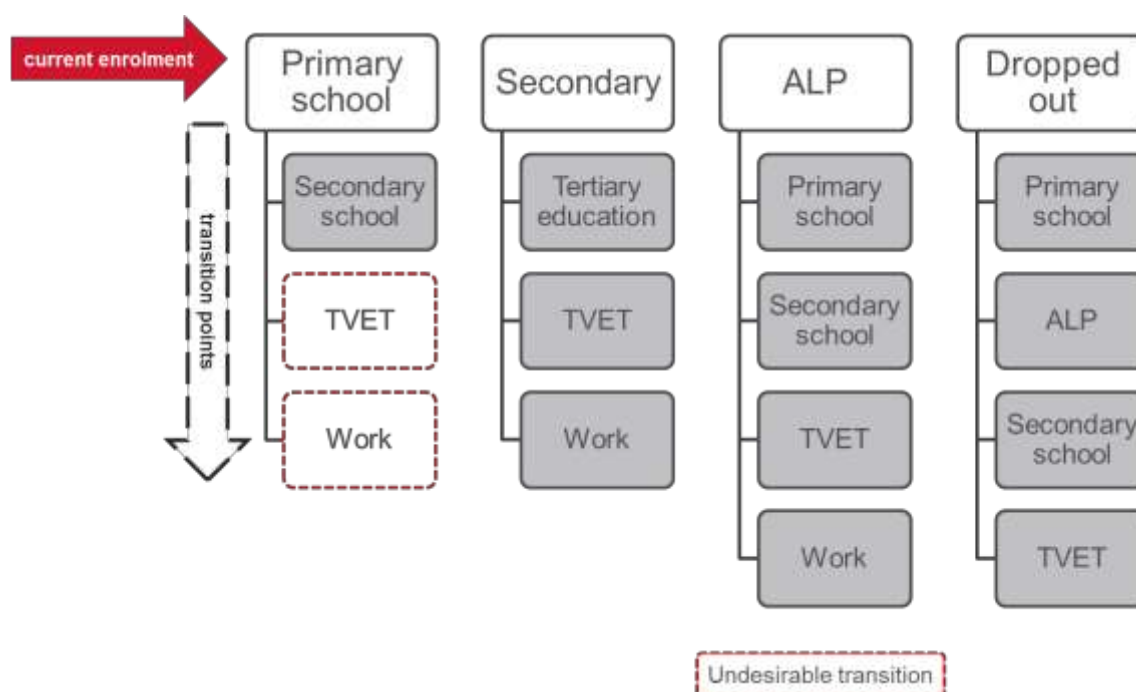
<sup>11</sup> This will depend on the age of beneficiaries and the stage they are at on their educational journeys. Projects will have to define clearly, within national contexts, at which point leaving school may not be considered a premature drop out.

<sup>12</sup> The potential adverse incentives for projects to inappropriately push children through school grades when they may not be prepared to meet the challenges of the higher grades is noted. There are two main responses to this issue: 1) the purpose of transition and the 'follow the girl' principle on GEC-T is to encourage projects to tailor ambitious but realistic transition pathways for each individual girl. We expect projects to consider the needs of a beneficiary ahead of the desires of an evaluation; and 2) the evaluation is set up to encourage ambition from projects. It should be the aim of a project to keep its supported children moving through educational stages at the normal rate expected of them. Where projects can argue that a repeat of grades is a good outcome for particular individuals, then with the right evidence and argument this could be agreed with the FM to be counted as a successful transition. Projects should note that such circumstances will be treated as exceptions to the rule.

<sup>13</sup> These definitions will vary depending on the individual contexts and theories of change for each project. Projects will be expected to articulate a 'successful' transition pathway for their beneficiaries. This will be explored more in the sections on ambition and measurement techniques below.

Projects should be able to express their ambitions for their cohort in a similar way to the figure below. Some projects may wish to organise this in terms of the age ranges they work with, or in the different 'current enrolment' positions as shown below. This gives clear indication of the different pathways beneficiaries are expected to follow, and allows evaluators to hold the project to account on these ambitions.

**Figure 1: Potential transition pathways**



Projects should look to recreate such an analysis in their MEL frameworks.

## 6.2 Ambition for transition on GEC-T

The transition measure has been designed in such a way as to encourage projects to be ambitious in removing barriers for girls' transition through education and on to paid work. The GEC considers the promotion and acquisition of life skills<sup>14</sup> as an essential enabler of transition, equipping and preparing adolescent girls for their transition into adulthood, particularly in contexts where access to appropriate information, guidance and role models is limited. As such, intermediate outcomes should help lay the groundwork and support successful transition e.g. acquisition of financial literacy skills. Transitions should not only be *supported* by project activities, but should be actively *facilitated* for each beneficiary according to their specific needs during and after the project's implementation

## 6.3 How transition should be measured on GEC-T

As suggested above, transition on the GEC-T will be measured using a 'survival rate' approach whereby a girl's current enrolment is compared to her enrolment in the previous evaluation point.

The general process that projects and evaluators should follow is as below:

1. Define appropriate gender-specific transition pathways for different ages and marginalised groups that can be applied to girls in both treatment and comparison groups.

<sup>14</sup> See intermediate outcome guidance on life skills for detailed definitions and measurement



2. Articulate barriers and enablers to transition for each beneficiary and each potential step of their pathway. Clear activities should be linked to each barrier and enabler. The FM expects projects to apply a gender lens to analysis and response to barriers and enablers.
3. At each evaluation point, follow up with the same cohort girls, and identify where they are currently 'enrolled'. At baseline, evaluators should also find out where girls were enrolled in the previous year as well.
4. Compare this enrolment against the previous year's status and consider whether this is a 'successful' transition for the individual.
5. Record the binary result (transitioned/not transitioned) in a separate 'transition' variable for that evaluation point.
6. Compare the proportions of individuals who have successfully transitioned in both the treatment and comparison groups.
7. Consider how the transition rate at the current evaluation point compares to the transition rates for the previous evaluation points. Calculate the difference in proportions for treatment groups and comparison groups between the two points.
8. Compare the difference in transition rates between the two evaluation points for both treatment and comparison groups to understand the difference in difference. Have treatment groups been able to improve transition at a higher rate than comparison groups?

While project evaluations will be expected to capture detailed information on the specific pathways which individual girls follow, the ultimate outcome measurement of transition is a binary 'transitioned/not transitioned' judgement. The decision to class an observed girl as transitioned or not transitioned will likely depend on project context, and should be based on a clear logic. This logic may be straightforward in many places, for example if a girl was in S1 the previous year, and in S2 now, she has clearly transitioned. There are examples, however, where this distinction is not as obvious: for example, if a girl has left school because she has got married, this may not be treated as a successful transition, but if she has got married and enrolled into a vocational course then she may be considered to have successfully transitioned. Projects should draft this logic in their MEL Frameworks, and evaluators should finalise the approach. Replication of this logic against the reported results will be a key part of the FM replication of results at future evaluation points.

These steps represent the methods by which evaluations can quantitatively estimate improved transition rates, but qualitative research is essential for understanding the enablers and barriers to transition, especially those that are specific to girls. Qualitative research should take place in the same community research clusters, and should seek the views of beneficiaries, caregivers, other family members, and/or community leaders on transition.

Specific qualitative research questions on transition may seek to understand may be:

- For those who remain in education through successive years, how have specific project activities affected the girls' abilities to stay on an educational pathway?
- To what extent have girls felt empowered to make informed choices on their future decisions, even when these pathways are not in formal education?
- What are the reasons for girls dropping out of education and training altogether? What could the project do differently to react to these challenges?

- How do girls' think about their future prospects? Has the project given enough guidance to consider *how* to achieve these ambitions?
- How do teachers view girls' abilities to transition, particularly in relation to in-school boys?
- To what extent have community attitudes, including families, siblings, and local leaders, encouraged girls' successful transition?

These qualitative questions should aim to get to the root of why girls' transition down the paths they do, and should probably be conducted using key informant interviews. Other qualitative techniques should be explored.

## 6.4 Establishing the transition outcome target

The target for transition will be set, in most cases, as a percentage achievement over and above the comparison group. This target will be set according to the existing transition rate, as identified through a transition benchmark (further details given in the *sampling for transition* section), and will be set in agreement between project, evaluator, and FM following the baseline data collection.

## 6.5 Indicators for transition

The quantitative indicator for transition is:

***Number of marginalised girls who have transitioned through key stages of education, training or employment (primary to lower secondary, lower secondary to upper secondary, training, employment or other)***

It is important to note that the wording of this indicator is specified as a 'number' of girls. This means that in order to complete project logframes, evaluators will have to take the transition rates from the sample and multiply this by the overall beneficiary number. If the overall beneficiary number is 5,000, for example, and the transition rate for the treatment sample sat at 80%, then the number to be input into this logframe would be  $5,000 \times 80\% = 4,000$ . The underlying assumption is that samples will be representative of the wider beneficiary population.

This quantitative indicator is, therefore, a sample-based estimate of the overall transition achievement for each project. Qualitatively, projects may be aiming to craft indicators to capture, for example, how:

- Girls have felt empowered to make informed and relevant choices on their transition pathways that best account for their individual circumstances.
- Schools have developed strong and effective career counselling approaches, including links with relevant TVET institutions, employers, or further education institutions, to provide girls with options and inspiration for successful transition.
- Communities express supportive attitudes to boys' and girls' transitions that emphasises opportunity and choice for all.

These indicators may have quantitative elements, but an in depth consideration of qualitative aspects will likely be more fruitful.

## 7. Outcome 3 – Sustainability

GEC-T is focused on achieving outcomes for beneficiary girls in terms of improved learning and transition. The third outcome is crucial to determine whether these improvements can be sustained for future generations of girls in these communities and schools, and in the education system more broadly.

The specific wording for the sustainability outcome for GEC-T is:

***Project can demonstrate that the changes it has brought about which increase learning and transition through education cycles are sustainable.***

To assess the degree to which a project has achieved progress towards gender sensitive sustainability, the FM has developed a **Sustainability Scorecard** (please see Table 8) which aims to measure key characteristics of sustainability at a given point. This is not an exact science, and the scorecard is based on a number of assumptions which should be tested further throughout the evaluation process.

The starting premise of the Sustainability Scorecard is that GEC projects have identified the key drivers which will deliver improved learning and transition and have articulated these as Intermediate Outcomes within their project logframe and theory of change.

At this Intermediate Outcome level, the evaluation will measure whether the anticipated change has taken place. In order to ensure learning and transition within target schools and communities continues after the life span of the project, or to take this to scale in the broader education system, these changes will need to be delivered in ways that can be sustained. Measuring this at outcome level will require an assessment of the nature and depth of the sustainability (and where appropriate scalability) of changes achieved at Intermediate Outcome level.

At GEC-T proposal stage, projects were asked to develop a Sustainability Plan, submitted alongside the GEC-T draft budget and logframe. This draft plan set out approaches and specific action to ensure the gender sensitive sustainability of the output level and outcome level changes. The plan set out gender sensitive indicators to measure sustainability of the project at school, community and system levels. **The Sustainability Plan will be a working document, which the project will need to refine and improve,** to further inform project work planning, and monitoring. You are not expected to resubmit the Sustainability Plan at this stage, as the focus is on developing your MEL framework and logframe.

**The priority for the MEL framework is to agree on project-specific indicators which can be used to measure sustainability against the scorecard, and to identify methodologies for collecting and analysing data against these.** These will form part of the project logframe.

### 7.1 Planning for sustainability

There are a range of stakeholders and institutions that projects work with and influence across the GEC portfolio; locally with individuals, households, communities, and schools and more broadly within education systems and with policy makers. Through this engagement, some projects have a potential influence on broader social norms. For the purpose of measuring sustainability at the outcome level, GEC-T focuses on the changes that occur at three levels: community, school and system. These three levels have been defined in the Table 7 overleaf.

**Table 7. Sustainability on GEC – planning for change at three levels**

<b>Community</b>	From the household/family level to broader community members, and especially leaders; including structures, groups, clubs, local businesses and other agents of change that the project establishes/works with to support girls.
<b>School/provider</b>	Includes government run/funded pre-primary, primary and secondary schools, vocational and other training providers and established non-formal education providers. This may also include private or community based/owned schools.
<b>System</b>	The education system at district, provincial and national levels, ranging from policy to delivery. This includes staff and units/departments the project may work and interact with, regulations within which the project works or may be aiming to influence. This may include private markets, or a broader set of networks that influence social norms.

Having three distinct levels to the Sustainability Scorecard allows sustainability to be measured on a more holistic basis, and to understand if projects are stronger or weaker across different domains of sustainability.

Projects may comprise a mix of interventions in order to achieve improvements in learning and transition for marginalised girls. In order to measure sustainability, the evaluation needs to look at each of the key changes interventions bring about with stakeholders and institutions, and form a view of how advanced or mature this process is. **The Sustainability Scorecard will be used as a tool for assessing this and tracking progress, but it is critical that projects conduct broader qualitative analysis to form a picture as part of this.** The FM recognises that this is a complex, non-linear process. However, projects should be able to design and deliver interventions in ways which optimise and promote sustainable change and replication, including factors of cost, staff or volunteer time and delivery roles. Critically, there needs to be evidence that a change is worth sustaining and stakeholders need to see these benefits.

Sustainable change may take a number of years, for example changes in social norms may not be expected to happen in the life span of the GEC-T. However, projects should aim to be precise around their expectations and what would demonstrate success or good progress by the end of the project.

#### *A model for sustainable change at community, school and system level*

At the local levels, in communities and schools, **a first step to sustainability will likely be changes in attitude among some key stakeholders** or those most directly benefiting from the intervention. **Over time this leads to change in practice and behaviour**, which gradually extends to others, ideally reaching a critical mass of stakeholders who see the benefits of the approach. Alongside this, projects may focus on ensuring structures, capacity and resources are increasingly in place, enabling local stakeholders to lead and maintain change independently of any external support. As change is more fully established, stakeholders can further develop existing or new approaches to support girls' education and respond to needs.

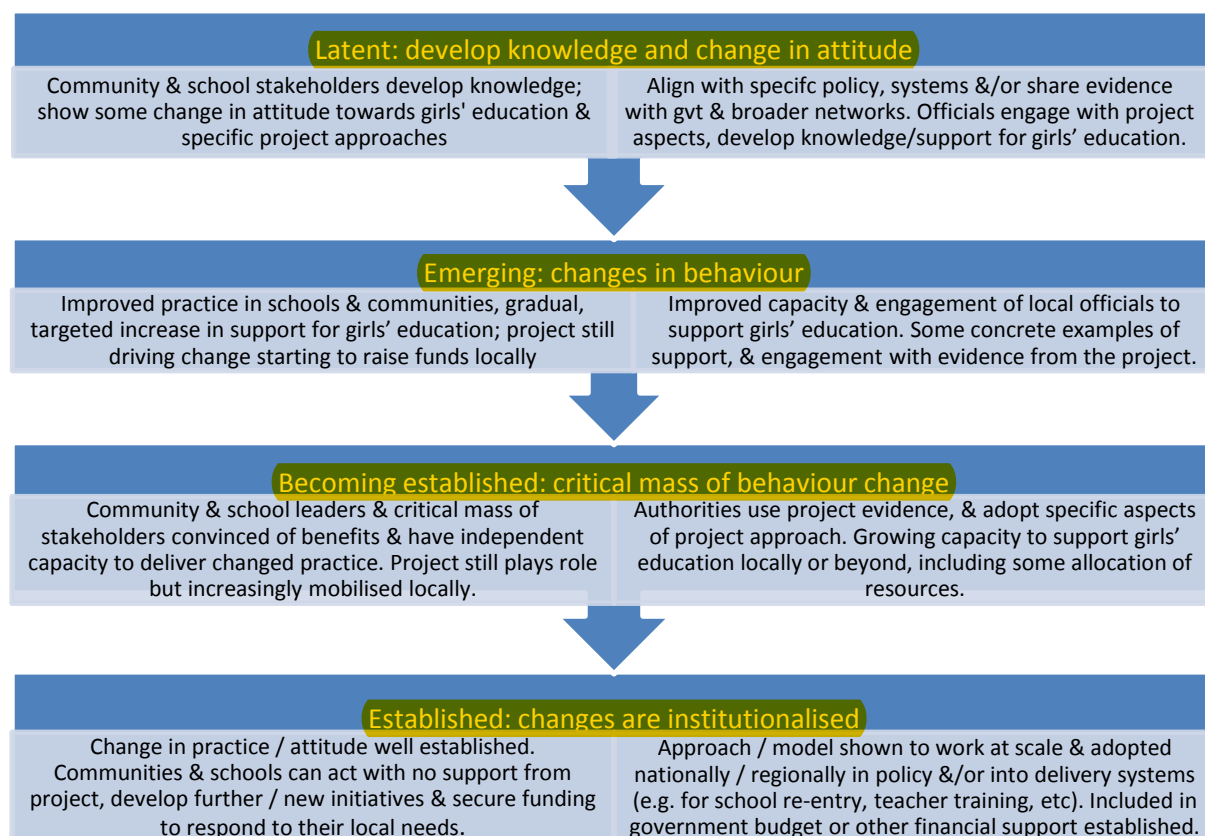
For some changes to become sustainable and for these to be taken to scale, there is a need for change at the level of the education system, which is more complex and often less within the control of projects.

Sometimes this system level change needs to be opportunistic, but it can be planned for. Initially, this may focus on involving local or national officials, or other networks and stakeholders, in the delivery or monitoring of the project or just sharing evidence as it emerges. Projects may also align with specific policies or regulations and work within existing

governance structures. By demonstrating the local benefits of an approach, a project can then seek to increase this level of engagement and the capacity and willingness of officials to support girls' education, using project evidence and adopting some aspects of the project approach. Ultimately, if an approach or model is shown to work it could be adopted in national policy and funded through the budget, or incorporated into specific delivery systems (e.g. for teacher training or school management).

The model sets out how change in practice becomes sustainable through four different stages of progress.

**Figure 5. Sustainability – a progressive model at school/community and system level**



This process will take place at different rates in the different communities and schools in which a project works. In assessing sustainable change it would be ideal to see this happening in a majority of locations. However, in the GEC-T evaluation, this also offers projects a chance to understand why it may vary and understand therefore how to adapt project delivery accordingly.

The Sustainability Scorecard incorporates all the four different stages of progress at the three levels, and has a scoring ranging from 1 to 4 depending on the stage of the progress. Please refer to Table X for further details.

## 7.2 Defining sustainability indicators and applying the Sustainability Scorecard

Projects will need to be clear on how each intermediate outcome contributes to sustainability, whether at community, school or system level, within their overall theory of change.

A specific intermediate outcome may contribute to one or more of these levels. The indicator chosen to measure change at intermediate outcome level may not capture the degree to which

this change is sustainable. Therefore, drawing on the draft Sustainability Plans, projects should identify precise sustainability indicators which will be used to assess progress against the Sustainability Scorecard. Underlying changes in gender power relations and gender norms and stereotypes can be particularly powerful in sustaining project successes beyond the project phase. Gender equity is therefore considered a key part of any sustainability plan and analysis and all indicators must be gender-sensitive.

There is a separate tab in the logframe for setting out sustainability indicators and their measurement, linking them to each intermediate outcome. All projects will need to set specific sustainability indicators for each of the three levels: community, school and system. These indicators should build on the measures identified at intermediate outcome level. There can be a maximum of three indicators for each level. Targets for each evaluation point should be used to set out expected progress and should relate to a point on the scorecard.

Projects should consider the following when developing their sustainability indicators:

- At school and community level, a measure capturing the achievement of a critical mass of behaviour change may be needed (i.e. among parents, both male and female, siblings, teachers, community members and/or leaders and others who see the benefits of change). The project should identify what would constitute this critical mass at a local level. For progress against the scorecard, it is expected that the process of achieving a critical mass of behaviour change would take place in a significant majority of project communities and schools.
- If a new practice or delivery model (e.g. clubs, classes, technology) is key to change, then sustainability may be determined by evidence of locally led or owned systems to incorporate this (e.g. in to school operations, community action) and provide local resources (including funds, staff or volunteer time). Evidence of action independent of project support should be sought, particularly from school and community leaders.
- At system level, projects should consider which specific authorities and actors, including gender and womens' issues authorities are expected to adopt a measure or approach and what they would need to do this (e.g. capacity, staff resources/time, funding, specific regulation, etc.). Projects should avoid setting indicators which look for a stated commitment (although this may be a good start), but rather some credible evidence that the commitment is being or will be put into practice. This would help projects to achieve a higher score on the Sustainability Scorecard.

Each sustainability indicator will need a **means of verification**, which draws on a mixed method approach. As part of this, good qualitative research needs to be planned and undertaken by projects' external evaluators. Indicators in the logframe will include quantitative proxy measures of change, but the qualitative work will be key to understanding sustainability better and justifying a score against the Sustainability Scorecard.

The MEL Framework will need to include sources of data for the sustainability indicators. There may be some overlap with intermediate outcome measurement and therefore it may make sense to use some of the same tools (e.g. household survey, school surveys, FGD, etc.). However, it is important to note that the sustainability indicators may require some specific modules and/or sections within these tools or may require their own data collection tools.

At the system level in particular, it is likely that specific tools to gauge government capacity and take up of approaches will be needed. The Sustainability Outcome tab in the logframe will need to specify what tools will be used to assess performance against the indicators and set a value against the Sustainability Scorecard. Each MEL framework will need to set out the



timing of data collection, an overview of the tools to be used, and who will collect the data for these indicators.

The external evaluator should score the project based on the data collection at baseline and all subsequent evaluation points, using the Sustainability Scorecard. The FM team will review and moderate this score as necessary, recognising that this requires a degree of judgement. While the data collection for the sustainability indicators will only happen at evaluation points, the sustainability outcome will be incorporated into project annual reviews to assess progress and adapt programming as needed.

### **7.3 Financial sustainability**

The GEC-T approach on sustainability set out in this guidance document assumes that financial sustainability is a key part of sustainability and is incorporated in the Sustainability Scorecard. For a project to achieve scores of 3 or 4 on the scorecard, for example, some level of financial sustainability is likely to be required. Where appropriate, projects should include specific indicators against which to assess the level of financial sustainability. The Sustainability Plan will be the place to set out in more detail how the project understands the current costs of delivering change and what these costs are likely to be for communities, schools and the broader education system. It needs to recognise that delivery by a project does not always involve the same costs as delivery by government or other stakeholders.

Some key considerations for financial sustainability include:

- Match funding at the start of the project (i.e. pre-committed funds) and at endline (i.e. funds leveraged during project implementation);
- Estimated annual cost of continuing core interventions in the years after GEC-T project funding ends (e.g. for continued payment of incentives, technology maintenance, refresher training, stipends, etc.), and who is expected to bear these costs.

The MEL Framework should make clear how financial sustainability is incorporated within the scorecard approach, in specific terms for each project at the three different levels. A suggested approach is that, where appropriate, one of the three indicators set for each level would be a financial or resourcing-related indicator.



**Table 8. GEC Sustainability Scorecard**

Rating	Community	School	System
<b>0 – Negligible</b> <i>(null or negative change)</i>	No evidence that community members accept the project approach, and changes in attitude or engagement with activities very limited. Stakeholders may even reject key aspects of project. Project not working effectively to build consensus or support, but focus only on activity implementation.	No evidence that school stakeholders accept the project approach, and changes in attitude or engagement with activities very limited. Stakeholders may even reject key aspects of project. Project not working effectively to build consensus or support, but focus only on activity implementation.	Very limited and ineffective engagement with system level stakeholders, including district or national authorities. Authorities do not see relevance of intervention. There is limited alignment to existing systems / structures and policies, or limited understanding by project of how it intends to influence change at this level.
<b>1 – Latent</b> <i>(changes in attitude)</i>	Community stakeholders (including parents, community leaders, and religious leaders) are developing knowledge and understanding and demonstrate some change in attitude towards girls' education. Appropriate structures are being put in place at community level, and there is some level of willing engagement and/or participation from the community.	School leadership, teachers and other stakeholders are developing knowledge and understanding and demonstrate some change in attitude towards girls' education in general and towards specific teaching practice and approaches, and the way schools are managed.	Local, district, and national officials are involved in delivery and/or monitoring; developing knowledge, and showing change in attitude towards girls' education and project focus areas. Project aligns with specific policy, systems and departments. Project's evidence is being shared with relevant stakeholders, including broader networks of organisations.
<b>2 – Emerging</b> <i>(changes in behaviour)</i>	There is evidence of improved practice and support for girls' education in specific ways being targeted by project. Change is not universally accepted among targeted stakeholders, but support is extending. Project staff and resources play key role in driving change, although there	There is evidence of improved support for girls' education in classroom practice, teacher management, and school management being targeted by project. The improved practice is not universal, but is extending. Project staff and resources play key role in driving change. School leaders	There is evidence of improved capacity of local officials to support girls' education through existing functions, adopting new approaches. Examples of support to project schools are being established. Government at local and/or national level has engaged with and understood

Rating	Community	School	System
	are activities in place to mobilise funding/other resources.	understand resource implications and mobilising funds locally.	evidence from the project. Resource implications are being made clear.
<b>3 – Becoming established</b> <i>(Critical mass of stakeholders change behaviour)</i>	Key community leaders and a critical mass of stakeholders are convinced of the benefits and have the capacity to lead and deliver changed practice independently. Financial and other resources are increasingly being mobilised locally. Project staffing and resources still play role but there is potential for this to be phased out.	Head teacher and critical mass of school staff and stakeholders convinced of the benefits and have the capacity to deliver changed practice independently. To the extent possible, existing financial and other resources are being used or mobilised. Project staffing and resources still play role but there is potential for this be phased out.	Authorities demonstrate active use of project evidence, uptake of specific aspects of the project approach and have a growing capacity to support girls' education locally or beyond. This may include limited support to a delivery model without fully adopting within a national system. There is an increase in allocation of resources and evidence of planning for required resource to upscale.
<b>4 – Established</b> <i>(changes are institutionalised)</i>	The specific change in practice and attitude is now well established. Communities demonstrate independent ability to act without support from project, are able to further develop existing and new initiatives and secure funding to respond to their local needs to sustain and build on the changes that have taken place.	The specific change in practice and attitude is now well established with school level systems to support this; schools demonstrate independent ability to act without support from project, have allocated and mobilised financial and other resources and are able to develop further initiatives to respond to local needs to sustain and build on the changes that have taken place.	An approach or model is shown to work at scale and is being adopted in national policy and budget as appropriate, and/or incorporated into key delivery systems (e.g. for teacher training, curriculum, school management etc.). There is an established track record of financial support.

## 8. Intermediate Outcomes

As detailed above, the core Outcomes of the GEC-T will be learning, transition, and sustainability. In addition to this, GEC-T introduces a level in logframes between outputs and outcomes - **Intermediate Outcomes**. These will be the key steps within the project theory of change (ToC) identified as essential enablers to improve learning, transition and sustainability.

Projects on GEC-T have one compulsory intermediate outcome: **attendance** (formerly an outcome indicator for GEC-1). In addition, each project is required to have their own key intermediate outcomes relevant to their ToC and the barriers to education they seek to overcome. Intermediate Outcomes form a core part of the evaluation for GEC-T, entailing robust and rigorous measurement methodologies undertaken by the independent, external evaluators. Projects have a minimum of three and a maximum of five Intermediate Outcomes including attendance, with one indicator for each of them.

### Intermediate Outcomes: core principles

1. Intermediate outcomes are a new level of indicator within the project logframe, and should measure the key steps to lead to learning, transition and sustainability.
2. Intermediate outcomes must be incorporated into the MEL Framework and be measured by the project's external evaluator at each evaluation point.
3. Attendance will be Intermediate Outcome 1 for all project logframes.
4. Projects have between two and four additional intermediate outcomes. There should be an indicator for each intermediate outcome.
5. Project logframes include gender-sensitive and disability-focused quantitative and qualitative indicators.
6. Comparison groups will not be compulsory for the measurement of intermediate outcomes, including for attendance however evaluations that do use comparison groups for intermediate outcomes, if this is possible, would be strengthened.
7. Projects will need to develop robust tools and methodologies for the measurement of intermediate outcomes alongside the MEL Framework, and refine this in collaboration with their external evaluator. These will need to be signed off by the FM before the evaluator goes to do baseline fieldwork.

### 8.1 Intermediate Outcome 1: Attendance

More complete guidance on developing strategies and indicators for measuring attendance as an intermediate outcome have been provided in earlier documents, so this section aims to build on the previous guidance. Particularly, this section builds on:

- Sampling and measurement techniques for attendance reporting
- Attendance and GESI standards for GEC-T projects
- Good practice and FM expectations for spot checks

#### 8.1.1 Measurement and sampling techniques for attendance

As described in further detail in earlier guidance, attendance data can be collected from multiple different sources. In their MEL Frameworks, projects should identify what sources they have available, and which will be most appropriate for measuring attendance. These sources may include:

- System-wide data from all project schools collected using technology-enabled registers
- Hand-completed registers from project schools
- Registers for attendance at other, non-school project interventions, such as girls clubs
- Other relevant registers that can track attendance, such as drivers' registers for those projects supplying school transport
- Self-reported attendance from girls, parents, and/or teachers based on survey responses

Projects should define the most accurate source for attendance data, and from this should develop indicators and measurements in line with what the source can measure.

If projects have dynamic, digital attendance records collected daily from all project schools, then an indicator can be developed on the average attendance rate in all schools, not just those schools sampled for learning. Likewise, for those projects with little reliable information from schools, an indicator based on a smaller tracked sample of individuals may be developed (e.g. those girls receiving one-on-one mentoring). This would require a different kind of logframe indicator, as one deals with school-wide averages, and the other deals with sampled individuals.

As a minimum standard, however, projects should be pushing their supported schools to collect and track attendance in more effective ways. Even if school data is currently unreliable, the project should react to this finding, and consider what actions are being undertaken to improve the school data collection systems.

Additionally, all evaluators will have to draw a sample of project institutions in which spot checks can be conducted. For simplicity, this sample will most often correspond with the sample of institutions selected for learning outcomes. Projects should note any exceptions to this in their MEL frameworks.

### 8.1.2 Attendance and GESI standards for GEC-T projects

As part of the GESI minimum standards laid out in the GEC-T recipient handbook, a requirement was made for projects to develop:

*“A retention strategy that captures the reasons for girls’ drop out from school and provides appropriate support to re-engage girls in response to the common issues is articulated in project activities.”*

The first step projects should take in responding to this is to define at what point girls are considered to have dropped out of each intervention. Dropout should be assessed both in terms of dropout from a formal or informal institution or class and project interventions e.g. a girls club. It is important that this information is considered in drafting the MEL Frameworks.

In the six monthly adaptation meetings being planned as part of the monitoring system, projects will be expected to comment on the extent of drop-out in the previous six months. Additionally, as the standard above suggests, capturing the reasons for drop-out is often more critical than the actual number of drop-outs.

Reporting on attendance as an intermediate outcome, however, means that in-depth data on attendance and transition (which should have some comment on the extent of drop-out) may only be collected and analysed at each project’s evaluation points. These points will be, at a minimum, one year apart, meaning that the six monthly reporting requirement may become difficult to fulfil.

This expectation of six monthly reporting on drop-out, however, remains a strong part of the monitoring approach, as it is important for projects to understand and respond to the movements of their beneficiaries in a dynamic way.

For projects without attendance systems that allow for simple regular reporting that captures numbers and reasons for drop-out, spot checks can be used as a more regular update for the monitoring approach. This is further described below.

### 8.1.3 Spot checks

The FM minimum expectations for spot checks are that:

- There should be a minimum of two spot checks carried out in project intervention institutions per year (i.e. one at least every six months).
- There should be at least one spot check per year carried out by the external evaluator, others may be conducted by project staff.
- The exact date of the spot check should be, ideally, unannounced.
- They should be carried out on a representative sample of project institutions (potentially utilising the same sample as identified for learning outcomes)

Spot checks should be designed on GEC-T to provide information on both attendance and retention. This represents a change of scope from GEC 1, where spot checks were often simply a head count comparison of who was in class against who was marked in a register. While this triangulation purpose remains a key part of any spot check, the monitoring requirements and GESI standards described in the handbook suggest a wider scope for spot checks may be required.

Spot checks on the GEC-T should aim to capture:

- A comparison of reported attendance and physical attendance on the day of the spot check, on order to give an indication of the reliability of school registers.
- The extent of enrolment, re-enrolment, retention, and drop-out in the period since the previous spot check.
- The reasons for drop-out e.g. migration, marriage, inability to pay school fees, increased domestic or caring responsibilities. These reasons should be captured through a survey to girls, teachers and head teachers.
- What follow up was done to reengage the beneficiary e.g. home visit from Project Officer of school Management Committee member. (This could be captured through qualitative interviews with girls, parents, a Head Teacher or School Management Committee).
- If reengagement was unsuccessful, what barriers remain to reenrolment/reengagement in the intervention? (This could be captured through qualitative interviews with girls, parents, a Head Teacher or School Management Committee)

Projects may wish to develop this spot check approach in their MEL frameworks, or may include it as a requirement for the external evaluator to develop the approach. At a minimum, projects should lay out the expectations for spot checks in their MEL Frameworks and evaluator terms of reference.

## 8.2 Other intermediate outcomes

In addition to attendance monitoring and reporting, recipients have proposed further Intermediate Outcomes as part of their proposal and these should be included in the logframe when submitted. Projects can have a maximum of five Intermediate Outcomes in their logframe including Attendance. Intermediate Outcomes will vary on a project-by-project basis and will need **to be aligned to each theory of change.**

Following a review of the portfolio-wide suite of interventions delivered through GEC-1, the following prominent themes have been identified. These provide an *initial* representative list of Intermediate Outcomes proposed by recipients, which link and contribute to the core outcomes of learning, transition and sustainability:

1. Teacher quality improvement;
2. Greater self-esteem and empowerment of marginalised girl;
3. Positive community attitudinal and behavioural change;
4. Economic empowerment of households and girls' families; and
5. Improved school management and governance.

This is not an exhaustive list, and projects have proposed a range of Intermediate Outcomes which are specific to their theory of change. Proposed Intermediate Outcomes can and should be further refined, and should be included in projects' Adjusted Full Proposals and Logframes, with methodologies for measurement developed alongside the MEL Framework in the first three months of GEC-T. **Data collection related to intermediate outcomes will need to be conducted by the project's External Evaluator.**

Each Intermediate Outcome should represent a critical 'step' / causal mechanism / essential enabler that is needed in order to drive results within the outcome areas of learning, transition and sustainability in the project context. For example, well-trained and motivated teachers who demonstrate improved teaching practice are critical to improving learning outcomes. The assessment of progress on these intermediate outcomes will be a key component of the evaluation at Baseline and subsequent evaluation points.

Intermediate Outcome	Notes
1 Attendance at school or another educational setting	This intermediate outcome is compulsory and should focus on attendance. This outcome links to both learning and transition outcomes.
2 Selected by the project	<ul style="list-style-type: none"><li>• Set out between 2 and 4 additional Intermediate Outcomes, which capture the key changes which will enable outcome level change. These are likely to include changes in attitudes, behaviour, capacity and use of resources by key stakeholders. The following are examples of the key enablers at this level which Intermediate Outcomes should aim to capture:</li></ul>
3 Selected by the project	
4 (Optional) Selected by the project	
5 (Optional) Selected by the project	
	<ul style="list-style-type: none"><li>• <b>Economic enabler:</b> change which enables learning and transition through reducing the economic barriers to girls' participation and progression in school. Indicators may focus on household level ability to cover the costs of education, or other ways in which resources are used effectively to help encourage this participation.</li><li>• <b>Social environment enabler:</b> positive changes to gender and social norms, for example the reduced incidence of child</li></ul>

Intermediate Outcome	Notes
	<p>and early forced marriage, or broader measures of increasing parental and community support for girls' education.</p> <ul style="list-style-type: none"> <li>• <b>Individual enabler:</b> improvements to girls' own motivation, aspiration, confidence and self-efficacy, which empower girls to negotiate their transition through and achievement at school.</li> <li>• <b>Learning environment:</b> changes in the teaching practice of teachers, the provision of safe and supportive learning environments, and improved use of teaching and learning resources in and out of the classroom environment.</li> <li>• <b>Sector and system enablers:</b> projects may seek to bring about changes to the broader enabling environment in school systems and policy, for example through teacher deployment and training, school governance which better supports the transition of girls, or specific areas of policy or resource allocation by government.</li> </ul>

### **Analysis – how does change happen in your context?**

If you haven't already done this, it's critical you conduct analyses of the context (social, gender, political, economic and technological) and power relations within which your project will operate. This provides a basis for understanding how change for girls happens in your context. This must include an understanding of child protection issues, including existing mechanisms and protection.

### **Theory of change – what change will your project effect?**

In order to develop your logframe Intermediate Outcomes, you will need to ensure the project Theory of Change is strong (see Section 1 above). Revisit if needed the TOC submitted at the proposal stage to ensure there is a clear line of logic which identifies the key changes at Intermediate Outcome to bring about outcome level change, and that project outputs link clearly to these Intermediate Outcomes.

## **8.3 Measurement of Intermediate Outcomes**

Indicators for Intermediate Outcomes should link to project outputs, and should capture the key enabling changes. For example, where an output includes training of teachers or school management committee members, the Intermediate Outcome should seek to measure the extent to which this has resulted in changes in behaviour and practice. The Theory of Change should then show how this expected change will lead to improved learning and / or transition.

As a result, each project will need to develop and set out what tools and detailed methodology are expected to be used to measure these indicators. This will be done in detail in the development of the MEL Framework, and in collaboration with the external evaluator when they are hired. Indicators and approaches will be reviewed alongside the theory of change, and will need to be signed off by the FM.

Alongside project and evaluator-developed tools, the FM will also provide 'core questions' for each of the major Intermediate Outcome areas identified. These will consist of a short set of questions per Intermediate Outcome to be administered as part of the household or school surveys. More details on the exact wording of these questions, and how they should be administered, will follow in future guidance when the full survey tools are shared.



The FM will also establish thematic discussion groups during inception phase with the aim to support projects, share good practice in measuring different types of indicators, and bring consistency in approaches to common intermediate outcomes wherever possible. However, the measurement of Intermediate Outcomes will need to be context and project specific and appropriate, and tools may therefore need to be tailored to project needs.

Measurement of Intermediate Outcomes will be carried out by the external evaluator in the Baseline and subsequent evaluation points. It will form a core part of the evaluation report alongside the reporting against Outcome indicators.

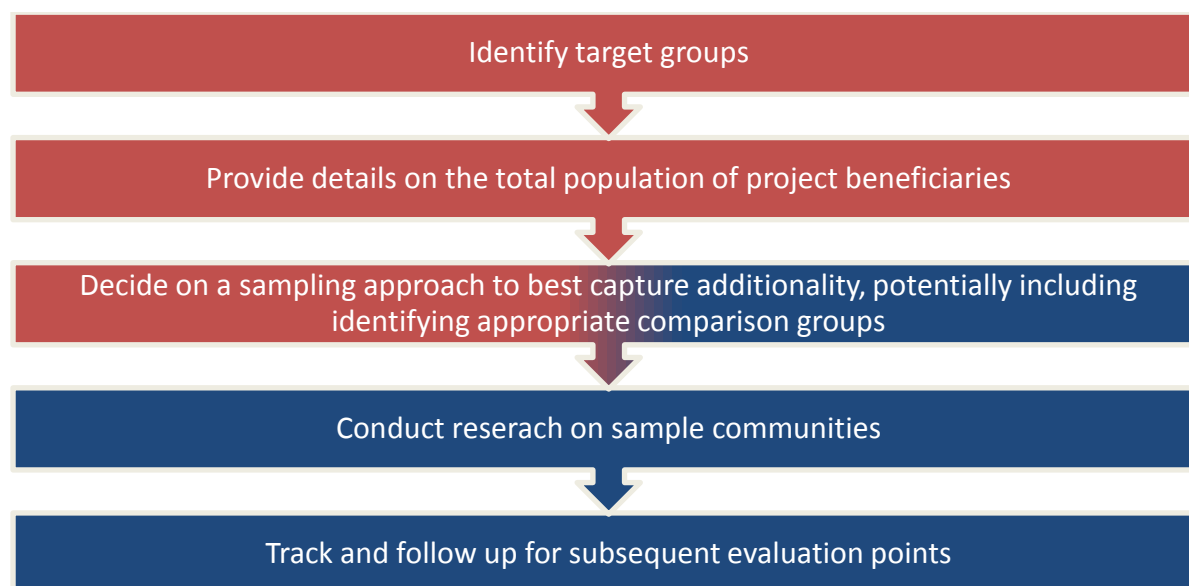
## Part 4: Methodology for outcome and intermediate outcome evaluation

The previous sections have described general ambitions and principles of the evaluation, as well as more on measurement expectations for learning, transition, sustainability, and intermediate outcomes. This section builds on the principles described, and lays out guidance on how to structure an evaluation sample to capture the required information.

The key aims for this chapter are to help projects and evaluators to:

- Describe the approach to measuring **additionality** for GEC-T evaluations
- Understand the key principles behind a representative and unbiased **sampling approach**
- Design detailed **sampling strategies** for each outcome and intermediate outcome that minimises sampling error while remaining unbiased and representative
- Set **minimum sample sizes** which are large enough to capture statistically significant results, but small enough to minimise evaluation cost and complexity
- Develop strategies for effective **cohort tracking**, especially for measuring progress against the transition outcome

The structure of this chapter is designed to follow the process of designing and carrying out an effective **longitudinal evaluation**:



The colour coding introduces the split in responsibilities between project and evaluator in crafting the evaluation approach. The initial steps, highlighted in red, should generally be carried out by the project, while the latter sections, in blue, are generally the responsibility of the evaluators. Step 3 on additionality and comparison groups is somewhat of a mixed responsibility, because projects should suggest appropriate comparison groups, while evaluators should ultimately select the comparison group.

More detail on these steps will be provided in the sections below. This part of the guidance should be read in line with the sampling framework template provided in Annex D.

## 9. Identifying target beneficiary groups

The project target beneficiary groups are the groups of girls who stand to benefit from the project's interventions across the learning, transition, and sustainability outcomes, as well as intermediate outcomes. The sum of the target groups is the project group of beneficiaries. Although all GEC-T projects have worked with their beneficiaries in the first phase, it is important that the target groups are clearly identified and described at the beginning of the second phase.

This Chapter sets out the steps that each project and its evaluator should take to identify the target groups, conduct a mapping of their intervention areas, track actual exposure of the beneficiaries to the project's interventions, and assess reach of the target groups.

### 9.1 Defining target groups

The project target groups are the groups of individuals who are expected to receive the direct benefit of the project interventions. It is important, for both sampling and wider project design, that projects can clearly articulate the defining characteristics of their specific target group.

#### *Definition 2 : Target Groups*



**Project target groups** are groups of outcome-level beneficiaries who are expected to benefit from project interventions.

Generally speaking, these groups will be characterised by:

- Geographical location of their project areas
- Project's marginalisation criteria
- Project's intervention targeting, including enrolment in specific 'intervention' schools
- Sex<sup>15</sup>

Different targeting and marginalisation criteria are usually justified by the focus on helping the groups who are at higher risk of poor educational outcomes in the long term. This risk can be associated via previous research or in-depth knowledge of a country or region, with a range of characteristics, for example those who are poor or displaced.

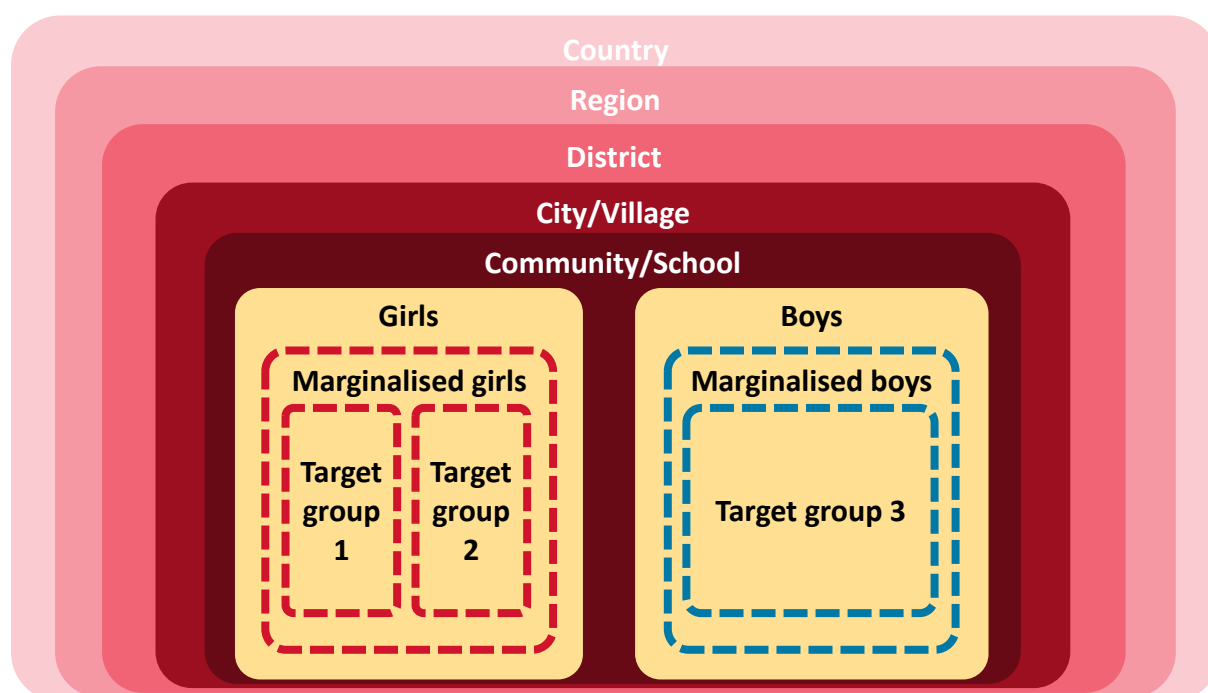
It follows that projects need to have clear definitions of each of the risk factors, targeting strategy, and marginalisation criteria that help define their target groups. This is important because if the project defines the various criteria inconsistently, then the evaluation is less likely to capture information relevant to the project's groups of interest. The criteria also need to be measured in the same way during the intervention as during the research which identified them<sup>16</sup>.

Figure 2 below gives an example of a project with three different target groups.

<sup>15</sup> Projects should note that boys are a welcome aspect of beneficiary target groups and – particularly within an in-school context – be included. However, this should not be included in an unduly disproportionate way if not part of current plans given the overall focus on girls.

<sup>16</sup> It should be noted that during the first phase of GEC, many projects first identified areas and even specific locations where most potential beneficiaries are educationally marginalised, and then targeted all girls in those areas (within a certain age range or in specific schools) with the same interventions. Even these projects should review all the targeting criteria for each of their interventions to correctly identify their target groups.

**Figure 2 - Identifying the project's target groups**



Whenever a project targets beneficiaries with different interventions, the evaluator needs to identify and characterise the different groups. To give an example, let's assume to have a project that identifies all girls and boys in Primary School classes 3 to 7 as marginalised due to low local economic development and low educational capacity of the schools. Within this group, the said project aims to deliver bikes to the specific group of girls that lives more than 5 kilometres away from the school, so they can travel to school safely and quickly and attend class more often. Then, for a different (and possibly overlapping) group of girls who are at risk of dropping out from school, the project aims to deliver bursaries to help them continue their education. Finally, the project aims to conduct training of the teachers to strengthen the numeracy outcomes for all students in the school, including boys. Such a project would have four target groups:

- **Target group 1:** In-school girls living more than 5 kilometres away from the school
- **Target group 2:** In-school girls at risk of dropping out
- **Target group 3:** All other in-school girls
- **Target group 4:** All in-school boys

In evaluation, knowing whether beneficiaries are **targeted** by different interventions is key to attributing any impact that is narrowly measured at the individual level, such as the GEC-T learning and transition outcomes. Understanding these target groups is also a key stage in developing a sampling framework: further described below.

For the purpose of developing a representative evaluation sample, therefore, projects need to conduct a detailed review of their target groups and targeting strategies. Projects should include this defining detail in their MEL Frameworks. The project also needs to inform its evaluator of its targeting strategies and regularly update about any change.

Further to the project's target groups' identification, the project's beneficiary numbers will also need to be disaggregated across dimensions such as age, school, social group and school status. The categories that identify the groups must be set in consistency with DFID reporting and are therefore common across the GEC-T portfolio. These groups are presented in the boxes below.

<p><b>By school</b></p> <ul style="list-style-type: none"> <li>• Lower primary</li> <li>• Upper primary</li> <li>• Lower secondary</li> <li>• Upper secondary</li> </ul> <p><b>By age</b></p> <ul style="list-style-type: none"> <li>• 6 – 8 years old</li> <li>• 9 – 11 years old</li> <li>• 12 – 13 years old</li> <li>• 14 – 15 years old</li> <li>• 16 – 19 years old</li> <li>• &gt; 19 years old</li> </ul>	<p><b>By social group:</b></p> <ul style="list-style-type: none"> <li>• Disabled girls</li> <li>• Orphaned girls</li> <li>• Pastoralist girls</li> <li>• Displaced girls</li> <li>• Slum-dwellers</li> <li>• Poor girls</li> <li>• Disadvantaged caste/ethnic minority</li> <li>• Affected by HIV/AIDS</li> <li>• Young mothers/expecting</li> <li>• Street Children</li> </ul> <p><b>By school status</b></p> <ul style="list-style-type: none"> <li>• Out-of-school girls: have never attended school</li> <li>• Out-of-school girls: have attended school, but dropped out</li> <li>• Girls in-school</li> </ul>
---	---

Projects and evaluators can track these key characteristics of their target groups using both management information trackers and evaluation tools such as the household survey. It is the responsibility of the project to record characteristics of its target groups; similarly, it is the responsibility of the evaluator to record characteristics of the sample of girls and boys tracked in the evaluation, and to check that the sample matches with the intended beneficiary population.

As a minimum standard on gender equality and social inclusion in GEC-T evaluations, evaluators should collect information on all sampled individuals that demonstrates, at a minimum, the individual's gender and disability status. Disability status should be derived using the Washington Group short set of questions, where disability is assigned to anyone with a ranking of 'has a lot of difficulty' or 'cannot do at all' in at least one of the six domains of functioning.

## 9.2 Geographical mapping of intervention schools and communities

The geographical mapping of the intervention schools and communities is a critical exercise not only to inform the approach to sampling and cohort tracking for the evaluation, but also for the project to get an understanding of the likely transition challenges. The project should conduct the mapping exercise at the MEL Framework stage and revise it in coordination with the evaluator after it's been recruited.

The geographical mapping should be conducted in four steps:

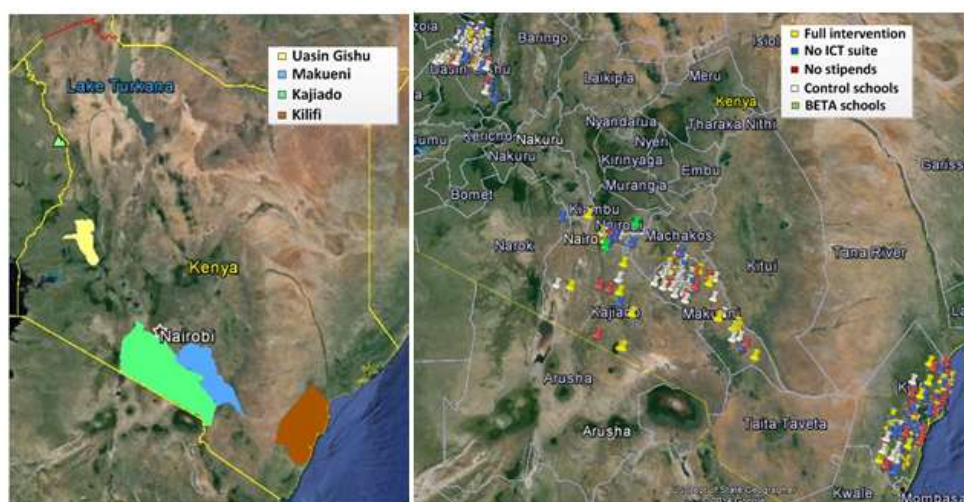
1. **Identifying GEC-1 geographical areas:** the project should list and map all the countries, regions, districts, villages and cities where it has worked in the first phase of GEC, and intends to continue working in through GEC-T.
2. **Identifying GEC-1 schools and clubs:** the project should list and map all the schools (primary and/or secondary) and physical clubs where it has implemented interventions in the first phase of GEC, and intends to continue working in through GEC-T.

3. **Identifying GEC-1 catchment areas and communities:** the project should identify the areas where their beneficiaries are likely to live; these are usually defined as areas within a given radius from the schools and clubs of the project. This distance crucially depends on transport modes available (foot, bike, bus etc.) and by the nature of the location, i.e. rural, urban, suburban, informal settlements etc.
4. **Identifying GEC-T transition areas and transition points:** lastly, the project should define transition routes for beneficiaries in all its communities and predict where they might be at different steps of their route. The project should identify the location of the schools where the beneficiaries can be expected to transit to, regardless of whether these are going to be GEC-T project schools or not. Again this mapping will depend on transport modes.

It should be noted that steps 1 and 2 above include an element of listing the intervention areas and schools. This exercise of listing (including crucial location information) should be completed as part of completing the sampling framework. Further explanation on this framework is given in the following sections.

Figure 3 below is a good example of a map of school locations from the Avanti iMlango project. Projects will be required to share with the Fund Manager the lists of the schools and their location, as well as the map in the highest possible resolution.

**Figure 3 - Map of iMlango counties and schools**



Mapping of project interventions should reveal the transition pathways which different regions and communities may be more predisposed to follow. For example, where target group maps have identified a community that has little access to secondary schools, it is likely that this community will struggle more with transitioning beneficiary girls into those secondary institutions. Projects should reflect on such findings, and consider how they may wish to adapt project activities to be most effective at addressing the various regional problems faced across their areas of operation. The section below gives a worked example of this.

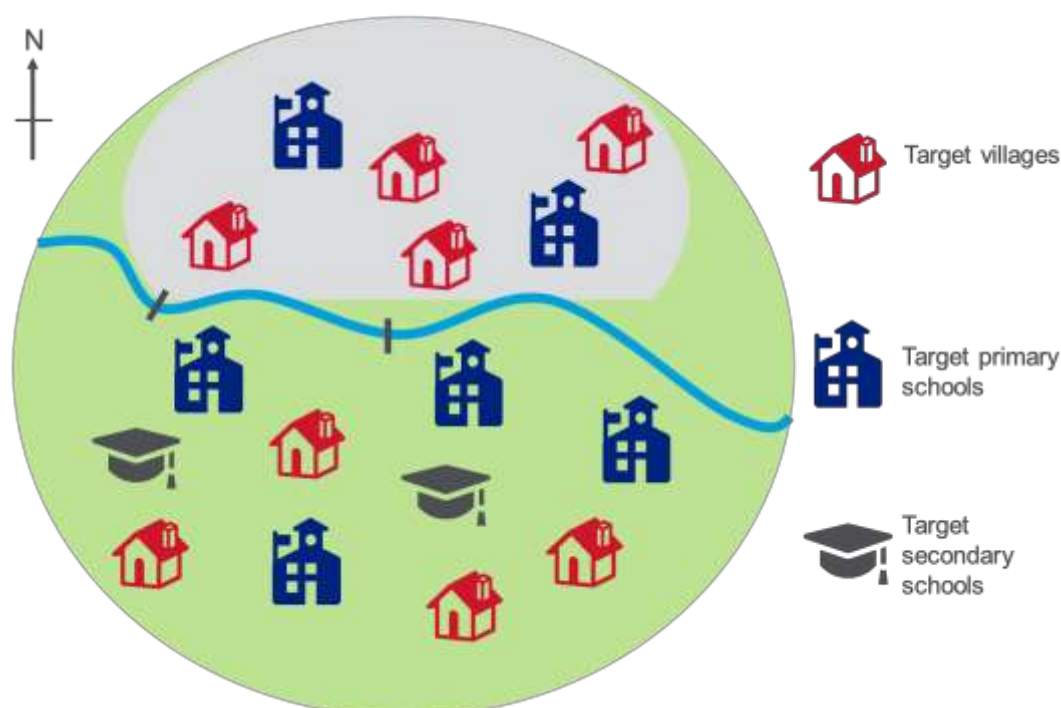
### 9.3 Mapping for transition

The figure below demonstrates the biases that must be considered when sampling villages for measuring transition. The figure represents an example project which is aiming to transition girls from their supported primary schools into, ideally, local secondary schools. They identify eight local communities which they could sample households from, the six local primary schools that they support, and the two local secondary schools which they intend for project beneficiaries to transition to. They map the locations of each of these, and present the findings



to the evaluator, who must develop a strategy to select which communities to select for sampling households from. A pattern quickly emerges:

**Figure 4: Example mapping of target communities**



Each of the identified villages is, more or less, equally likely to have beneficiaries currently attending the intervention primary schools, as they are relatively equidistant from target primary schools. Assuming that the project's activities targeting transition are concentrated in their primary schools, this means that each target village is equal in terms of how likely they are to have girls receiving the transition intervention.

There are only two local secondary schools, however, in the identified region, and they are both located in the South. This means that, in general, it will be easier for those in the Southern community to successfully transition. Those individuals in the North, marked in grey, will have more distance to travel to reach a secondary school, and are therefore less likely to successfully transition at the end of their primary education.

If the evaluator were to select the four Southern villages as their sample, then there would be a significant bias in this approach: girls in these communities are geographically pre-disposed to be more likely to transition. Such a sample risks overstating the project's true achievement, as the more challenging transition pathways of the Northern community is ignored.

A more balanced sample would, for example, select two communities from the Northern communities (in grey) and two from the Southern communities. In this way, the evaluation is likely to be able to say more about how the project has addressed distance from school as a barrier to successful transition.

Projects could take several actions on the back of such a finding:

- The MEL framework should stipulate restrictions on the evaluator's sampling strategy, stating that sampling should be geographically balanced: taking account of Northern and Southern communities.



- Project staff should reconsider their planned activities and theories of change on how they are addressing the different barriers to transition for the girls in these Northern villages. Have activities been organised to help reduce the geographical barriers to transition?
- What plans has the project got in place to deal with the possibility of the two local secondary schools becoming saturated and unable to support the transition of more girls from the target primary schools and communities? Projects should consider how they could react to support the continued transition of girls who, for supply-side reasons beyond their control, have no ability to go on to secondary school.

While this example is somewhat simplified, a similar logic needs explored in the mapping of target groups for each project. On a case by case basis, potential avenues for geographical bias, and potential blind spots in the theory of change, need to be explored and mitigated against.

Crucially, the logic explored above applies equally as well to comparison groups as it applies to treatment groups. Before selecting individual communities to sample from in comparison areas, evaluators should look to provide such an analysis, in order to demonstrate that the comparison groups are not inherently biased.

As a minimum in the initial MEL framework design, however, projects should provide maps of their intended target groups in terms of regions, districts, communities, schools, and other planned intervention locations.

Tools for mapping accurately, which projects and evaluators may wish to explore, include:

- **QGIS:** a free open source software with statistical and analytical capabilities: <http://www.qgis.org/en/site/>
- **Google Maps:** A wide range of software is available, some free and some paid-for. Could be especially useful if road travel is a key factor in determining transition: <https://developers.google.com/maps/documentation/api-picker>
- **MapInfo:** A paid-for software that has a simpler user interface than other mapping software: <http://www.pitneybowes.com/us/location-intelligence/geographic-information-systems/mapinfo-pro.html>
- **ArcGIS:** A paid-for software offered by ESRI, this is a powerful and flexible software. Still a learning curve to pick up, but simpler than the open source QGIS: <http://www.esriuk.com/software/arcgis/arcgisonline>

## 10. Evaluation approach to additionality of outcomes

The previous section introduced the need to identify and closely define target groups, while this section introduces the importance of comparing these groups to a counter-factual.

In the GEC the central principle of the selection of the evaluation approach is that it should be based on measuring the achievement of additional results. This means outcomes achieved for intervention groups that can be shown to be over and above any change that would likely have been experienced in the absence of the intervention<sup>17</sup>. The project's evaluation approach

---

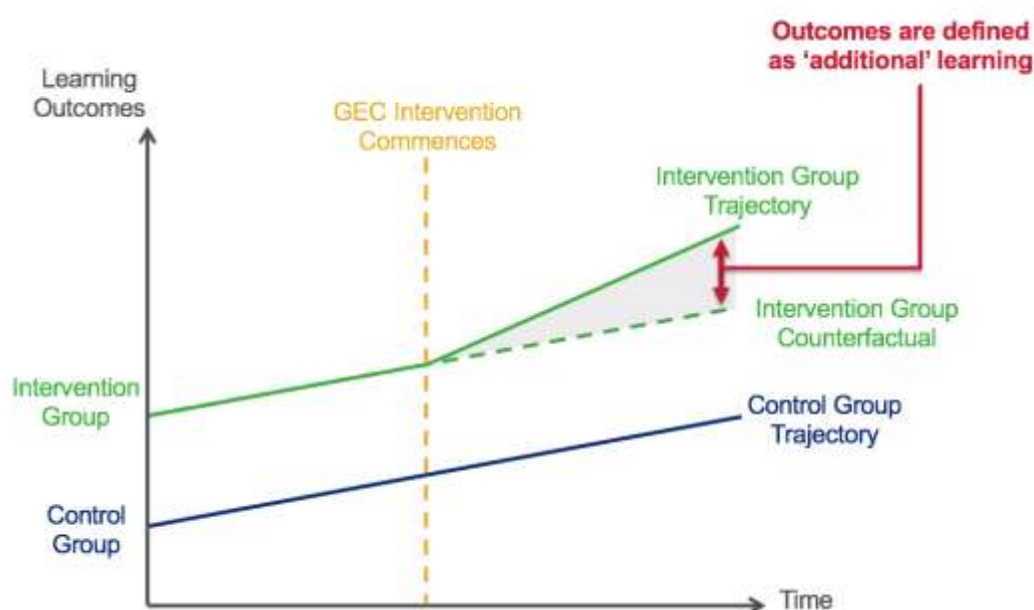
<sup>17</sup> The control or counterfactual group may have interventions from other, non-GEC donors, and the Recipient should not seek to restrict access of other interveners to control groups. Any non-GEC activities in control areas should be reported to the Fund Manager, simply from a results reporting point of view.

describes the *counterfactual scenario* to the intervention, i.e. a state of the world where key conditions are reasonably similar to the intervention scenario except for the intervention.

## 10.1 The Difference-in-Difference approach

**The key approach to estimating additional results in the GEC is Difference-in-Differences.** This approach is founded on attributing to the interventions any change in the outcome experienced by a group of beneficiaries before and after the intervention and over-and-above the change experienced by a comparison group of comparable non-beneficiaries in the same period. The change happening in the comparison group provides the counterfactual scenario to the interventions. Figure 4 shows a diagram of the overtime changes in outcome for the intervention and comparison group and identifies the change that can be attributed to the intervention (or interventions).

**Figure 5 - Additionality in the GEC: the Difference-in-Difference model**



**The standard, unadjusted Difference-in-Difference approach delivers an unbiased estimate of the causal impact of the project under the following assumptions:**

*Assumption 1. The outcome trajectories of the intervention and comparison groups would be parallel in the absence of the intervention.*

This assumption can reasonably be accepted when:

- The intervention and comparison groups share similar socio-economic, demographic, cultural, geographical, and educational characteristics. This allows to assume that, on average, individuals in the two groups would react in the same way to the interventions.
- Prior to the start of the project, the outcomes in the two groups parallel outcome trajectories. Although past performance is not always indicative of the future, common trends suggest the two characteristically comparable groups are affected by similar shocks and external factors.

*Assumption 2. All individuals in the target groups are reached by the same intervention or interventions at the same time and no individual of the comparison group receives any of the interventions.*

This assumption cannot be tested before implementation but it can be verified ex-post through detailed exposure data (as explained in Chapter 9 on Target Groups).

Both assumptions can be rigorously tested with the appropriate evidence in the MEL Framework design stage using secondary data as well as at baseline, midline and endline with primary data.

**DID delivers a robust statistic for the Average Treatment Effect, that is, the average effect of the intervention across all targets.** However, the researchers should be aware that beneficiaries often respond to policies and interventions in different ways. For example, among the beneficiaries of bursaries, there might be some for whom the economic barriers to continue education are still too high and have to drop from school. Another example is when teacher training affects disproportionately the sub-group of the targets who start from low learning levels. While appropriate sub-group analysis can produce insights into different effects of the projects, there can be extreme cases where the distribution of the effects is so skewed that DID leads to conclusions that are not true (more often than it should, based on given statistical power of the samples).

The section below gives an overview of evaluation approaches and related procedures to ensure that there will be the required confidence in the results.

## **10.2 Valid approaches to apply Difference-in-Difference**

Recipients should include **details of their proposed method of demonstrating causality within their MEL framework.** The table below provides a summary of approaches to show additionality in consistency with the DID framework. These approaches represent the minimum level of rigour that is required for projects to estimate their causality for GEC outcomes and for PbR purposes. There may be some exceptions where other methods may need to be used, but it is vital that the methodology for outcome targets and measurement are agreed with the Fund Manager at the MEL framework stage.

It is understood that the context in which each Recipient is operating is different and that it may be challenging to identify a counterfactual or comparison group. However, it is expected that the Recipient should make every effort to employ at least a quasi-experimental approach. The Recipient should discuss any challenges/concerns with the Fund Manager through the submission of the MEL framework and the subsequent dialogue and iterations. Any deviations from the minimum standard shown below must be approved by the Fund Manager prior to commencing the baseline.

**Table 9: Approaches to additionality**

Approach	Description	Acceptability for GEC-T
Randomised control trials (RCTs)	Experimental designs randomly assign subjects to intervention or comparison groups to ensure that there is no selection bias. While RCTs are the 'gold standard' of evaluation designs they are often not practical because random selection of sites and students can be contentious or impossible given political considerations.	Treatment cannot be randomised in GEC-T since projects are to continue with the same beneficiaries of GEC-1. However those projects who adopted a RCT in GEC-1 can continue with this approach and are encouraged to do so.
Quasi-experimental designs	Quasi-experimental designs are based on comparing outcomes of a treatment group and comparison group. Comparison groups should be matched to the greatest extent possible on observable characteristics that were used for project selection criteria. Where comparison groups differ noticeably in terms of their characteristics (observed or unobserved) then the Recipient should demonstrate a reasonable attempt to avoid selection bias.	Acceptable approach to demonstrating additionality for GEC-T.
Pre-post-experimental design	Pre-post experimental designs do not have a comparison group. These typically involve measurements pre and post-intervention and allow us to measure changes in outcomes over time. While statistical methods can be used to demonstrate if changes are significant, we cannot say if this change would have occurred even without the application of the programme. The counterfactual and the targets in these designs are usually based on projections of historic data from the group of beneficiaries or other similar subjects.	Not an acceptable approach to implying additionality for GEC-T. This approach may be employed only in extreme circumstances with agreement by the Fund Manager and would require additional accompanying evidence. <sup>18</sup> This approach would not be eligible for the PbR upside.

### 10.3 Expected challenges to selecting comparison schools and communities

Projects who employed a quasi-experimental design or an RCT during GEC-1 might be interested to use the same schools and communities to sample a comparison group of individuals who are similar to their beneficiaries. However it is important to recognize that there could be major limitations to this approach:

<sup>18</sup> Pre and post-test methods may be considered only when the use of a comparison group violates DFID's policy of "do no harm" – that is, where comparison groups may cause major safety problems for the Recipients or project staff involved. This will need to be agreed with the Fund Manager.

- **Common transition routes:** When GEC-1 control schools and communities are particularly close to the project's schools, then it is likely that girls from both intervention and comparison group will share similar transition routes in the future. In some cases this may mean that girls from the control schools will benefit to some extent from the interventions. Projects who will implement interventions in new Secondary School need to be particularly aware of this potential limitation.
- **Consent to participate:** The GEC-1 control schools and communities might not be willing to participate to the evaluation any longer, especially where they have not received anything in return.
- **Contamination:** The GEC-1 control schools and communities might have been "contaminated" by other donor-funded projects to an extent that they are no longer suitable comparisons for the intervention schools. Where it becomes difficult to identify communities with no educational interventions as comparison groups, evaluators should look to select communities with interventions that may be significantly different from the proposed GEC activities.
- **Saturation:** The project might have saturated control schools and communities by rolling out some interventions to them (following end of the evaluation).

## 11. Quantitative and Qualitative methods

A key principle for GEC-T project reporting is regular independent evaluation of project interventions. These evaluations should employ the use of mixed methods, and ensure rigour in both quantitative and qualitative analysis.

This section motivates the need for mixed methods analysis in GEC-funded evaluations and provides guidance on the minimum standards expected for the quantitative and qualitative research presented in evaluation documents for GEC-T. The standards presented are designed to accompany the terms of reference for evaluators of GEC-T projects. Projects should ensure their evaluators are fully aware of the need to adhere to these standards.

### 11.1 Why mix methods?

The approach of drawing on both qualitative and quantitative evidence is referred to as 'mixed methods'. Taking such an approach can strengthen an evaluation, both through triangulation of findings and by building a deeper understanding of how and why change has/hasn't occurred.

'Triangulation' is when different data sources and methods are used to shed light on an issue or programme. Triangulation can be achieved either by gathering data from different research participants or by examining an issue using different data collection methods. For example, it is possible to compare the perspectives of teachers, students, and parents on the quality of schooling or to gain an understanding of student perspectives through a questionnaire, focus group discussions and participant observations. Triangulation can strengthen evaluation conclusions and help to identify meaningful areas for further work.

Qualitative and quantitative methods can also be used to supplement and build on each other as one research approach can rarely fully address the evaluation questions that are posed. Depending on the evaluation questions being addressed, sequencing of quantitative and qualitative techniques may differ. As a result, an evaluation may, for example, use qualitative research to develop and guide the selection and design of questions in a quantitative survey. Conversely, a statistical analysis of quantitative survey data may identify variances, trends, and patterns, which can then be explained and explored further through subsequent

qualitative data collection and analysis. Identifying an appropriate approach to sequencing can add considerable value to an evaluation and result in findings which are more nuanced and based on a greater depth of analysis. There is no right or wrong sequence. The most important thing is that evaluators choose a strategy that can best answer the evaluation questions of interest.

## 11.2 Quantitative methods

There are several quantitative options available for projects to use. Some must be used to measure certain outcomes as stated in the FM guidance (i.e. learning and transition) and others are optional, but all will require some tailoring to suit the projects particular needs.

### 11.2.1 Learning tests

Learning tests will be administered at the school or intervention level to collect data on literacy and numeracy. Learning tests are compulsory but should be adapted to suit the project, e.g. language in which it is administered and cultural relevance of the tests. See section 3.6 for more details on learning tests. Data from these learning tests will be used to carry out difference-in-difference calculations and provide a measure of additional learning achieved. Other types of learning relevant to projects ToC, e.g. knowledge of sexual and reproductive health can be measured as deemed appropriate by projects. This could take the form of multiple choice questions in a specially designed test and be administered alongside the learning tests (see section on school based surveys below).

### 11.2.2 Household based survey (HHS)

All projects are required to carry out a HHS at baseline. The HHS can also be used to collect relevant information on other project outputs and outcomes.

#### *HHS format*

The HHS will be made up of core and optional modules. Modules refer to sets of questions linked to particular areas, e.g. a module on household demographics or a module on girl's self-esteem. Core modules should be administered by all projects and must be asked of all girls (see section 13.6 for guidance on calculating minimum sample sizes for the HHS). The optional modules should be administered where they are relevant to the activities and outcomes of the project. Optional modules could for example be used to collect data on intermediate outcomes expected at a community or parental level, e.g. attitudinal change towards girls' education. These optional modules can be asked of the entire sample or a subset of the sample (see section **Error! Reference source not found.**). Information collected in this survey could be matched with learning results, depending on the tracking approach taken by evaluators and the level of overlap between the learning and transition sample (this is further explored in section 16).

A final draft of the survey (with core and selected optional modules and revisions) must be submitted to the FM for comment before going into the field. Once the baseline is complete, the HHS will be updated by the FM for use at subsequent evaluation points by adding a recontact/substitution section to allow cohort tracking. This allows for a longitudinal view of changes in outcomes.

#### *Timeline*

A finalised HHS template accompanied by guidance on how to use and administer the survey will be shared with projects by **30<sup>th</sup> June 2017**. The project should adapt the survey as appropriate before translating and piloting. The FM recommends that baseline data collection is complete by November 2017 to allow for analysis and submission of the baseline report and



data by March 2018. Data collection dates should be appropriately aligned with school and exam dates.

### *Training interviewers and field staff*

Projects and their external evaluators should ensure all enumerators are trained in the correct use of the HHS. Enumerators must know how to administer the survey in the field; and for subsequent evaluation points be familiar with the recontact and substitution protocol (see section 9: sampling). Interviewers also need to be aware of cultural sensitivities and ethical issues, for example around singling out a specific girl in the family without interviewing her siblings, or asking sensitive questions about girls, especially when they hit adolescence. In many contexts, it may be advisable to use female interviewers to the extent possible.

### *Pre-baseline survey*

Projects may also wish to design and carry out a pre-baseline survey in order to help identify barriers to education, to identify potential beneficiaries or to inform their sampling framework.

### 11.2.3 School/intervention based survey

Projects also need to administer a short school/intervention based survey to those girls that form the learning cohort and take the learning tests. The core survey should consist of a very small number of questions related to demographics such as age and grade, and questions around marginalisation as defined by the project. These school/interventions based surveys could however also be used to collect intermediate outcome information such as for quality of teaching or self-esteem. Information collected in this survey can be matched with learning results, and therefore provides the potential for interesting analysis of links between learning and other factors. A template with core questions for the School Based Survey will be issued alongside the HHS template.

School-based surveys may also be administered to teachers or head-teachers in order to triangulate information provided by girls or to collect data on additional outputs and outcomes.

## 11.3 Qualitative methods

This section presents minimum standards guidance for qualitative research in six specific areas, which are summarized in **Table** below.

**Table 10: Summary of minimum standards for qualitative research in evaluation**

Area	Minimum standards <i>At a minimum, the evaluation team must outline</i>
Personnel Involved in Qualitative Data Collection and Analysis	Who was involved in the qualitative work presented, and whether those collecting and analysing qualitative data had necessary skills, experience and training
Data Collection Methods	Which data collection methods were selected and why, and who was intended to participate in each of them
Data Collection Tools	The data collection tools used for each of the methods included in the evaluation
Qualitative Sampling Strategy	A comprehensive, and well-articulated sampling strategy for the qualitative work undertaken
Method of Qualitative Analysis	The process for analysing qualitative data, and any associated challenges



Area	Minimum standards <i>At a minimum, the evaluation team must outline</i>
Biases related to qualitative research	The main biases related to the qualitative work undertaken, including those related to data collection, analysis and interpretation

This document proceeds as follows: Section A provides introductory material on the rationale for qualitative research. Section B describes each of the six minimum standards for qualitative research in further detail.

### 11.3.1 Why qualitative research is needed in evaluations

Not only can qualitative research give voice to people who were intended to engage with and/or benefit from an intervention, it can also help explain ‘how’, ‘why’, and ‘under what circumstances’ intended or unintended changes are taking place.

As such, qualitative research can be used in evaluations to obtain information about:

- Local **knowledge** and understanding of a given issue or programme.
- People’s **perceptions** and **experiences** of an issue, or a programme intervention.
- How and why people **engage** with and **react** to a programme, or an organization.
- Local **responses** and the **acceptability** and **feasibility** of a programme.
- **Social processes** and **contextual factors** (for example, social norms, values, behaviours, and cultural practices) that marginalize a group of people or have an impact on programme delivery and effectiveness.

Qualitative research could be used to help exploring why a specific intervention was successful or unsuccessful in improving learning, or how a specific component of programming helped enhance the transition of girls. Qualitative research can also help build narrative links between project outputs, intermediate outcomes, and outcomes by exploring the mechanisms that support these links. So, for instance, qualitative analysis could shed light on how and why having a school improvement plan in place for a particular school improved learning results.

Because qualitative research is usually better suited than quantitative research to investigating sensitive topics, it can also be used in some cases to capture evidence related to issues such as sexual violence or corporal punishment. It could likewise be useful in understanding barriers experienced by hard-to-reach groups, such as children living or working on the street.

### 11.3.2 Minimum standards

This section focuses on the minimum standards that must be fulfilled for the qualitative research conducted as part of GEC-funded evaluations. In some cases, evaluators and projects may feel the need to diverge from the more specific guidance presented below. In general, such divergence must be explained in the methodology section of evaluation reports, which should also highlight the rationale for this divergence. As part of best practice, projects should consider discussing more substantial divergences with the FM at the time of planning and implementation of fieldwork.

All GEC-funded evaluations must also meet child protection and ethical standards. These standards are outlined in the GEC project handbook.

### 11.3.3 Personnel involved in qualitative data collection and analysis

At a minimum, the evaluation team must outline:

*Who was involved in the qualitative work presented, and whether those collecting and analysing qualitative data had the necessary skills, experience and training*

- Evaluation reports should include a summary of the personnel involved in the collection and analysis of all data, highlighting where the evaluation used raw data or processed qualitative data from other sources. **Evaluation reports must note any possible biases that may have occurred** due to improperly trained researchers or non-objective sources of qualitative information.

*Who was involved in the collection and recording of qualitative data*

- In order to lend objectivity to the evaluation, we encourage evaluators to only use qualitative data collected by themselves or a firm who they have partnered with.
- In some cases, the evaluator may also need to use qualitative data collected by project staff. Where this is the case the evaluator must make it explicit in all reporting. The evaluator must also attempt to validate this qualitative data in some way, and should highlight these attempts in their reports.
- It is imperative that all reports explicitly state who collected each type of qualitative data used and the training which was provided to them.
- Evaluation reports should also note any challenges encountered which may have undermined the independence and reliability of the qualitative data.

*Who was involved in the analysis of qualitative data*

- In order to lend objectivity to the evaluation we encourage evaluators to analyse qualitative data themselves or for this to be done by a firm who they have partnered with.
- In some cases, this might not be possible. Evaluators using qualitative data analysis from another source must attempt to validate the analysis in some way, and should highlight these attempts in evaluation reports.
- In some cases, qualitative analysis will be done by researchers hired by the evaluator, and then compiled at a central level. In others, one or two qualitative specialists will analyse all of the data. In others still, evaluators may use qualitative data analysed and processed by another party. In any case evaluators should be explicit about who undertook the analysis and any implications for the reliability of the findings.
- Evaluation reports must also outline the skills and experience of those involved in analysing the qualitative data and the training which was given to them.
- In many cases, translation of raw qualitative data will be required. The evaluator should explain who conducted these translations and any checks that were in place to ensure the accuracy of them.

### 11.3.4 Qualitative data collection methods

At a minimum, the evaluation team must outline:

*Which data collection methods they have selected and why, and who is intended to participate in each of them.*

- This should include a clear rationale for each data collection method used, the groups or individuals who are intended to participate in each of the methods, the type of data they are intended to generate and the sensitivity of the topics to be covered. As well as an explanation of why each individual method was selected, the value of the combination of methods should also be explained, as well as the rationale for any sequencing of methods.

#### *Rationale for selected data collection methods*

- Evaluators should explain where they are using tried and tested methods which are being used in their pure form and where they are drawing principles from a range of methods in order to develop a new method tailored to the evaluation.
- It is good practice to use a combination of methods within an evaluation and evaluators should explain the value not just of individual methods but of the combination of methods used and any linkages between them. This may include sequencing of methods, for example to support triangulation.
- Methods may range from FGDs and KIs to more innovative participatory methods, including peer research and methods which use technology, for example video and photography. Direct observations are also likely to provide valuable data, particularly in classrooms. The use of participatory methods is particularly encouraged because of the valuable data they can yield and the empowering effects they can have on participants.
- As noted in Chapter 16 guidance, ethical considerations should inform the selection of data collection methods and these should be reflected in rationales provided.

#### *Biases arising from certain data collection methods*

- Biases can result from certain methods, for example social desirability bias with methods that take place within groups. Potential bias should be noted for each of the data collection methods included in the evaluation.
- Evaluators should demonstrate that efforts have been made to minimise these biases and strengthen the quality of the data collected.



#### **Further reading (optional)**

- **Research methods** - <http://www.ccs.neu.edu/course/is4800sp12/resources/qualmethods.pdf>
- **Participatory methods** - [http://archive.unu.edu/hq/library/Collection/PDF\\_files/CRIS/PMT.pdf](http://archive.unu.edu/hq/library/Collection/PDF_files/CRIS/PMT.pdf)
- **Most Significant Change** - <http://www.mande.co.uk/docs/MSCGuide.pdf>

### **11.3.5 Qualitative Data Collection Tools**

At a minimum, the evaluation team must provide:

- The data collection tools used for each of the methods included in the evaluation.

#### *Development of qualitative tools*

- These tools should be accessible and appropriate to the groups and/or individuals who are intended to use them.

- The evaluator must share all qualitative data collection tools with the FM and should explain whether any subsequent feedback or advice from the FM was taken on board.
- Any sensitive topics should be appropriately worded and scripts for field researchers should be included to carefully introduce sections on sensitive topics and to forewarn participants about the nature of the questions being asked. Prompts should be included to give respondents the opportunity to stop or pause the data collection activity if they want to.
- All data collection tools should be field-tested and any findings from this should be incorporated in the final version of the tools
- Tools should be developed using appropriate and accessible language. In many cases, data collection tools will need to be translated from English into at least one other local language. Poor quality translations can undermine the value of the data collected. In order to check the accuracy of the translations, it is advisable to conduct 'back translation' with a sample of data collection tools in order to validate their quality.
- Where possible, tools should be developed based on examples of good practice from other evaluations, incorporating lessons learned from elsewhere. Efforts to do this should be clearly outlined by evaluators.
- Tools should be developed in a way that prioritises the data that is needed and minimises respondent and field researcher fatigue.

### 11.3.6 Qualitative sampling strategy

At a minimum, the evaluation team must outline:

#### *A comprehensive and well-articulated sampling strategy for the qualitative research*

- Evaluators should describe who participated in the data collection for each qualitative method and explain how participants were selected. This should include information on sample sizes and the representativeness of samples selected, as well as any potential biases as a result of sampling approaches used.

#### *Sample size*

- A rule of thumb for qualitative methods is to stop collecting data when saturation point is reached, meaning nothing new is being raised by participants. However, it is not always possible to manage data collection in this way and sample sizes often have to be defined in advance. Previous experience should guide decisions about when saturation point is likely as a way of determining the sample sizes to be used. Practical considerations such as cost and time also need to be factored in.
- Evaluators must clearly delineate sample sizes for each method and provide a rationale for why these sample sizes were adequate and appropriate.
- Where focus group discussions are being used, these should include on average between five to 10 participants.

#### *Method of participant selection*

- Evaluators should highlight the sampling techniques employed for each type of method and explain why certain sampling techniques were chosen, e.g. probabilistic sampling vs. purposive, quota or snowballing.

### **Representativeness of sample/s**

- The validity, meaningfulness and insights generated from qualitative work are related to the sampling approach used. Evaluators should outline the key characteristics of samples. For instance, in a focus group with school council members, the evaluation team should note how many were women. In semi-structured interviews with parents for instance, evaluators could highlight how many parents were educated themselves.
- In qualitative research, the aim is not to be representative of the population. However, if evaluators are excluding certain types of beneficiaries, or focusing on those with certain characteristics, they should note any implications of these choices. For instance, if a project has beneficiaries who are disabled but an evaluation team does not interview any disabled girls, then this should be acknowledged and explained.

### **Biases as a result of sampling**

- A biased sample consists of respondents who don't represent the group of interest. Poor screening and recruiting causes biased samples. It is important that evaluation teams carry out appropriate screening.
- Gatekeeper bias arises when individuals or groups of stakeholders take/are given control of sampling and decide who should participate in the evaluation without using a pre-agreed sampling framework.

Evaluators should outline any biases they feel arise out of the sampling strategy or implementation method.



#### **Further reading (optional)**

- **Sampling** - [http://www.ajmse.leena-luna.co.jp/AJMSEPdfs/Vol.2\(2\)/AJMSE2013\(2.2-21\).pdf](http://www.ajmse.leena-luna.co.jp/AJMSEPdfs/Vol.2(2)/AJMSE2013(2.2-21).pdf)
- **Sampling** - <http://legacy.oise.utoronto.ca/research/field-centres/ross/ctl1014/Patton1990.pdf>
- **Focus groups** – <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.468.3527&rep=rep1&type=pdf>

### **11.3.7 Method of qualitative analysis**

At a minimum, the evaluation team must outline:

#### **The process for analysing qualitative data, and any associated challenges or biases**

- Evaluators should describe their overall approach to qualitative data analysis as well as specific techniques used, highlighting any software employed. Any sequencing of the analysis to inform further quantitative or qualitative data collection should be made clear.
- Evaluators should explain how, if at all, interpretation and findings were cross-checked and validated. They must also note any possible biases in the methods of analysis or any challenges faced during the analysis process.

### ***Process for analysing qualitative data***

- In general, qualitative data will need to be categorized and sorted through pattern or thematic analysis as the primary basis for organizing and reporting the evaluation findings. Evaluator should be clear about how themes were identified.
- Evaluators must also describe how the raw qualitative data was handled and explain whether this was done manually or using software such as Nvivo.
- Any cross-validation techniques employed to ensure the final analysis is accurate should be highlighted.

### ***Sequencing of the quantitative and qualitative data collection***

- As noted in section B2, sequencing of data collection and analysis can strengthen an evaluation and the value of the findings. Either quantitative or qualitative data collection and analysis can come first. In general, the two should not be conducted simultaneously as one should be used to inform the other so that initial findings can be explored in greater depth.
- Evaluators should be specific about the sequencing used in the evaluation, and explain why the sequencing approach employed was appropriate in the context.

### ***Biases related to analysis***

- Bias related to analysis occurs when the views presented are not objective, and do not present a reasonable representation and interpretation of the qualitative data.
- Wherever appropriate, evaluators should include alternative explanations and divergent views in order to build nuance in to the qualitative analysis.

### **11.3.8 Biases related to qualitative research**

At a minimum, the evaluation team must outline:

#### ***The main biases related to the qualitative research undertaken, including those biases related to data collection, analysis and interpretation.***

Biases could relate to:

- Interviewer/researcher bias: particular attention should be paid to interviewer bias, confirmation bias, and cultural biases.
- Inadequate training: Bias can occur if personnel involved did not know how to correctly collect or analyse the data, for example as a result of inadequate training.
- Social desirability bias: This is particularly common in group settings. Questions should be phrased so it is clear that it is acceptable to answer in a way that is not socially expected or desirable and which runs counter to prevailing social norms. Evaluators should also consider including indirect questions which ask how other people might think or behave.
- Leading questions: Evaluators should ensure tools are designed in a way which minimizes the risk of leading-questions bias and question-order bias.
- Sampling bias: A biased sample consists of respondents who don't represent the group of interest. Poor screening and recruiting causes biased samples. It is important that evaluation teams carry out appropriate screening.

- Gatekeeper bias arises when individuals or groups of stakeholders take/are given control of sampling and decide who should participate in the evaluation without using a pre-agreed sampling framework.
- Bias in analysis: Bias related to analysis occurs when the views presented are not objective, and do not present a reasonable representation and interpretation of the qualitative data.

### **Documenting biases**

- Most research will be subject to some form of bias. It is critical that evaluators document possible biases in order to help readers interpret the findings appropriately

### **How to increase credibility of analysis**

- Data triangulation is the most commonly used method of establishing trustworthiness in qualitative research. Evaluators are encouraged to triangulate qualitative data obtained from different sources, as well as triangulating qualitative and quantitative data in order to enhance the credibility of the evaluation
- There are many other ways of establishing trustworthiness and confidence in qualitative research, including: member check, interviewer corroboration, peer debriefing, prolonged engagement, negative case analysis, auditability, confirmability, bracketing, and balance. Evaluators are encouraged to employ such techniques as and when appropriate to improve the credibility of the evaluation findings



### **Further reading (optional)**

- Kielmann, K, Cataldo, F, Selley, J, 2001, Introduction to Qualitative Research Methodology, DFID and Evidence for Action  
[https://assets.publishing.service.gov.uk/media/57a09ddfe5274a31e0001ac6/qualitative\\_researchmethodologymanual.pdf](https://assets.publishing.service.gov.uk/media/57a09ddfe5274a31e0001ac6/qualitative_researchmethodologymanual.pdf)



## 12. Evaluation ethics and child protection

Child participation in evaluation activities is critically important in order to understand the impact of a project. However, children's participation need to be approached with the utmost care to ensure that inclusion is meaningful and that children are not put at risk through their participation. While safeguarding children that organisations come into contact with through project activities are widely recognised as important, this is not always thought about within the context of evaluation activities. The section of the guidance aims to help GEC projects think about child protection and the risks to children through their research and evaluation activities.

### 12.1 Child protection policy requirements for external evaluators

The FM places high importance on the safety and wellbeing of all children that come into contact with any GEC funded activity, including any GEC commissioned evaluation activity. The FM requires all GEC funded projects to ensure that all evaluation activities are conducted in the best interest of the children involved and ensure that they are safeguarded through different evaluation activities, including data collection, data analysis, report writing and dissemination. Ethical issues associated with these are discussed more in detail in the next section.

Data collection, in particular, needs to be carefully considered due to the close contact between the researchers and the children being consulted. The inherent power difference between the researcher and children and the often private nature of the interview must be understood in order to ensure children's participation is safe. It is therefore crucial that **adequate safeguards are put in place to minimise the risk that researchers behave inappropriately or insensitively.**

*A CPP is one indication that an organisation is mindful of the special needs of children in research and that they recognise that international best practice dictates that organisations that come into contact with children should have CPPs in place that safeguard those children.*

*However, it is not a magic bullet and GEC grantees are accountable for ensuring robust safeguards are in place to protect children who take part in evaluation activities.*

#### **What child specific safeguards does the FM expect to be in place?**

- That the external evaluator has a CPP in place and that all research activities, whether sub-contracted or not, adhere to the safeguards outlined below:
- That recruitment of all research team members are guided by safe recruitment practices. All GEC grantees need to be mindful that adults with abusive intentions may use research as a means of gaining access to children.
- **A comprehensive code of conduct** that outlines how to protect children from inappropriate behaviour perpetrated by staff/contractors is in place for all in the research team to follow.
- **All members of the research team have had CP training** to understand how CP features in different evaluation aspects including developing tools and research methods, informed consent, code of conduct, incident reporting mechanisms, data protection etc.
- **A reporting and response mechanism is in place** to respond to children in distress or to a researcher breach of the code of conduct.

As stated in the GEC-T minimum standards for child protection (see section 2.1, 7.11, 7.12 and annex B in the handbook), all GEC-T grantees are expected to ensure child protection standards are upheld across all project partners and contractors. This includes external evaluators. The FM encourages all GEC grantees to take this into consideration when drafting

TORs, advertising and selecting their external evaluator. However, it is important to note that **a Child Protection Policy (CPP) is not a magic bullet and GEC grantees still have the overall responsibility to ensure that all evaluation activities, including through evaluation sub-contractors, are conducted with the best interest of the child and that CP protocols are firmly in place.**

**Additional safeguards are needed when researching violence, sexual and reproductive health and rights and other sensitive topics**

Asking children about violence can generate much needed data to inform programming efforts. However, if the research is done poorly, there is a real risk of harm. Children may become distressed or re-traumatised, their safety compromised and they can be put at risk of further violence.

GEC projects that are considering researching these topics or measuring their impact on these, should be aware of the **need for higher ethical standards in order to keep children safe.** In particular the FM encourages projects to place a heavier emphasis on:

- Ensuring that the **external evaluator who is selected has the appropriate expertise to research these sensitive topics with children.**
- **Ensuring enumerators are recruited with the correct skill set and appropriate safety checks.**
- Limiting data collection on these topics only to what the programme can realistically make use of and **avoid overburdening children.**
- Ensuring that **questions are framed sensitively and are age appropriate** to minimise distress to children.
- Ensuring that **protocols are in place to refer children to needed support or services.** Where a project assesses that this may be difficult due to lack of services, the project should carefully review whether the information is in fact needed.

## 12.2 Ethical standards in the evaluation process

Ethics is a crucial consideration when conducting any research or evaluation activity. However, ethical consideration are often not straightforward and vary depending on different settings and cultural differences. Several standards have been developed to help researchers

““Ethics” refer to choice making around “right” and “wrong” values and behaviour in evaluation and research. They guide us in our behaviours and relationships – from commissioning through to design and data archiving. They are subject to differing interpretations and complex judgements that are context specific. Ethics are one distinct piece of what constitutes good research and evaluation. Not following principles of ethics can harm evaluation and research outcomes.”

Source: Groves, L. (2016), ‘Review of Ethics Principles and Review and Guidance in Evaluation and Research’, available at:

[https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/524635/Ethics-principles-report-2016.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/524635/Ethics-principles-report-2016.pdf)

and evaluators carefully think through ethical concerns. For example DFID has 10 ethical principles that they expect all their research and evaluation work to adhere to (DFID, 2011). Other international actors have gone further by developing ethical standards specifically looking at the ethical issues of conducting research with children (Graham et al, 2013).

While ethical research can have different meanings in different settings, the FM expects that all GEC commissioned research should be guided by international best practice and seek to uphold the highest ethical standards at all times. In order to do so the FM expects all projects to ensure that ethical considerations are taken into account in the planning stage of research, itemised in the project budget and factored into the timeline, in order for ethical principles and practices to be implemented throughout the research. This guidance aims to help projects think through some of the ethical implications related to both their quantitative and qualitative data collections as well as report writing and dissemination.

GEC projects should be able to demonstrate how they have considered ethical protocols across the following:

### **In the planning and data collection stages:**

#### **Projects should carefully consider the scope and quality of collected data.**

- Projects should carefully consider what data they need to collect and why they need this. It is important to acknowledge the opportunity cost for participants to take part in the research (such as potential impact on income generating activities, homework, childcare, delivery of education).
- If the project decides that sensitive subjects such as experiences of violence and/or SRHR will be included, additional safeguards need to be considered throughout the research process (see section on child protection for more information).
- Projects should also carefully consider whether the data they are planning on collecting or the approach to data collection may introduce bias, such as how marginalisation factors may be picked up or excluded in the evaluation process.

#### **Projects should carefully consider what skills and experience the external evaluator need in order to conduct the evaluation to a high standard and ensure it is safe for participants take part in.**

- Projects should clearly communicate the need for external evaluators to be able to conduct the evaluation activities to the highest ethical and child protection standards.
- Projects should ensure that the correct expertise is recruited for the evaluation, including gender, child protection and inclusive education as needed.
- Projects should ensure external evaluators have processes in place to select enumerators and researchers based on:
  - Safe recruitment practices ensuring that none of them have a criminal record involving child abuse or other inappropriate behaviour.
  - Previous experience of collecting data with children, including with vulnerable children where appropriate.
  - Their experience of conducting research on the topics that will be explored in the evaluation.

- They have the appropriate socio-demographic profile where this is appropriate (such as ethnic belonging, gender and age where either the context or the research topics make these more pertinent).

Projects and external evaluators should develop an approach to research ethics and ensure they have considered risks across the different research activities, including both quantitative and qualitative research methods.

- The evaluator should identify a named individual who has overall responsibility for ethics within the evaluation team.
- A comprehensive risk register should be developed with appropriate mitigating actions identified. Risks could include:
  - What potential risks are there for participants who are involved in the GEC evaluation? Are there any potential physical, psychological or disclosure dangers that can be anticipated?
  - What procedures have been established for the protection of participants and the oversight of any information gained from them or about them?
  - Have particularly marginalised groups been identified to take part in the research for example women (in some contexts), children or adults with impairments, children living or working on the streets or married girls etc? If so have appropriate safeguards and support been put in place to ensure they can meaningfully participate.
  - What are the safety risks for enumerators and researchers?

Projects and external evaluator should ensure they adhere to national and local ethics approval processes.

- Projects are responsible for identifying the need for and securing any necessary ethics approval that may arise out of their research, including from national or local ethics committees as appropriate.
- Projects need to demonstrate that they have a good understanding of pertinent ethical issues are from local perspectives, e.g. particularly sensitive topics, vulnerable participants, taboo issues, and how these topics are understood. Project staff and external evaluators should then take these into account throughout design and implementation of evaluation activities.

External evaluators should ensure survey instruments and data collection methods are age, gender and culturally appropriate and tested.

- External evaluators should ensure the survey instruments adhere to best practice and are sufficiently tested before being used to ensure they are age and culturally appropriate.

External evaluators should ensure all researchers are appropriately trained before they begin data collection

- All personnel involved in data collection and analysis have received training in all of the evaluation's child protection policy and ethical approach and protocols.
- All personnel involved in data collection have been trained to detect signs of distress or trauma and to pause or stop data collection activities as appropriate. All researchers must also be trained on what further protocols must be followed if such a situation would arise.

## **Data collection**

### **Selection of methods and locations for data collection**

- The external evaluator should explain how an approach to research ethics has informed the selection of methods used for data collection and analysis.
- The evaluator must also demonstrate that careful thought has been given to the location and set up of data collection approaches, including considerations related to privacy and who is present.

### **External evaluators should ensure that the safety of participants is paramount at all times**

- External evaluators should ensure comprehensive child protection protocols are in place throughout the data collection period to safeguard participants from inappropriate or harmful behaviours perpetrated by researchers and to be able to respond to new threats or concerns raised by children.
- The safety of the research participants must be paramount at all times and risks to their participation and physical, social and emotional wellbeing should well identified and mitigated against throughout data collection.
- Researchers should be trained on how to report and respond to different situations that may arise (see CP section above) as well as understanding when and how to end an interview appropriately when a child is distressed.
- Researchers should select a safe and private location for their interview with children, to ensure the interview is conducted confidentially, while minimising any time they spend alone with children.

### **External evaluators should ensure critical incident protocols are in place before data is being collected**

- External evaluators and projects should ensure they have robust processes in place to support children who become distressed during the data collection.
- Critical incident protocols should also be in place to rapidly respond to any safety issues affecting the research.

### **External evaluators should ensure they Do No Harm**

- Researchers should try to make sure that they do not inadvertently put research participants at risk or reinforce unequal power structures or discrimination against some groups.

### **External evaluators and projects should ensure the inclusion of stakeholders is meaningful and not based on discrimination**

- External evaluators should ensure that children with any special circumstances (for example married girls or children without caregivers) are identified and adequately supported to participate in the research where appropriate.
- External evaluators should ensure their methods and questions are designed in such a way to encourage and promote the participation of marginalised groups or individuals in communities to take part in the research.

- Researchers should be aware of and able to mitigate the stigma and discrimination experienced by marginalised children, such as those with disabilities.
- All research participants should be able to participate fully in the research, meaning that the location chosen must be fully accessible, the data collection methods used should be adapted where necessary and researchers should be trained and sensitised in regard to including children with disabilities, for example.
- Researchers should regularly review whether children might be inadvertently excluded from a participatory activity because of a physical disability or learning difficulty and most importantly take steps to prevent this from happening.

**External evaluators should ensure participants' participation rests on informed, voluntary and ongoing consent/assent:**

- The evaluator must **define a meaningful process for gaining informed, voluntary and renegotiable consent from adults and assent (agreement to take part) from children under the age of 18.** Consent must not be sought from children but assent must be sought before any research activities take place.
- In most cases parental/caregiver consent should also be sought. However, the FM recognises that in some cases this may not be appropriate and it is therefore not required for all data collection. Where parents/guardians are asked to consent, and children have declined to take part, the rights of the children should be respected.
- The evaluator needs to take an informed view about whether consent from adults should be written or verbal.
- Care must be taken not to put potential participants under any pressure to give consent or assent to take part in evaluation activities. Evaluators must ensure that participants feel they can say 'no' at any point in the process.
- Additional consent/agreement should be sought for the use of voice recorders, video equipment or cameras in accordance with a project's CPP.
- Evaluators must ensure that sufficient information is provided to potential participants. This information provided to potential participants must be appropriate and accessible, including to accessible to children and those with impairments. At a minimum the information provided should include:
  - The purpose of the evaluation
  - The funder of the evaluation
  - Contact information for the evaluation team
  - Why the individual has been selected for participation
  - What participation in the evaluation will entail
  - Any risks or benefits of participating in the evaluation
  - Provisions for privacy, confidentiality and anonymity and any limitations
  - Future use of information given
  - Right not to participate and to withdraw at any point

**External evaluators should ensure research participants can take part anonymously**

- The limits to confidentiality should be defined and explained to participants (e.g. where there is a safeguarding concern).

## Data analysis, storage and report writing

Projects and external evaluators should ensure confidentiality of participants' data at all times and ensure strict data protection protocols are in place

- Projects should be able to demonstrate administrative, technical and physical safeguards to protect the confidentiality of beneficiary data. For instance, when longitudinal sampling or studies are carried out, it is essential that personal information is separated from the panel participants' data.
- Projects and external evaluators should also be able to articulate how physical and electronic data is stored and disposed of to maintain the privacy and confidentiality of all project participants.
- Projects should ensure the data shared with the FM is anonymised at all times.

Projects should ensure they are aware and reflect of biases and limitations of the research findings

- Any limitations or biases should be outlined in evaluation reports and the research findings should be located within these, with any necessary caveats noted.

## Dissemination

Projects should ensure to the greatest extent possible research participants and communities are informed about the evaluation findings.

- Projects should consider how findings can be made available, including to children themselves, illiterate individuals and community members with impairments.

### Useful resources

Bennouna, C., Mansourian, H. and Stark, L., (2017), 'Ethical considerations for children's participation in data collection activities during humanitarian emergencies: A Delphi review', Conflict and Health 201711:5, available at: <https://conflictandhealth.biomedcentral.com/articles/10.1186/s13031-017-0108-y>

Devries, K, Dipak Naker, Adrienne Monteath-van Dok, Claire Milligan and Alice Shirley, (2016), "Collecting data on violence against children and young people: need for a universal standard", available at: <http://inthehealth.oxfordjournals.org/content/8/3/159.full.pdf+html>

DFID, (2011), 'DFID ethics principles for research and evaluation', available at: [https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/67483/dfid-ethics-prcpls-rsrch-eval.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/67483/dfid-ethics-prcpls-rsrch-eval.pdf)

Graham, A., Powell, M., Taylor, N., Anderson, D. & Fitzgerald, R. (2013). Ethical Research Involving Children. Florence: UNICEF Office of Research – Innocenti, available at: <http://childethics.com/>



## 13. Sampling for outcomes in GEC-T

For most projects on GEC-T, it will not be feasible to collect outcome level data on every single beneficiary in the population. This section describes in detail the general approach to sampling being suggested for GEC-T, and is written on the assumption that evaluations are following a standard difference in differences approach as described in the previous section. For projects making exceptions to this approach, the logic of sampling will remain the same, however, some of the application is likely to change given the unique circumstances.

The major sampling technique for most GEC-T evaluations will be multistage, first consisting of cluster sampling, to select relevant communities and schools to identify individuals within, followed by stratified, systematic, or simple random sampling to draw out individuals from within those clusters. This multistage process is described in much greater detail below.

The preferred sampling approach for GEC-T evaluations would be:

- Learning, transition, and intermediate outcomes are all evaluated using the exact same sample of individuals.
- Individuals are first sampled from school clusters, and have learning tests and the school survey modules administered to them.
- Information is collected on each of these individuals in order to follow up with the exact same person at their household.
- Follow up surveys are carried out at the household of each of these individuals from the learning sample and the members of their household (caregivers, siblings, etc.)
- The school and household-collected information is linked together for analysis, and the same individuals are tracked and recontacted for all future evaluation points.

This approach is the preferred method, as it allows for a deeper exploration of the causal relationships between learning, transition, and the intermediate outcomes. Also, in many ways it simplifies the sampling approach for transition. Further details on the sample development, and principles for linking the samples are given in sections 14, 15, and 16 below.

There are practical difficulties with this preferred approach, particularly regarding locating and interviewing the same individuals at both school and household. Sections 14 and 15 give advice, therefore, on how a two-sample solution can be developed: with one in-school sample for learning, and a separate household-based sample for transition.

Projects should make a clear case in their MEL Framework as to whether they intend to follow a linked-sample approach, or if they wish to follow a two-sample approach, and evaluators should likewise be prepared to reflect on the feasibility of the proposed approach through their research development.

Before going into detail on the approaches for sampling for learning, transition, and intermediate outcomes, it is worth detailing general principles that all evaluations, regardless of the linked-sample / two-sample choice, should follow.

### 13.1 GEC-T sampling principles and responsibilities

In order to ensure the credibility of project evaluations, recipients are required to track a cohort of marginalised girls to form a sample that is:

- (i) Representative of the population of beneficiaries

- (ii) Large enough to be able to detect statistically significant improvements in learning, transition, and intermediate outcomes for beneficiaries
- (iii) Diverse enough to be able to capture information on a variety of different transition pathways
- (iv) Designed to minimise sampling error in the treatment group, and minimise contamination in treatment and comparison groups
- (v) Suitably structured to be tracked and re-contacted for subsequent evaluation points
- (vi) Unbiased in its propensity to achieve project outcomes

Theoretically, the more data evaluators are able to collect, the more likely that the criteria above will be met. However, it will not likely be logistically or financially feasible to track all individual beneficiaries recipients intend to engage with. Hence, projects should select a subset of beneficiaries to form the sample for cohort tracking, which is a more pragmatic approach to deliver the goals of project evaluation.

In order to create a sample, however, planning and research needs to go into the creation of a sampling framework.

### *Definition 3: Sampling Framework*



**A sampling framework** is a tool used to identify the total population of regions, communities, and schools that could be targeted for data collection in both treatment and comparison areas.

The responsibility for creating sampling frameworks is mixed. While projects will be expected to create accurate frameworks and lists for their target group populations, it will be the responsibility of the evaluators to select the unique areas, households, and individuals that comprise the treatment sample.

For comparison groups, projects should complete a high level sampling framework for regions, communities and, if possible, schools. It will be the role of the evaluator to verify if these suggestions are appropriate (in line with the guidance on representative sampling below), suggest others if necessary, and ultimately to apply a strategy to identify which areas will compile the comparison sample.

The following section on representative sampling lays out some of the questions that projects and evaluators should be asking themselves in the development of these sampling frameworks. **Error! Reference source not found.** of this guidance provides a template for such a framework, and could be used to help guide the preparation for each project evaluation.

The basic principle of using a sampling framework is that the framework should be populated with a long list, the proposed treatment and comparison *populations*, along with key demographic and stratification variables that should be balanced in the final sample composition. From these population lists, a sampling technique should be applied in order to identify which communities or schools from the total framework should be included as clusters from which to draw the final sample.

This sample technique effectively becomes a filtering exercise by which evaluators narrow down from a full list of potential sampling units (the sampling framework), to a more manageable set of clusters (schools and communities) from which households and individuals

can be identified and interviewed. It is the remit of the external evaluators to apply a sampling technique to the project-prepared sampling frameworks to find this smaller list.

#### *Definition 4 : Sampling technique*



**A sampling technique** is a method to select individual units from the population to form a part of the treatment or comparison sample.

Appropriate sampling techniques to narrow down the sampling frameworks into a list of cluster schools and communities include:

- Simple random sampling
- Stratified random sampling

In both cases, the randomisation is key. Stratified sampling may be preferred in the case, for example, where it is crucial for both urban and rural communities to have representation. A sample consisting of only urban community clusters may under-represent the intervention activities designed to help those in rural communities to overcome their particular barriers to transition. A stratified approach would ensure that such occurrences do not happen, as a number of clusters would be selected randomly from the urban strata and then from the rural strata. The weighting of urban to rural clusters need not be equal in this example, and it is up to evaluators to define the weighting in a stratified sampling technique if it is chosen.

#### *Definition 5 : Sampling cluster*



**A sampling cluster** is a subset of a population from which smaller units can be drawn through a second sampling technique, and which can be used as geographical hubs for research.

Through random selection, the selected clusters should be demographically comparable to those not selected. Evaluators should check that this comparability holds after selecting the clusters. If an underlying pattern appears to have emerged, then re-sampling, or considering stratifying, may be appropriate remedies.

Once a list of treatment and comparison sampled clusters has been decided, evaluators should provide details of a further sampling technique for selecting individuals from within these clusters. This 'secondary' sampling technique is most likely to be some form of stratified systematic sampling technique based on the demographic criteria identified for each project's target group.

Further details on these 'secondary' sampling techniques will be explored in sections 14 and 15 on learning and transition sampling techniques.

### **13.2 Representative sampling**

Representative sampling on the GEC-T should be understood as the principle by which treatment samples are truly reflective of the breadth of project target groups. It should also be understood as the principle by which the comparison sample should match, as closely as possible, to the treatment sample.

The table below lays out in more detail some of the specific questions that projects and evaluators should ask themselves when considering sampling. These questions are made specific to the different strata to be sampled from, and the communities and schools questions are reflected in the sampling framework template in Annex D

The key principles of representative sampling that projects and evaluators should follow are described in the table below.

**Table 11: Principles of representative sampling**

Treatment			
Regions	Communities	Schools	Individuals
<ul style="list-style-type: none"> <li>Where the number of regions being targeted is small (roughly 3-5), a representative sample would take individuals from all regions.</li> <li>Treatment regions should be already pre-defined by the project's target groups.</li> </ul>	<ul style="list-style-type: none"> <li>This community-based sample will primarily be used for setting a benchmark for transition, and so it is important that these sampled communities are unbiased in terms of: <ul style="list-style-type: none"> <li>Density of primary/secondary schools within community</li> <li>Average income</li> <li>Language and ethnic diversity</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>The number of institutions sampled should be calculated using the minimum sample size methodology explained in section 13.6.</li> <li>Care should be taken to avoid bias in terms of, for example: <ul style="list-style-type: none"> <li>Size of the schools</li> <li>Primary/secondary split</li> <li>Government/comm unit ownership and management</li> <li>Average educational attainment</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>While the sampling framework is not expected to provide direct information on the population of individual beneficiaries (this would likely be too much detail), for the purposes of planning evaluation fieldwork, clear criteria for which individuals should be included in the sample should be described: <ul style="list-style-type: none"> <li>Is the child in the appropriate grade or age ranges for the intervention?</li> <li>Are they enrolled in project schools or other project interventions?</li> <li>Does the individual or household fulfil the specific eligibility criteria for the intervention?</li> </ul> </li> <li>These characteristics should be measured and tracked at each subsequent evaluation point.</li> </ul>
Comparison			
Regions	Communities	Schools	Individuals
<ul style="list-style-type: none"> <li>Comparison region should share similar demographic characteristics to the treatment region(s). Appropriate control region(s) should match with the treatment regions in terms of: <ul style="list-style-type: none"> <li>Average income</li> <li>Language and ethnic diversity</li> <li>Rural / urban split</li> <li>Population density</li> </ul> </li> <li>Comparison regions should also be analysed for the likelihood of contamination, e.g. will a similar intervention be carried out in that region through other programmes?</li> </ul>	<ul style="list-style-type: none"> <li>Comparison communities should be selected from within the identified regions to be sampled from. The sampling should take into account similar characteristic and demographic information that is applied to sampling framework for treatment groups, for example: <ul style="list-style-type: none"> <li>Average income</li> <li>Language and ethnic diversity</li> <li>Rural / urban split</li> </ul> </li> <li>Where possible, the chosen comparison communities should have comparable statistics against all of these characteristics, and have minimal identifiable risk of contamination.</li> </ul>	<ul style="list-style-type: none"> <li>In general, the comparison schools and grades should be selected in a similar way as the treatment schools and grades are sampled. If the treatment schools are affordable private institutions, then it is important that comparison schools should match this.</li> <li>In identifying potential comparison schools in a community, the sampling framework should record and seek to balance: <ul style="list-style-type: none"> <li>Size of the schools</li> <li>Primary/secondary split</li> <li>Government/comm unit ownership and management</li> <li>Average educational attainment</li> </ul> </li> <li>Again, the risk of contamination should be considered.</li> </ul>	<ul style="list-style-type: none"> <li>Comparison individuals need to match the treatment individuals as closely as possible in order to form a good comparison. The sampling framework for comparison individuals should include identical criteria to what was used for identifying individuals for the treatment sample, e.g. if the treatment group is formed of children with disabilities, the comparison group should also be formed of children with disabilities.</li> <li>The sampling framework should also specify criteria around: <ul style="list-style-type: none"> <li>Grade / age ranges</li> <li>Enrolment in comparison schools</li> <li>Marginalisation criteria used for selection of treatment sample</li> </ul> </li> </ul>

The principles of representative sampling laid out in

**Table** generally suggest that the same sampling criteria should be used for both treatment and comparison regions, schools, communities, and individuals.

In their sampling frameworks, projects should focus on providing the relative information on the 'cluster' variables in the above table, i.e. regions, communities, and schools. Where projects feel that further clustering levels exist, or where clearer definitions could be given, they should feel free to amend the guidance above, but keep with the same principles. If there are project-specific targeting categories for communities, for example, then these specific criteria should be measured in the additional column provided for 'sampling point stratifiers.'

### 13.3 Sampling error, compliance, and contamination

GEC-T evaluations should be structured to identify samples that are not only representative in terms of demographics and regional characteristics, but also have a treatment sample that is made up of actual beneficiaries, and a comparison sample with little contamination.

The idea of minimising sampling error is the key in making sure that the treatment sample only captures those individuals with whom the project intends to reach with the interventions.

#### *Definition 6 : Sampling Error*



**Sampling error** occurs when individuals sampled as treatment are not part of the target population, and hence there is no intention to treat.

All sampling approaches, at cluster and individual levels, should be designed to minimise this sampling error. Communities, for example, should only be considered by a sampling technique if the project is targeting interventions in that community (or, at a minimum, with schools in that community). If sampling error is large and unidentified, then it could significantly dilute the final evaluation results.

A related concept is compliance, however this can only be estimated from evaluation points beyond the baseline.

#### *Definition 7: Compliance*



**Compliance rate** is the proportion of individuals selected in the treatment sample who have received the effects of the intervention.

This differs from the basic sampling error above. Sampling error was a forward looking measure designed to estimate the proportion of individuals in the treatment sample with whom the project has an intention to treat. Compliance is a backward looking measure that seeks to measure the success of the project in rolling out activities to reach the targeted sample. It is possible to have an evaluation where all individuals in the treatment sample were due to receive the benefits of the intervention (i.e. a minimised sampling error), but that by the first evaluation point, none of the sampled girls had successfully been reached by the project activities (i.e. zero compliance). Any changes in outcomes between the two evaluation points would therefore not be attributable to the project activities, as the sampled girls did not actually receive these activities.

Contamination is a related concept, in that it measures the extent to which the *comparison* group has remained isolated from the effect of the GEC-T project, or other similar projects' interventions.

#### *Definition 8: Contamination*





**Contamination** is the proportion of individuals selected in the comparison sample who have received the effects of the intervention, or of similar interventions.

Highly contaminated comparison samples can lead to evaluation results that underreport the actual impact of the project. The difference in difference approach relies on the comparison group acting as a counter-factual for what would have happened to the beneficiaries had the project activities not occurred. If the comparison group has also received the benefits of the intervention, then this counter-factual is void. In such a case, we would expect to see improvements in the treatment group matching closely with the improvements in the comparison group, because both groups are effectively receiving the same intervention. This would give a difference in difference value of zero, suggesting that the project intervention has had no impact whatsoever on its beneficiaries. This is clearly misleading.

In sample framework development, efforts should be made to select comparison areas where contamination can be minimised. Selecting a comparison community that is geographically close to an intervention one may open up risks of contaminating the comparison community through spillover effects, for example.

It is also important to note that contamination in comparison samples can also occur where similar interventions from other organisations are being carried out in a sampled area. Projects should never act to stop other interventions from moving into an area that they have identified as part of the comparison sample, but in sampling framework development, projects and evaluators should be aware of regions where other interventions are being planned or carried out in, and avoid these where possible.

Evaluators should look to use sampling techniques that minimise sampling error and contamination in both treatment and comparison groups. They should verify any anticipated contamination in treatment and control areas through consultation with relevant stakeholders, for example with local DFID country offices. They should also make plans to collect detailed exposure information, relevant to each project's theory of change, during their fieldwork and analysis at each evaluation point.

### 13.4 Sampling points

Many of the following sections talk about choosing appropriate *sampling points* for the relevant research. Sampling points should be understood as the locations where the research will be carried out at each evaluation point. A household sampling point, for example, means that researchers will visit individual households and administer whatever surveys or tests should be taken there.

#### *Definition 9: Sampling points*



**A sampling point** is the location from which data is physically collected during fieldwork at each evaluation point.

Sampling points are important to define in a longitudinal study like the GEC-T. To facilitate robust comparisons over time, it is important that the sampling point from which that information has been collected has been consistent between the two periods. If a certain question on self-esteem, for example, was asked at the school at baseline, and then asked at households in the subsequent evaluation point, inherent biases in how girls' answer at each sampling point would mean that the comparison may be void.

It is important, therefore, that MEL frameworks define the sampling point for each tool very clearly. The following sections on outcomes and intermediate outcomes lay out suggestions

for sampling approaches that maximise the chances of a successful evaluation, including comments on the most appropriate sampling points for collecting relevant information.

The major sampling points for individuals on GEC-T are expected to be the **school**, and the **household**.

In GEC 1, it was expected that projects would select one of these two options as their primary sampling point. For GEC-T, this distinction is less clear. As is described in the sampling for learning, and sampling for transition sections below, in order to maximise the chances of collecting data that can say something powerful and useful about learning, transition, and intermediate outcomes, a more mixed approach to sampling will likely need to be followed. This would ideally be formed of linked household and school-based samples, with the same individuals interviewed in both locations, and more detail will be given below.

To draw individuals from these sampling points, evaluators will be expected to use a clustered sampling approach: generally with communities as the cluster for selecting individuals from households, and schools also being selected from the communities.<sup>19</sup>

This will require evaluators to be prepared to develop separate sampling frameworks: one for schools, and one for households/communities. This is reflected in the design of the sampling framework template in Annex F, where separate tabs are provided for communities and schools. The principles of representative and unbiased sampling applies equally to both approaches, as well as to the separate sampling strategies that should be developed at each evaluation point for qualitative research.

### 13.5 Suggesting a comparison group

Projects will have to make suggestions on which regions, communities, and schools they would consider as a fair representation of the intervention groups. Projects will be expected to provide a list of these suggestions in the sampling framework template, and evaluators will be expected to test these suggestions for relevance and accuracy.

In making this suggestion, projects should look to consider factors such as those listed in the table below (but also not necessarily limited to only these factors):

**Table 11: Considerations for suggesting a comparison sample for learning**

Community	School
Average household income	Levels of education (primary/secondary)
Rural/urban split	Management of school (government/private)
Linguistic diversity	Average distance girls travel
Levels of marginalisation	Average achievement in national exams

The sampling framework is designed to be flexible to project and evaluation needs, and so the above table should not be viewed as exhaustive. The sampling framework template in Annex D reflects this, with additional column included for those project-specific characteristics that should be matched between treatment and comparison groups.

<sup>19</sup> Interestingly, schools also act as clusters for selecting individuals from, so this can be understood as a multi-stage clustering approach, where communities are used as clusters to select schools, and then schools are used as clusters to sample individuals. Further detail is provided in the sampling for learning sections below.

Evaluators will be expected to verify the appropriateness of the project suggestions. If the evaluator disagrees with the project suggestions in the framework, then they will be expected to provide and test alternatives.

Key things for evaluators to look for in testing the appropriateness of comparison schools and communities include the attributes explained in the table above, but also include the potential risks of contamination in the comparison communities and schools. If relevant stakeholders, including DFID country office staff, suggest that communities identified as potential comparison regions may come into the realm of other interventions during the course of GEC-T funding, then these areas should not be considered appropriate.

Likewise, projects and evaluators should liaise with other projects in GEC-T to make sure that selected control communities and schools do not overlap with other projects' interventions.

### 13.6 Power calculations and sample size determination

As described in Section 10 the default estimator of impact for GEC project level evaluations is the difference-in-difference estimate. This will be the case for both learning and transition outcomes, and any intermediate outcomes which apply comparison groups<sup>20</sup>.

This estimate aims to capture the average outcome effect based on a difference-in-differences coefficient (DID): the average *changes* in learning or transition outcomes before and after implementation for girls between the treatment and comparison groups. In practice, this estimator can be computed in two equivalent ways:

$$\beta = (\bar{Y}_{treat,after} - \bar{Y}_{treat,before}) - (\bar{Y}_{control,after} - \bar{Y}_{control,before})$$

$$\beta = \bar{y}_{treat} - \bar{y}_{control}$$

Where  $\bar{Y}$  are averages of the *levels* of the outcomes (i.e. average raw learning scores for learning, or the survival rates for the transition outcome), and  $\bar{y}$  are averages of the *changes* to the outcomes before and after implementation. These averages are taken across individuals in the same group, according to the subscript. In GEC terminology, the beta coefficient ( $\beta$ ) is referred to as *achievement* of the projects.

Sample size selection should be derived using power calculations and presented in project MEL frameworks. The sampling methodology set out in the MEL framework should be adhered to during the baseline and subsequent evaluation points. The Fund Manager should be informed of any changes to the sample size or sampling frame during the course of the evaluation.

<sup>20</sup> For pre-post-evaluations, where the exclusion of comparison groups has been agreed with the Fund Manager, a single-difference estimator would be the default. The exact formula for calculating outcomes will change to account for this, however the remaining advice on sampling in this chapter remains relevant for these projects. The beta coefficient for these evaluations with no comparison groups will either be:

$$1) \quad \beta = (\bar{Y}_{treat,after} - \bar{Y}_{treat,before})$$

or:

$$2) \quad \beta = \bar{y}_{treat}$$

Where, as above,  $\bar{Y}$  are averages of the *levels* (i.e. average raw learning scores) of the learning scores and  $\bar{y}$  are averages of the *changes* of the learning scores before and after implementation. Projects not following a comparison group approach have been made on an exceptional basis. All other GEC-T projects are expected to use the difference in difference estimators described above.

In order to estimate the sample size required to determine whether a given effect is statistically significant, projects need to be aware and consider the following terminology and factors:

- A **hypothesis test**: is the basis of the test of whether a change has taken place. The *null hypothesis* will be that a project has made no change to an outcome. The evaluation should therefore test whether this null hypothesis can be rejected in favour of an *alternative hypothesis*, in which the project has a statistically significant change in the given outcome.
- The **minimum detectable effect (MDE)**: this is the effect that a given sample size can calculate. For GEC-T projects samples for transition and learning should be determined to at a minimum detect an effect equal to the target (0.25 standard deviations per year for learning, and a change to transition rates to be agreed). The effect size is the difference-in-difference estimate – i.e. the change in outcome of the treatment group vs. the comparison group.
- **Binary and continuous variables**: the expression of the effect varies depending on the type of variable. For GEC-T, learning is likely to be measured by a continuous variable, meaning the value of learning levels (i.e. scores on a test) can take a range of values. Change for a continuous variable is expressed in standard deviation (SD) terms as well the MDE, which for GEC-T should be 0.25SD per year. For Transition, the variable of interest is likely to take a binary form – i.e. a girl either transitions ( $T=1$ ) or she does not ( $T=0$ ). For a binary variable the change is therefore expressed in proportions, where a baseline proportion may be, say, 40% of girls transitioning, and with a target of 10% improvement, the midline value would be 50% of girls transitioning. In this case for sample size purposes,  $p_1=0.4$ , and  $p_2=0.5$ .
- **One-sided / two-sided test**: the difference between these two test types is whether the effect size could go in one or both directions and whether this would be a meaningful effect. In general, when researchers can predict in which way (positive or negative) the intervention is expected to have an effect a one-sided test is used. This is likely to be the default position for GEC projects which expect to have a positive effect. However, in some cases where a project is testing different types of intervention against one another they may prefer to use a two-sided test to compare outcomes more thoroughly.
- **Level of significance (or alpha)**: is the principle measure of statistical significance and whether a false positive result has been measured. It should be set at 0.05 or 5%. The converse of this measure is the **confidence level** which would be 95% for a 5% level of significance. This latter is the degree of statistical confidence an evaluation can have that the effect measured did not occur simply by chance.
- **Power**: is the statistical measure of the degree of confidence an evaluation did not measure a false negative result. This must be at least 0.8 or 80% for GEC. Sometimes it may be appropriate to choose a higher level of Power (e.g. 0.9 or 0.95). Setting the Power higher means there would be less chance of a false negative result. Increasing power is thus beneficial for the evaluation, but it would typically require a larger sample so it may need to be discussed with the Fund Manager.
- **Allocation ratio** – is the relative balance in size between treatment and comparison groups. All else equal it is better for this to be a 1:1 ratio with an equal number of treatment and comparison communities, particularly to keep the overall sample size down. However, for practical reasons this may not be possible. A 2:1 ratio would mean that the treatment group size is twice the comparison group size, and will lead to a different sample size in calculations. This might be justified if there are cost savings from going to more treatment than comparison schools.

- **Clustering** – is used in circumstances where it may not be feasible to visit all schools to conduct sampling. Using cluster sampling, projects may visit a **subset of the schools to conduct research**. However, this requires a revision upwards on total sample size as a larger sample is required to take into account clustering effect on a school (or school catchment area) level (design effect). This technique is often used in social research as outcomes tend to cluster around a geographical location or community.
- **Intra-cluster correlation (ICC)**: is the key measure of how much clustering will impact the sample size (if applicable). It is a measure of the degree to which geographical specificity (for example a given region, district or school) determines the outcome levels or changes to outcomes. **The design effect** is a number to multiply the sample size by, in order to allow the survey to achieve the same power and precision despite the effects of not sampling in every location, and is determined by the ICC and the number of clusters that will be visited in the research. *Projects with effective clustering data could use their GEC-1 data to examine what the ICC should be.* This can be found using the Stata command `icc`.
- **Attrition** – is the loss in the sample size over time. In all cases, the sample size given by the power calculations provides the minimum number of sample required at each evaluation point in order to detect a statistically significant learning improvement. However, at baseline, projects will need to increase the sample size in order to take in to account sample attrition. Projects will need to make an assumption regarding the annual attrition rate that could be encountered given the context of intervention and given their experience from GEC-1. This could range from 30% - 50% across different projects depending on the challenges they face. The baseline sample size will need uplift in order to take into account attrition.

### 13.7 Methods to calculate sample sizes

Optimal sample sizes could be determined by statistical programmes and available formulas. A few example tools recommended by the Fund Manager that can be useful for calculating sample sizes are:

*Stata statistical software (not free)*

- The Fund Manager would suggest using **Stata** if your statistician / evaluation consultant already has this software. In particular, the command **Power**, which can be used to select sample sizes. Or the command: **Sampsi** (or use the interactive menu), with slightly different options to those used when working with proportions (i.e. including the relevant standard deviations).
- For a treatment and comparison group assessment then the additional command used should be **twomeans** for continuous variables (i.e. learning in GEC-T), and **twoproportions** for binary variables (i.e. transition in GEC-T).
- The **ICC** can be measured in **Stata** using the command **loneaway X Y**, where X is the outcome variable of interest, and Y is the cluster definition of interest (i.e. school code, community code etc.)

**G\*Power** (free)

- If not, the Fund Manager would recommend **G\*Power**. This can be downloaded for [free](#).
- The command for continuous variables is: **t tests-> Means: difference between two independent means (two groups) -> A priori: Compute required sample size – given  $\alpha$ , power, and effect size.**

- The command for binary variables is: **z-tests-> Proportions: difference between two independent proportions -> A priori: Compute required sample size – given  $\alpha$ , power, and effect size.**

#### *Optimal Design (free)*

Optimal Design can be used when the evaluation strategy involves clustering methodology.

#### **Minimum standards for sample sizes for learning and transition outcomes**

Minimum detectable effect	0.25 standard deviations for Learning sample; 10% for Transition sample <sup>21</sup>
Level of significance	5%
Power	80%
Attrition buffer	30%
Intra-cluster correlation	0.1

Any parameters required for the sample size calculation that are not supplied as an FM minimum expectation in the table above, should be based on the available evidence for these parameters as collected through secondary data such as census information. This will be the responsibility of the project evaluator to finalise.

For those evaluations which intend to link the learning and transition samples as described in section 16, sample sizes for both learning and transition should be calculated as if the samples were intended to be collected independently. Evaluations should then be designed using whichever calculation gives the higher sample size. Evaluators should identify this number of individuals in schools and then attempt to recontact them at their households. More detail is provided following the individual sections on learning and transition, respectively.

## **14. Sampling for Learning**

The following section discusses the FM expectations for designing a robust sampling approach for measuring learning throughout the course of the project. While the previous section on the learning outcome focused more on the definitions, tools, and ambition for learning as a GEC-T outcome, this section focuses on who to administer the tools to, in both treatment and comparison groups.

The main content of this section covers:

- Selecting baseline samples for treatment and comparison learning cohorts
- Benchmarking for learning outcomes
- Tracking approaches for the learning cohort

Projects and evaluators should note that the following guidance relates equally to literacy, numeracy, and any optional third learning outcome<sup>22</sup> that the project has chosen to measure progress against.

<sup>21</sup> This 10% will not likely be the transition target for each project, as the targets for transition will be set after baseline using the methods explained in below sections in this guidance. This 10% is merely a reference point to calculate an initial sample size on, and the target set for each project is likely to differ from this.

<sup>22</sup> The working assumption for GEC-T evaluations will be that third learning outcomes are also examined in comparison schools. It may be inappropriate for all third learning outcomes to be tested in comparison schools, especially where there are no opportunities for girls to develop skills in the relevant area. For projects selecting ICT as a third learning outcome, for example, if a comparison school does not offer access to computers then it may be inappropriate to test them on their ICT literacy. Projects



## 14.1 Baseline learning treatment sample selection

The learning treatment sample should be rooted in the target groups identified for each project. Project evaluations should aim to draw a treatment sample that best matches this target group.

To maximise this, baseline sampling for the learning treatment cohort should happen primarily around the schools or institutions in which an intervention is planned. This means that a sample of beneficiary girls should be identified at schools<sup>23</sup>, examined using an appropriate learning test, have the school survey administered to them, and then tracked longitudinally through the course of the evaluation.

With this in mind, the baseline learning sample should be constructed using a clustered approach, where schools are selected as the clusters for individuals to be drawn from. The population of potential treatment schools should be identified in the sampling framework, an appropriate technique applied to narrow down to a list of sampled schools (potentially random or stratified sampling, depending on what the evaluator feels is most appropriate), and then a further sampling technique used to draw individuals from these clusters.

Assuming that a representative and unbiased list of sample schools has been selected using random or stratified sampling, which fits with the number of clusters recommended from the minimum sample size calculations described above, the evaluator should then look to develop a secondary sampling techniques used to draw individuals for the learning sample from the cluster schools.

This sampling technique for individuals should be based on the target group criteria the project has identified. Using lists of girls provided by each school based on age, grade, and then marginalisation status relevant to the project's theory of change, for example, evaluators should draw a random sample of individuals which meets the relevant beneficiary criteria.

If a project works with five different grade ranges, then stratified sampling should be applied to draw a balanced number of individuals from each of these ranges. Each school/cluster should ideally have an equal number of individuals drawn from it. Evaluators should carefully consider the treatment criteria, and apply what balancing stratification they feel appropriate.

The baseline treatment sample for learning, therefore, is comprised of individuals who fulfil the relevant target group criteria, drawn randomly from grades in each sampled school/cluster.

Once schools and individuals for the learning baseline have been identified, the expectation is for evaluators to administer the learning tests and record results accurately. Additionally, as will be explained more in the section on sampling for transition, it is also of vital importance that detailed information that could help evaluators locate and re-contact girls at their households is collected. This could be through, for example, asking girls, teachers, head teachers, or community leaders for contact information.

Furthermore, exposure data should be collected that helps identify which precise interventions (if any) a girl has received. At baseline this is likely to be indicative of the extent to which the sampled girls represent the project's GEC 1 beneficiaries. This exposure data should also

---

should consider the viability of a comparison group for their third learning outcome, and provide justification in their MEL framework if they believe it should be exempt from using a comparison school.

<sup>23</sup> Projects and evaluators should note that when the term 'school' is used, the definition is very broad. We intend the term 'in-school' to refer to when a girl is enrolled in an institution or intervention that is key to the project's theory of change, including, for example, informal or ALP courses. A more accurate term may be 'in-intervention', but for simplicity we will continue with the terms school and in-school.



look to show whether or not individuals from sample treatment communities are recipients of other educational interventions.

Individuals in the baseline learning treatment sample are expected to be longitudinally tracked through the course of the project.

## 14.2 Baseline benchmarking for learning

As described in Section 8, in the general case target setting for learning will follow a similar methodology to GEC 1, where benchmark results will be taken from older grades that learning beneficiaries are expected to move into. Based on the standard deviations of the collected results, a target will be calculated. Progress against this target is measured using a difference in difference approach.

Of importance to the sampling approach is that this target setting is based off a benchmark learning sample. This means that the baseline learning sample includes not only beneficiaries expected to be tracked over the course of the project, but also a set of 'one-off' sampled individuals where results from their learning tests are taken solely for the purposes of establishing a benchmark.

This benchmark sample should be identified and sampled using the same clusters as for individuals the project and evaluator decides to track longitudinally. Suitable demographic information should be collected to check that any bias in the sampling technique has been minimised.

Benchmark learning tests need only be administered to higher grade girls in beneficiary schools. There is no requirement for benchmarking in comparison schools.

A generalised summary of the steps behind benchmarking for learning is:

1. Identify the upper boundary of your learning beneficiaries: assuming that they will stay enrolled in the interventions through the life of the project, what is the last stage that the project could actively work with these individuals to improve their learning? Evaluators should collect baseline learning information for all grades between the lowest baseline grade and this upper limit benchmark grade.
2. Identify the necessary sample size needed for treatment and comparison areas (see further detail above for exact methods). From this total, decide how many individuals will need to be sampled from each grade. Benchmark grades should need fewer individuals sampled than the grades intended to be tracked.

In summary, the benchmarking process for learning requires a separate benchmark sample to be established, and this sample will not be tracked.

## 14.3 Baseline learning comparison sample selection

The general principle for establishing a comparison group is that it should match closely with the criteria for selecting the individuals from the treatment group. The general approach to securing this should be similar to the sampling approach for the baseline learning treatment sample: the project identifies the population of treatment communities, the schools within these communities, and the evaluator applies a sampling technique to filter these down to sampled schools which act as clusters to draw individuals randomly from. The process should be fairly similar to the process for selecting appropriate treatment schools and individuals. Projects should aim to provide:

- The same selection criteria used for identifying individuals in treatment schools

- Details (including geographical information) of suggested comparison communities, and schools within those communities
- As with the selection of treatment schools, any balances that the sampling choice needs to make, e.g. does the sample need a balanced geography, is there a certain ratio of primary to secondary schools that needs selected? These can be guaranteed through setting a stratified sampling technique to account for these things.
- The minimum sample size for the learning comparison group

Evaluators should be prepared to:

- Further evaluate the appropriateness of project suggested comparison areas and schools, by analysing relevant demographic variables for comparability
- Investigate the potential effects of contamination in selected comparison communities. Through engaging with DFID, other donors, policy makers, and local stakeholders, is their evidence to suggest that similar educational interventions will move in to the suggested comparison areas?
- Decide on the number and location of selected comparison schools
- The number of learning tests that need to be administered in each school
- The final selection criteria for individuals in these institutions (should generally be consistent with treatment group criteria)

Evaluators should collect the same information for comparison individuals as the treatment individuals. This includes questions on exposure, contamination, and demographics, as well as vital re-contact information to allow evaluators to follow up comparison individuals at the household in a similar way to the expectations for treatment individuals.

#### 14.4 Sampling and cohort tracking for learning beyond baseline

The baseline learning sample should be established using the methods articulated earlier in this chapter, paying particular attention to geography, marginalisation criteria, and the potential comparability and biases of using different comparison areas. Beyond this baseline, expectations for projects and evaluators will be that:

- Learning tests for treatment and comparison samples in subsequent evaluation points will be conducted in schools, in line with the approach for the baseline learning sample
- Benchmark grades do not need to be assessed again after the baseline
- Cohort tracking is preferred, i.e. it should be the priority to re-identify the same girls from the baseline learning sample and administer tests again to understand the individual changes in their scores between the two periods
- Where the sample size for a school cannot be reached solely by recontacting the same girls, substitution should take place using a substitution protocol built from the same sampling criteria used to identify suitable individuals for the baseline sample. Further guidance will be given on this
- Detailed household re-contact information will be collected from the newly substituted girls, in order to feed into the transition outcome for the next evaluation point

Cohort tracking is still expected as the aim for learning evaluations. Evaluators should structure their record keeping and data collection in such a way as to maximise the likelihood of identifying the same girl at the next evaluation point if she has stayed in school. Strategies for achieving this may include (among others):

- Clear and consistent creation of a unique ID for each sampled individual
- Careful recording of names and ages, checking for accuracy with teachers, school records, and the sampled individuals themselves
- Recording (where appropriate) information such as community, address, or telephone numbers that can be used at subsequent evaluation points for individuals to confirm if they are the same person as was interviewed in the previous round
- Taking note, and storing clearly, other features as appropriate for the context of the project intervention

Projects and evaluators should note the need for cohort tracking in both treatment and comparison schools.

Cohort tracking should be attempted as a minimum standard for all GEC-T projects, however, with the added challenges of GEC-T moving to focus on older individuals, it is understood that chances for drop-out and attrition will get progressively higher as the project continues. This is especially true for projects straddling a formal educational transition point, such as children moving from primary school into secondary school. In these cases, it is unlikely that all children from sampled primary schools will enrol in the list of sampled secondary schools, and so replacement strategies should be developed to account for the potential drop in sample sizes across the life of the project and evaluation.

Replacement strategies should be developed by evaluators before research is conducted at each evaluation point. More guidance on this will be provided at each evaluation point, along with the key documents such as report templates and outcomes spreadsheets. These strategies should aim to replace a girl with another in-school girl who closely matches the demographics of the lost girl, as well as mirroring her level of exposure to the project intervention. If a girl has been lost from the sample, but had been enrolled in the intervention for two years, for example, then she should be replaced by an in-school girl, preferably in the same class, who also had been enrolled in the intervention for two years. In this way, the replaced girls should match the lost girls in terms of demographics, marginalisation status, and level of exposure to the intervention.

Evaluators, in future evaluation points beyond the baseline, should look to collect information on the attrition rates, measuring how many individuals have been recontacted successfully, how many have been lost from the previous evaluation, and how many have been substituted into the sample. Further guidance on this will be given as project approach the second evaluation point, however evaluators should be aware of this expectation in designing their research approach.

## 15. Sampling for Transition

As described in Section 6, transition is a measurement that requires a different sampling approach than learning in order to capture the relevant information on the many pathways that children can take.

Sampling for transition must account for the journey made by girls that drop out of, or graduate from project interventions. It is important, therefore, for the evaluation to be able to capture

information from individuals who move beyond the initial intervention institutions, and to track these pathways effectively. Sampling from a household level is more likely to allow evaluators to capture this varied information than from sampling at a school level.

The household-based sampling for transition is favoured for several reasons:

- Sampling at school level may inherently bias transition measurement as evaluators are more likely to only capture information from children in school, and therefore not capture the rich information on the variety of pathways that individuals may follow beyond the intervention.
- Transition is longitudinal by nature, measuring the change in an individual's enrolment status from one period to the next. Sampling from households makes it easier for evaluators to locate and recontact the same girls across different periods.
- Transition is a more open-ended measure than learning, as girls can transition to multiple different pathways over the course of the project. Starting at the household should allow evaluators to capture varied information on where girls start from and go to over time. This is built on an assumption that while institutions may vary, the household is likely to be a more permanent location.
- Household based sampling also opens up more opportunities for qualitative exploration on household attitudes towards education and transition.

The more appropriate sampling point for measuring transition outcomes, therefore, is the household. Successful tracking of individuals at a household level should allow projects and evaluators to capture the movement of these individuals over time, and this is crucial to understanding the transition pathways that such individuals follow.

Before considering how to select individuals for treatment and comparison samples, however, it is important to consider the geographical biases which may be at play when sampling communities for transition.

### 15.1 Baseline transition treatment sample selection

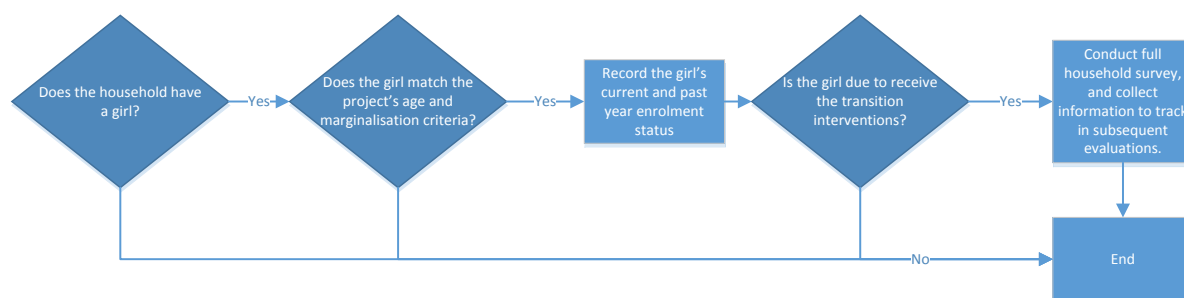
With the household as the key sampling point for transition, the cluster from which these households should be drawn is the community. While learning samples were created using schools as the ultimate cluster, communities are key for transition.

Projects should provide the full population list of targeted communities in their sampling frameworks. Evaluators will be expected to apply a sampling technique to identify which communities should act as research clusters for transition. With these clusters identified, evaluators should define an exact point in each community (i.e. the 'centre' of the community) from which a systematic sampling approach can begin. An example of such an approach might be, 'conduct interviews at every third household from the centre of the community.' In this case the 'centre of the community' needs to be clearly defined. This should be applied consistently across all sampled communities (treatment and comparison) and across all the evaluation points scheduled for the project.

The main research aim for the transition outcome is to understand how many treatment girls have successfully transitioned relative to their counterparts receiving none of the project's interventions. As such, it is important that the transition treatment sample captures information from beneficiaries the project intends to work with, i.e. that it minimises the sampling error in a similar way that has been described for learning.

The project should identify the interventions that have been designed to aid beneficiaries' transitions, and which groups these interventions will be aimed at. Evaluators should use this information to develop screening criteria for individuals and households: if a household is visited which does not fall into the transition target groups identified by the project, then they should be excluded from further research and should not have the full household survey

If the project's transition activities are being carried out in project schools, for example, then the defining feature for a household to be administered the full survey will be if they have a qualifying individual presently enrolled in a project school. The below diagram demonstrates the kinds of questions evaluators should ask when they arrive at a household to estimate its appropriateness for the evaluation:



For transition, it is key that information is collected on two points in time, i.e. where each interviewed individual is currently enrolled<sup>24</sup> and where they were enrolled last year. This will be the same principle for all subsequent evaluation points as well.

## 15.2 Baseline transition comparison sample selection

Projects and evaluators should follow the same sampling process for comparison communities as was carried out for treatment communities:

- Project identifies potential communities in a comparison region
- Evaluator verifies these suggestions, and applies a sampling technique to select a subset of communities to act as research clusters
- Evaluator defines a systematic sampling strategy for selecting individual households from within these sampled communities
- Households are sampled until the minimum sample size is reached

For treatment households, it was important to identify if they were due to receive the benefit of the intervention before they were administered the full household survey. Comparison communities will not have this same criteria to fulfil, however it is important that they match the general characteristics of the treatment sample.

If the transition activities are aimed at in-school individuals, for example, then the treatment sample will likely consist of entirely in-school cohort. The comparison sample should, therefore, also be comprised of largely in-school girls. This should be worked into the sampling criteria for individual comparison households, and should be defined clearly in the project's MEL framework.

<sup>24</sup> This 'enrolment' is broadly defined in order to capture where the individual spends the most of their time. This could be, for example, in a formal school grade, in non-formal education or vocational training, in employment, or out of school. Transition will aim to capture the movement between these enrolment points.

### 15.3 Baseline benchmarking for transition

Whereas learning benchmarks should be collected at the same time and in the same location as the general learning baseline is carried out, i.e. in the school, the transition baseline benchmarking cannot be carried out in the school.

A transition benchmark should aim to collect information on the general levels of transition for girls in the beneficiary communities. In particular, it should capture the transition levels of girls within the complete age range the project intends to work with. If a three year project commences working with girls aged 9-12, then the transition benchmark should aim to collect information on girls aged 9-15. Projects should define this age range clearly in their MEL framework.

While the transition treatment sample focused on finding girls who would receive the benefit of the intervention, the transition benchmark does not require such precision in its sampling. The aim of the transition benchmark is to define the existing picture of transition for the entire beneficiary population, in order to set realistic targets. The transition sampling criteria should therefore be much broader, in most cases based only on the relevant age range.

This benchmark should be sampled representatively to avoid geographical bias. If transition benchmarks are taken solely from communities situated very closely to secondary schools, for example, it will appear that the benchmark transition rate is very high. This may not represent the full picture, however, if a project's interventions stretch over a much broader geographical area, with some regions with very poor supply of secondary schools.

The survey administered for this benchmark sample need not be very long. The key pieces of information to collect from the households will be the current enrolment/employment status of the sampled girl, her previous year's enrolment/employment status, and the critical demographic information needed to check how closely the benchmark matches the transition sample. This should be a very short survey to administer.

The transition benchmark should consist of girls who match all the relevant intervention criteria (age, location, etc.), but are not enrolled in a project intervention. These girls should have the short benchmark survey administered to them, whereas girls who are also enrolled in the project intervention should have the full survey administered, and also be included as part of the transition cohort for tracking.

The benchmark sample captured here need not be tracked again in future periods, especially as there is no idea of the level of compliance in this sample, i.e. it is unclear how much these sampled individuals may gain from project interventions. As such, it is unnecessary to track this sample, and it is advised that this baseline benchmark for transition is used only for the purposes of setting baseline targets and for comparing progress in the treatment and comparison groups.

### 15.4 Tracking for transition beyond baseline

It is expected that the evaluation will follow up with all sampled households in subsequent evaluation points. It is crucial, therefore, for information to be collected that will allow evaluators to locate each household again. Such information may include, for example, phone numbers, addresses, and notable features about the neighbourhood.

Replacement for transition will be much more difficult than replacement in schools, as the nature of transition relies on comparing where individuals are enrolled in one year compared to the previous one. Therefore, initial sample sizes should be inflated to account for potential attrition in the transition sample. Projects should make clear in their MEL frameworks the



extent of additional attrition buffer being suggested, and how this impacts the overall sample size for transition.

## 16. Principles for linking learning and transition samples

As a minimum standard, evaluators can treat the transition sample as separate and distinct from the learning sample.

As stated above, however, the preferred sampling method would be to link the learning and transition samples, through following up with girls who sat learning tests at school at their homes. This would have several benefits, including potentially simplifying the sampling approach. This section will explore more on how this linking could be achieved.

For clarity, however, it should be stressed that this linking of transition and learning samples is not the expectation, but a recommendation.

Benefits of linking the samples include:

- Minimised sampling error for the transition sample, as all individuals in the treatment group would be beneficiaries of the project interventions.
- Greater ability to link learning outcomes to changes in transition, and other characteristics that are best measured at the household (e.g. parental attitudes). This increases the questions that the evaluation can answer, for example through understanding if successful transition occurs as a result of improved learning outcomes, or if learning is dependent on successful transition.
- Potentially reduce the cost of identifying a brand new sample to collect information from, working with existing sampled individuals.

The process of linking learning and transition relies on the collection of detailed contact information on each individual who sits a learning test. If such information is collected, then evaluators should aim to recontact these girls at their households, administering the relevant household survey and collecting key information on transition outcomes.

If these girls are successfully 'followed' back to the household after sitting the learning test, they now constitute a data point in both the learning and transition sample.

A more ambitious evaluation approach will attempt this link, the process to do so could look like this:

1. Evaluators use a learning sampling framework to locate and identify sample of in-school learning beneficiaries who will sit the evaluation learning tests and school-based survey
2. Detailed contact information should be collected on each girl who sits the tests. This contact information should be triangulated with school management information, details collected from girls, teachers, head teachers and others who may have relevant information to locate each girl's household<sup>25</sup>

---

<sup>25</sup> Evaluators should consider carefully the importance of consent in gathering such information. Caregivers should receive adequate notice and time to consent to the study, especially if evaluators are attempting to follow up with girls at their households. Information on addresses and locations of households, for example, should not be collected or used without prior notice being given of the evaluation's intent to do so.



3. Evaluators schedule follow-up research at the households of the girls from the learning sample. Ideally, every girl who sat learning assessments, from both treatment and comparison groups, would be followed up at the household.
4. Researchers conduct household data collection using household surveys, and collect further information on households to aid with recontacting the exact same household at the next evaluation point, for example, collecting caregivers' phone numbers, GPS coordinates (these would also be extremely useful for spatial analysis of the sample) or distinctive location features to help locate the same house.
5. Results from household survey should be recorded using the same unique ID assigned to the girls at the school. Her results from both school-based research (tests and school survey) and the household-based research (household survey) should be merged and stored as one record in the evaluation dataset.

In terms of the timing of this research, it is important for the research to start at the school, and then move to the household. This is important to minimise sampling error in both samples. If the evaluation began at the household, for example, then it would be likely that evaluators would collect information on many girls who were not actually enrolled in the intervention. This would potentially increase the sampling error of both the learning and transition samples.

One major downside of this approach, however, is the difficulty to interview out of school girls. For those projects with a significant re-enrolment component with girls who are not currently in any project interventions, it may be important to purposively sample for out of school girls in a selected community. This could be done alongside the linked samples, where additional out of school girls are added to the transition, household-based sample, but with no learning results to correspond.

### Linking learning and transition sampling - example

A three-year project is targeting their learning and transition interventions initially at individuals about to move into grades S2 to S4 in 20 schools.

A sampling framework identifies 10 of the 20 total schools as being eligible for the learning sample. Consent is sought from the school management, individual teachers, and notes of consent are sent to the individual families of each of the girls fulfilling the relevant sampling criteria. With consent secured, a sample of individuals from classes S2 to S4 is drawn.

One of the girls identified in this sample is successfully identified in the school research visit and has both learning tests and the school-survey administered to her. She is assigned a unique ID that will also mark her information across all the remaining evaluation points.

Researchers also collect from the girl herself, her teacher, and the head teacher, relevant information on the address and location of that girl's household, as well as important contact details on her caregivers.

One week after the initial school research visit, the household research is organised based on the address and location information captured from all of the individuals in the sampled schools. Research visits are organised according to the community groupings of the identified individuals. Community leaders are contacted in advance to confirm the exact whereabouts of each identified household.

The same girl who was identified at school, and sat the learning tests, is successfully followed up at her household. The girl and her caregivers are asked the relevant questions from the household survey, including the key questions relating to transition. This information is collected and stored against that girl's individual ID that was assigned following the learning tests and school survey.

One year later, for the project's second evaluation point, the school research visit finds that the girl has not remained in school, and so cannot sit the learning assessment. Researchers collect basic information on the potential whereabouts of the girl, and potential reasons for her not being enrolled (e.g. drop-out, graduation, or employment). In order to maintain the required sample size for learning, the girl is substituted using the evaluator's replacement strategy, based on the original sampling criteria. The substituted girl is assigned a new ID following her learning tests, and is recorded as being a substitute. The original girl is recorded as having been lost from the learning sample.

Researchers then follow up at the original households, and successfully find the girl who was assessed at baseline, but had dropped out from the learning sample by the second evaluation point. The household survey is administered to the girl and her caregivers, in order to gauge transition. She should then be sought again at the final evaluation point, but at the household only.

One important nuance to note is that girls successfully contacted at both school and household for baseline should again be followed up at school and baseline for the next evaluation point. The ordering of the research at future evaluation points is potentially less important as it is at the baseline: it matters less whether researchers collect information from schools for learning first, or from households for transition first. Ultimately, however, it is important that both sampling approaches take place at each evaluation point.

Initially creating a sample from schools facilitates low sampling error in both learning and transition, and fulfils the ‘follow the girl’ principle from GEC 1 as the previous beneficiaries of learning interventions are more likely to be found in intervention schools. Following up at subsequent evaluation points at the household allows for greater tracking of the different transition pathways, more robust information on intermediate outcomes around community attitudes and engagement, and a greater ability to longitudinally track the same GEC 1 beneficiaries over time as they move out of GEC school environments.

## 17. Sampling for Intermediate Outcomes

Intermediate outcomes are a new addition for the GEC-T evaluation, and so it is important to consider how best to collect information and establish a sample to provide insight into each one.

This guidance will not go into depth on sampling strategies for each intermediate outcome because, in theory, each should follow a similar logic:

- For each intermediate outcome, choose a sampling point that is likely to capture the most insightful views on progress.
- Add the relevant intermediate outcome module to the school or household survey as appropriate.
- When administering the surveys for learning and transition, administer the relevant intermediate outcome modules as part of the household or school survey.

The aim for intermediate outcomes would be that, in the evaluation, they fit into existing structures of data collection. This allows projects to be able to make links more closely between learning, transition, and each of the intermediate outcome elements being targeted by the project. Where appropriate, projects and evaluators should aim to use the same sampling clusters that have already been established.

### 17.1 Sampling points for each intermediate outcome

Intermediate outcomes are designed to be evaluated rigorously at each evaluation point through a module of either the school or household surveys. Each intermediate outcome is likely to have different sampling point which is more appropriate, depending on where the interventions for each intermediate outcome are aimed in the project theory of change. This list is a suggestion of what locations may be most appropriate for each intermediate outcome.<sup>26</sup>

Household survey	School survey
Attitudes and perception (boys, family, community)	Attendance
Self-esteem	School governance/management
Economic opportunities/economic empowerment	Teaching quality
Life skills	

This list is suggestive, and projects will be expected, in their MEL framework, to clarify what

<sup>26</sup> This list of intermediate outcomes is not exhaustive, but has been made from looking through all the suggested intermediate outcomes that projects submitted in their final GEC-T proposals. Final intermediate outcomes should be decided in the final logframe submission, and indicators should be agreed in communication with each project's FM Evaluation Officer.

they regard as the best location for evaluating each of their intermediate outcomes. This should be justified by reference to the relevant project activities: who they are aimed at, and how they are expected to contribute to progress against the intermediate outcome. Examining teaching quality in schools, for example, may make more sense than trying to measure it at households.

## 17.2 Baseline sampling for intermediate outcomes

The core questions for measuring progress against intermediate outcomes are all designed to be asked towards the ultimate beneficiaries of the activities: the girls themselves. The learning and transition outcomes are likewise evaluated by asking questions of the beneficiary girls (and their comparison group). A suitable sample of beneficiaries and comparison, therefore, has already been formed.

There is no expectation for projects to measure intermediate outcomes using comparison groups. In line with previous guidance, however, where it is appropriate, and of minimum extra cost, projects may wish to extend intermediate outcome measurement to the comparison group. This would increase the rigour of the analysis, and would allow stronger conclusions to be reached.

Since the sampling of individuals for intermediate outcomes is due to overlap with the sampling for transition and learning, it may be possible to expand the measurement of intermediate outcomes to comparison groups at minimal extra cost. The extra cost would come in the form of additional questions in a survey already due to be administered. Projects should consider this trade-off in designing their MEL Frameworks and the terms of reference for their external evaluation.

Similarly to learning and transition, the quantitative exploration of progress will be insufficient for telling the full story. An analysis of teacher training, based solely off beneficiaries' quantitative answers in the school survey, for example, would not be enough to say how much has truly changed in teacher practice, what particular elements work well, and what barriers to good practice have been overcome. Teacher observations, interviews with teachers, and focus group discussions alongside trainers are all qualitative methods in which teacher training, as an example, may be further evaluated. Further guidance on sampling for qualitative research is given in Chapter 11 below.

Again, similar to learning and transition, detailed exposure data should be collected. This should demonstrate the extent to which the sampled individuals have benefitted from the interventions specific to the intermediate outcomes. Evaluators should prepare their approach with this in mind.

## 17.3 Sample sizes for intermediate outcome samples

In most cases the minimum sample size for intermediate outcomes will be identical to the sample size for learning and transition, as in most cases the samples will be the same.

There may be some exceptions, for example where only a subset of girls are receiving interventions related to the intermediate outcome, and in these cases, the specific sampling criteria should be developed and worked into the survey logic. Unless in exceptional circumstance, we do not expect additional individuals to be sampled for intermediate outcomes beyond the individuals identified in the schools and households for learning and transition outcomes, respectively.

## 17.4 Sampling for intermediate outcomes beyond baseline

The same cohort approach applies to intermediate outcomes as it does to full level project outcomes.

## 18. Summary of Sampling Approaches

Table 12: **Sampling Summary** shows a summary of the school-based and household-based sampling requirements.

**Table 12: Sampling Summary**

	<b>Learning</b>	<b>Transition</b>	<b>Intermediate Outcomes</b>
Sampling Point	School	Community	Defined as most appropriate for each IO
Tools	Learning tests and school survey	Household survey (including transition module)	School or household based surveys
Sample Size	Defined by standard deviation target (0.25), the number of schools to be sampled, and the learning results from the GEC 1 Endline.	Defined by assumed percentage improvement target (10%) <sup>27</sup> , the number of communities to be sampled, and any nationally available estimates of the existing transition rate	Same sample size as for learning and transition
Benchmarking Sample	Treatment schools only, taken from grades up to the maximum for eldest in cohort to reach, learning tests applied, not longitudinally tracked	Treatment communities only, individuals within the project age range, short questionnaire on enrolment applied, not longitudinally tracked	No benchmarking required
Cohort Tracking	Cohort tracking expected, but replacement strategies should be developed at each evaluation point to overcome attrition	Cohort tracking expected, but chances for replacement smaller than for learning, so tracking even more important	Cohort tracking expected, and will follow protocols for learning and transition.

<sup>27</sup> This 10% will not likely be the transition target for each project, as the targets for transition will be set after baseline using the methods explained in below sections in this guidance. This 10% is merely a reference point to calculate an initial sample size on, and the target set for each project is likely to differ from this.

## Appendix A: Useful resources

Theme	Link
Developing SMART indicators	DFID, <a href="#">SMART Guide</a>
Evaluability	ODI, <a href="#">Evaluability Checklist</a>
Theory of change	DFID, <a href="#">Review of the use of ‘Theory of Change’ in International development</a> DFID, <a href="#">examples</a> of Theories of Change
Evaluation	World Bank, 2016, <a href="#">Impact Evaluation in Practice – 2<sup>nd</sup> Edition</a> . HM Treasury, <a href="#">‘The Green Book: Appraisal and Evaluation in Central Government’</a> . 2011 update. J-PAL, <a href="#">Introduction to Evaluations</a> Esther Duflo, 2006, <a href="#">Field Experiments in Development Economics</a> .
Gender and power analyses	Sida, 2013, <a href="#">Power Analysis: Experiences and challenges (Concept note)</a> . Stockholm: Swedish International Development Cooperation Agency (Sida). European Bank for Reconstruction and Development, <a href="#">Gender Tools and Publications</a> . DFID, 2009, <a href="#">‘Gender and Social Exclusion Analysis How To Note’</a> , A Practice Paper, Department for International Development, London, UK
Experimental analyses	Banerjee and Duflo, 2008, <a href="#">The Experimental Approach to Development Economics</a> . National Bureau of Economic Research.
Child protection	Save the Children’s <a href="#">‘Assessing effectiveness in child protection’</a> <a href="#">j-pal e</a>
Educational pedagogy in developing countries	Centre for International Education, 2013, <a href="#">Pedagogy, Curriculum, Teaching Practices and Teacher Education in Developing Countries</a> . Education Rigorous Literature Review. Department for International Development.
Sustainability and community engagement	World Bank, 2012, <a href="#">What have been the impacts of World Bank Community-Driven Development programs?</a>

## Appendix B: MEL Framework template

The first draft of your MEL framework should be submitted to the FM by **30<sup>th</sup> June 2017**. It is required that a completed MEL framework should include each of the following sections. Makes sure that for each of these sections you are referring back to the detail in the preceding guidance:

### Cover Sheet (1 page)

The cover sheet should state:

- Project number, name and implementing agency;
- Name of the author (including details of the External Evaluator if used to put together the MEL framework);
- Version number; and
- Date.

### 1. Introduction (1 page)

Please provide a brief description of the context for the project, summarising key details including:

- The projects objectives.
- The main contextual factors that have influenced the project design.
- A brief summary of the project activities.
- An overview of the Theory of Change (ToC) and underlying assumptions
- A brief description of the areas and beneficiaries with which the project will work.
- Start date and length of the project.
- Overarching principles and aims of your MEL framework (which will help you integrate it into the overall design and management of the project and learn what is working and why).

### 2. Learning from GEC 1 (1 page)



Please provide a brief description of any lessons you have learned from the GEC 1 evaluation, including what actions you are planning on taking to mitigate against similar problems occurring for GEC-T.

### 3. Monitoring (for guidance see Chapter **Error! Reference source not found.**) (4 pages)

Provide details of your monitoring system, following steps in guidance Chapter **Error! Reference source not found.**. Consider how you will subsequently use this data to assist in the management, evaluation and learning of your project.

For each logframe output indicator, please provide level at which measurement will take place, tool and mode of data collection, rationale, and frequency. Summarise this information in table 1.

*Table 1: Outputs for measurement*



<b>Output</b>	<b>Level at which measurement will take place, e.g. household, school, study club etc.</b>	<b>Tool and mode of data collection</b>	<b>Rationale, i.e. why is this the most appropriate approach for this output</b>	<b>Frequency of data collection, i.e. annually, per term</b>
Output 1.1				
Output 1.2				
Output ...				
Output 2.1				
Output 2.2				
Output ...				

#### **4. Key evaluation questions** *(for guidance see Chapters Error! Reference source not found., and 5-8) (1 page)*

This section should outline how the ToC links to the key evaluation questions at both programme level and project level. These questions should be consistent with your Theory of Change and logframe output indicators, and feed into the overall project outcomes. Intermediate outcomes must also be integrated within the key evaluation questions.

In this section please also describe the logic to choosing evaluation questions as part of your learning strategies (more info to be provided under section 11 of this template).

#### **5. Evaluation design** *(for guidance see Chapters 3, 5-8, 9-17)*

This section should set out the overall evaluation approach and present the strategies for measuring all relevant outcomes and intermediate outcomes for the project. An explanation should also be provided for why this strategy is considered to be the most suitable for each outcome.

##### **5.1 Research design** *(for guidance see Chapters 10, 11, and 13) (1 page)*

Describe whether you will be using a randomised controlled trial (RCT) or quasi-experimental technique, stating the rationale and implications (sample size, budget etc.) of your choice. This section should also explain how you will ensure in your analysis that you can disaggregate the respective influence of the different elements of your intervention and the impact on different subgroups of your treatment population.

This section should also detail whether:

- The same cohort from GEC-1 will be tracked or a new sample will be selected.
- All schools/classes/areas will be sampled or a subsample will be selected.
- The same sample will initially be used for learning and transition (tracking from school to home at baseline) or two completely separate samples will be selected.

##### **5.2 Measuring outcomes** *(for guidance see Chapters 5-8) (4-8 pages)*

Explain the key intermediate outcomes and higher level outcomes you will be measuring, whether data will be collected quantitatively or qualitatively and the means by which they will

be measured. Please also state how frequently this data will be collected and the processes in place for verifying the data and results (see Chapter 3).

This information should be summarised in Table 2, which should be adapted according to your projects specific activities and measurables.

*Table 2: Outcomes for measurement*

<b>Outcome</b>	<b>Level at which measurement will take place, e.g. household, school, study club etc</b>	<b>Tool and mode of data collection, e.g. HH survey, school based survey, focus group discussions etc</b>	<b>Rationale, i.e. why is this the most appropriate approach for this outcome</b>	<b>Frequency of data collection, i.e. per evaluation point, annually, per term</b>
Literacy	<i>School/study clubs</i>	<i>EGRA</i>		
Numeracy	<i>School/study clubs</i>	<i>EGMA</i>		
Transition	Household	HH survey		
Intermediate outcome 1: attendance	School	<i>e.g. school register, spot checks</i>		
Intermediate outcomes 2:				
Intermediate outcomes 3:				
Intermediate outcomes 4:				
Intermediate outcomes 5:				

### 5.2.1 Sustainability (for guidance see Chapter 7) (1-2 pages)

Sustainability will be measured at three levels – school, community, and system – against a Sustainability Scorecard (ratings 1-4, for details please see Chapter 7 of the MEL guidance). Ratings will be determined by the External Evaluator at each evaluation point, based on progress against the indicators you chose, and the qualitative, quantitative, and financial data provided to support such progress.

Please describe what sources, both qualitative and quantitative, you will use to verify progress against your indicators for each level, and where measurement will take place. Please note, qualitative analysis will form a vital part of your sustainability rating. Where appropriate, add the source for financial verification. The sources chosen must correspond logically to your chosen indicators for each level and support these. Please explain the rationale for your

chosen approach to each level of sustainability and the verification sources chosen, and how frequently you will collect data.

*Table 3: Sustainability outcome for measurement*

Sustainability Level	Where will measurement take place?	What source of measurement/ verification will you use?	Rationale – clarify how you will use your qualitative analysis to support your chosen indicators.	Frequency of data collection
School	Household/School	e.g. Household Survey, FGDs		
Community				
System				

### 5.3 Ethical protocols *(for guidance see Chapter 12)* (1-2 pages)



#### 5.3.1 Child protection

Describe the project's approach to child protection throughout the different stages of the evaluation process. This should include reference to both your project's safeguarding standards but also how you will work with the external evaluator and any sub-contractors to ensure children are protected. Please also highlight key risks that may emerge as part of your evaluation activities and your mitigating strategies, including additional safeguards you plan to implement for particular groups of children if any.

#### 5.3.2 Ethics

Describe the approach to research ethics through the different evaluation stages. This may include how you plan to ensure meaningful inclusion of different groups of participants, safety for participants and researchers, ethical considerations with regards to your quantitative and qualitative data collection methods, your planned approach to consent, and ethical considerations in relation to report production and dissemination.

## 6. Sampling framework *(for guidance see Chapters 13-18)*

A completed Sampling Framework template should be submitted alongside the MEL Framework. See Appendix D for the template and guidance on how to complete it. The information in this template should be consistent with the details provided in this section.

### 6.1 Target groups *(for guidance see Chapter 9)* (1-2 pages)



This section should:

- Describe how you are defining marginalised girls and how you will identify your beneficiaries for both learning and transition.
- Provide detailed lists of intervention locations and schools (see Mapping guidance and Sampling Framework template) and expected challenges to transition.

- Outline the criteria used to select your sample from both the treatment and control groups to ensure a sample that can be relied on to generate statistically significant results.

#### 4.2 Control groups / Counterfactual scenario (for guidance see Chapter and 13) (1-2 pages)

Explain how you have identified and selected your comparison group, e.g. matching techniques. Consider potential challenges (e.g. outdated government population data; security implications of collecting such information; other interventions external to your project). How will you mitigate against these issues? If your design is based on a quasi-experimental technique, describe how you will use counterfactual data to measure additionality. Discuss how possible contamination has been accounted for in the evaluation design.

#### 4.3 Cohort tracking (for guidance see Chapters 14, 15, and 16) (2-3 pages)

Describe the (treatment and control) cohorts you will be tracking longitudinally for both learning and transition. How are they defined and how closely do they relate to your identified sampling framework? How will these cohorts be tracked successfully throughout the duration of the project?

##### 4.3.1 Learning cohort (for guidance see Chapter 14)

How will the learning cohort be tracked across grades, particularly when they are moving from primary to secondary school? How will learning beneficiaries of non-school based interventions be tracked? If your project works with out of school girls, how will these girls be tracked successfully in the control group? See *Chapter 5 of the Evaluation Guidance*.

##### 4.3.2 Transition cohort (for guidance see Chapter 15)

If tracking from school to the household at baseline, describe how this will be carried out successfully. What are the possible challenges with this approach and what mechanisms will be put in place to ensure this tracking can be done successfully? If you are following a separate learning and transition sample, how will the transition sample be selected to ensure only those respondents that could be potential beneficiaries are surveyed? Also, how will the two cohorts be matched to identify overlaps between the two samples? See *Chapter 6 of the Evaluation Guidance*.

##### 4.3.3 Replacement strategy (for guidance see Chapters 14, 15, and 16)

Outline your replacement strategy for girls from the original cohort who cannot be found at subsequent evaluation points. How will replacement be carried out in a way which maintains the integrity of the original sample.

#### 4.4 Power calculations and sample sizes (for guidance see Chapter 13) (1-2 pages)

Sample sizes should be provided for both the learning and transition cohort. This section should:

- Be specific about how you have defined your overall minimum sample size (including details of any power calculations used and use of clustering; see *section 13.6 of the Evaluation Guidance*), your planned split between treatment and control samples and your sampling points.

- Consider response rates and attrition and be clear about any assumptions and evidence behind these.
- Provide sample sizes split by grades (in school and out of school) and cohorts if applicable.

#### 4.5 Benchmarking *(for guidance see Chapters 14.2 and 15.3)* (2 pages)

This section should explain how data will be collected to benchmark learning scores and transition rates (*see Chapter 6*). This should include:

- Details of how girls will be selected for benchmarking.
- Which grades will be tested/surveyed at baseline.
- Approximate sample sizes for higher grades that must be benchmarked for subsequent evaluation points.

#### 7. Baseline study *(for guidance see Chapters 13-17)* (1-3 pages)

In addition to the information provided in earlier sections which relates to all evaluation points, given the baseline is key for selecting the initial sample any additional details relevant specifically to baseline data collection should be included in this section. For example, if pre-baseline data will be collected to help select a sample full details of the approach should be provided here.

#### 8. Evaluation governance *(for guidance see Chapters Error! Reference source not found., 11, and 12)* (1-2 pages)

##### 8.1 Evaluation steering group

This section should set out how the whole evaluation will be governed, including details of any partners or organisations who will be involved in the process (*see section X on best practice*).

##### 8.2 External evaluator

This section should state whether the project plans to use their evaluator from GEC-1 or procure a new evaluator via competitive tender. If the GEC-1 evaluator is to be re-contracted for the next phase of GEC provide details of why this is felt to be the best approach, and how you have ensured they can meet the revised requirements of the next phase. Where a new evaluator will be recruited, please outline your procurement approach. A ToR template is provided in annex C of the evaluation guidance. You may wish to include a draft of the completed ToR as an annex to your M&E framework.

##### 8.3 Data validation

Include a description of your plans to ensure that your evaluation will be undertaken in an independent and impartial way. Provide details for how you will ensure data collection is transparent, and how you will validate its veracity.

#### 9. Data quality assurance *(for guidance see Chapter 3)* (1-2 pages)

##### 9.1 Training

What training will be provided to evaluation staff and enumerators to ensure the data is collected correctly and consistently, and that ethical procedures are followed.

##### 9.2 Piloting

How will tools and data collection approaches be piloted to test they work correctly. How will learnings/findings from pilots be fed back into the design.

### 9.3 Data cleaning and editing

What steps will be carried out to check, clean and anonymise data before it is shared with the FM?

## 10. Risks and risk management (1-2 pages)

This section should set out a risk assessment of the potential risks that could impact on your ability to undertake and complete a robust and rigorous MEL strategy. For each risk a suitable mitigation strategy should also be developed. The assessment of risks should include an accurate assessment of the impact of the risk and the likelihood of it arising. Identified risks can be either external or internal to the project.

This information should be filled out in Table 4 below.

Table 4: Risks and mitigations

Potential risks	Probability of risk occurring over the course of the project	Potential impact on project's success	Proposed actions to mitigate risks that have both significant probability and impact/importance
	<i>Low/medium/high</i>	<i>Low/medium/high</i>	

## 11. Learning (for guidance see Section *Error! Reference source not found.*) (2 pages)

### 11.1 Learning strategy

Describe your approach and plans to learning internally and externally. Describe what GEC Learning Clusters you will contribute to and how.

### 11.2 Stakeholder engagement, dissemination and influencing

Describe your plans for stakeholder engagement in the context of your MEL functions. Consider the preferred means, times and protocols for communicating with stakeholders, disseminating information to them, and influencing their action.

## 12. Evaluation workplan (1-2 pages)

### 12.1 Timetable

Please provide a timetable of key MEL activities throughout the lifetime of the project, including procurement of an external evaluator, tool development and testing, pre-baseline data collection if relevant, data collection at the various evaluation points and analysis and report writing.

School term dates and key examination dates should also be highlighted where relevant.

### 12.2 Responsibilities

This section should describe the key MEL responsibilities of all parties involved, and should clearly state who has the overall responsibility for each aspect.

**Annexes:**

- Logframe
- Draft evaluation tools (if already available)
- Completed ToR for evaluators
- Draft Sampling Framework
- any other relevant details



# Appendix C: Terms of reference for external evaluator

## *Standard Terms of Reference template*

Recipient note: the purpose of this ToR is to provide Recipients with a standard template for commissioning evaluation services post development of the MEL framework draft. The recipient may choose to include their MEL framework draft as an annex (making it clear that this draft may be subject to revisions) or insert information from the framework as required. This template should be adapted and tailored according to the specific needs and priorities of the project by deleting /revising and adding content as appropriate.

Evaluation services for the [Recipient to insert name of project], implemented by [Recipient to insert own name].

### *Background to the GEC Programme and the Project*

GEC programme background:

- The Department for International Development (DFID) leads the UK's work to end extreme poverty. DFID is tackling the global challenges of our time including poverty and disease, mass migration, insecurity and conflict. DFID's work is building a safer, healthier, more prosperous world for people in developing countries and in the UK too.
- DFID is working to reach the Sustainable Development Goals (SDGs) by 2030. Progress on girls' education is critical to the achievement of these targets. SDGs 4 and 5 specifically relate to education and achieving gender parity. SDG 4 specifically notes '*inclusive and quality education for all and promote lifelong learning*'.
- Globally 31 million primary age girls, have never been to school<sup>28</sup>. And the majority of these girls come from the poorest and most marginalised communities in the most disadvantaged locations, ethnic groups etc.<sup>29</sup> Over the last 20 years primary enrolments for girls have improved along with boys but completion rates are equally low for both sexes. At the secondary level the differences between boys and girls participation rates really start to show. Significant disparities exist within countries, with the poorest girls from rural areas most severely subject to educational disadvantage - even at the primary level<sup>30</sup>.
- The Girls' Education Challenge (GEC) is helping the world's poorest girls improve their lives through education and supporting better ways of getting girls in school and ensuring they receive a quality of education to transform their future.
- PricewaterhouseCoopers LLP (PwC) and alliance partners have been contracted as the dedicated Fund Manager (FM) and is responsible for the day-to-day operation of the GEC. This includes establishing the recipient tendering process, supporting bidders, sifting and scoring proposals, monitoring Value for Money (VfM) and making project funding recommendations for DFID approval. The FM also manages the relationships with the selected projects and oversees their Monitoring, Evaluation, and Learning operations.

<sup>28</sup> United Nations, 2015. *The World's Women 2015: Trends and Statistics*. New York: United Nations, Department of Economic and Social Affairs, Statistics Division. Sales No. E.15.XVII.8.

<sup>29</sup> Idem

<sup>30</sup> Idem

- Through the GEC, DFID provided £355m between 2012 and 2017 to the FM to disburse to 37 individual projects across 18 countries across sub-Saharan Africa and South Asia to help girl's education. In 2016 the GEC Transition window has been set up with additional DFID funding to support the original GEC beneficiaries continue their journey through stages of education and further improve their learning<sup>31</sup>.

Project background:

This should cover project context, aims and objectives, beneficiaries, outcomes and activities. This information could be lifted directly from section 1 of the MEL framework template.

1. Overview of the project budget and implementation timescales (including number of evaluation points required for the project): [Recipient to insert]

Recipient note: The background section should be short and concise and focused on the information that a bidder would need to know to develop an appropriate methodology and budget for the evaluation assignment. If this requires a long description of the background or if there is other information that would be useful for a bidder to know, consider attaching this as an annex to the ToR.

### *Rationale for the Evaluation*

2. The findings from the evaluation will primarily be used:
  - By the project management team, project partners and stakeholders to inform improvements in the delivery of the project during its lifetime;
  - to demonstrate accountability for the funding received to DFID, other UK Government Departments, UK tax-payers, UK media;
  - by the project management team to leverage additional resources from existing and new partners and stakeholders in order to scale-up and sustain the activities /benefits delivered by the project;
  - by the project management team to support the on-going development and implementation of the project's sustainability and succession strategies;
  - by partners, stakeholders and the Government to learn lessons from the project for the purpose of replicating what works elsewhere and/or taking up approaches and activities that have proven to work in order to scale up the project;
  - by the Fund Manager to feed into and identify insights in order to inform programme level questions; and
  - by other donors, academic institutions and education networks to inform the wider policy debate concerning the education of girls and marginalised girls.

### *Evaluation Objective*

3. The project is seeking to procure the services of an independent External Evaluator to conduct a **mixed-method**, gender-sensitive evaluation that is inclusive of persons with disabilities of the [Recipient to insert name] project over the next [X] years. The evaluation will assess the delivery, effectiveness, VfM and impact of the project and report the findings and lessons learnt throughout the process.

### *Evaluation Questions*

---

<sup>31</sup> <https://www.gov.uk/international-development-funding/girls-education-challenge#overview>

4. The Evaluation Team will be required to develop an evaluation approach that answers the following overarching questions as a minimum:
  - *Process* – Was the project successfully designed and implemented?
  - *Impact* – What impact did the project have on the learning and transition of marginalised girls, including girls with disabilities? How and why was this impact achieved?
  - *VfM* – Did the project demonstrate a good VfM approach?
  - *Effectiveness* – What worked (and did not work) to increase the learning and transition of marginalised girls as defined by the project.
  - *Sustainability* – How sustainable were the activities funded by the GEC and was the project successful in leveraging additional interest and investment?

Specific project and programme level evaluation questions are outlined in *Section 2* of the MEL Framework (*or lift text from this section*). These questions help define the scope and focus of the project evaluation process. The successful bidder will be expected to work with the Project Management Team to review and revise these questions as appropriate at the outset of the project. Project specific context is important in this respect.

### *Overall Evaluation Approach*

5. The overall evaluation approach requires the Evaluation Team to design, plan and conduct a mixed-methods evaluation that is longitudinal in nature. More details on evaluation approach can be found in *section X* of the FM's Evaluation Guidance.
6. A proportionate amount of time and resources should be allocated to the evaluation given the type of project interventions, operational context and the reporting requirements of the GEC.

### *Research design*

7. Comparison groups: bidders are required to outline their approach to evaluating the impact of the project. This should include consideration of the most rigorous approach to establishing a counterfactual. This should enable **comparison** of the outcomes achieved by a target group who were affected by a project intervention with the outcomes achieved by a group who are similar in every way to the target group, except that they have not in any way been exposed to or affected by the project intervention i.e. a comparison group. Careful consideration should be given to the use of experimental or quasi-experimental methods for this purpose.
8. Cohort tracking: the project is required to track a learning cohort and a transition cohort – defined as a group of individuals who progress through life (community or school) together. Bidders should outline their approach to tracking these cohorts in both the control and intervention areas. See section 5 of the Evaluation Guidance for more information on cohort tracking.
9. Measuring outcomes: bidders are expected to understand the projects key and intermediate outcomes and suggest the most appropriate data collection approach to evaluate each outcome. This should include a mixture of quantitative and qualitative approaches. Refer to the Evaluation Handbook and section 3.2 of the MEL Framework. The Evaluator will be expected to pilot tools that will be used for data collection and refine as necessary.

10. Project sampling framework: The Evaluation Team will be required to help finalise the sampling frameworks for both qualitative and quantitative samples. These should be of a sufficient size and representativeness to allow:

- reasonable levels of certainty that the findings are representative for the target population;
- reasonable ability to generalise the intervention's effectiveness to similar contexts; and
- reasonable ability to generalise the insights into what works and why for similar contexts.

Refer to the Evaluation Guidance for further information on sampling and section 12 of the MEL framework.

11. Baseline Study: The Evaluation Team will be required to design and implement a gender-sensitive mixed method baseline study as an integrated part of the overall MEL strategy and plan for the project. This may include pre-baseline data collection to identify the target group and barriers to education. The baseline study should identify the number of beneficiaries with disabilities as well as the type and severity of their disability, following the UN Washington Group methodology<sup>32</sup>. Bidders should set out their approach to the baseline study. See section 5 of the MEL Framework.

#### *Ethical protocols*

12. The evaluation approach must consider the safety of participants and especially children at all stages of the evaluation. The evaluation team will need to demonstrate how they have considered the protection of children through the different evaluation stages, including recruitment and training of research staff, data collection and data analysis and report writing.

13. Research ethics plan: bidders are required to set out their approach to ensuring complete compliance with international good practice with regards to research ethics and protocols particularly with regards to safeguarding children, vulnerable groups (including people with disabilities) and those in fragile and conflict affected states. Consideration should be given to:

- administrative, technical and physical safeguards to protect the confidentiality of those participating in research;
- physical safeguards for those conducting research;
- data protection and secure maintenance procedures for personal information;
- parental consent concerning data collection from children or collation of data about children;
- age- and ability-appropriate assent processes based on reasonable assumptions about comprehension for the ages of children and the disabilities they intend to involve in the research; and
- age-appropriate participation of children, including in the development of data collection tools.

#### *Risk and risk management*

14. Risk management plan: It is important that the successful bidder has taken all reasonable measures to mitigate any potential risk to the delivery of the required outputs for this

---

<sup>32</sup> <http://www.washingtongroup-disability.com/>

evaluation. Therefore, bidders should submit a comprehensive risk management plan covering:

- the assumptions underpinning the successful completion of the proposals submitted and the anticipated challenges that might be faced;
- estimates of the level of risk for each risk identified;
- proposed contingency plans that the bidder will put in place to mitigate against any occurrence of each of the identified risk;
- specific child protection risks and mitigating strategies, including reference to the child protection policy and procedures that will be in place; and
- health and safety issues that may require significant duty of care precautions.

### *Data quality assurance*

15. Quality assurance plan: bidders are required to submit a quality assurance plan that sets out the systems and processes for quality assuring the evaluation and research process and deliverables from start to finish of the project. This plan should include the proposed approaches to:

- Piloting of all research activities;
- Training of enumerators and researchers conducting the mixed-methods primary research, including in research ethics;
- Logistical and management planning;
- Field work protocols and data verification including back-checking and quality control by supervisors; and
- Data cleaning and editing before any analysis.

### *Existing Information Sources*

16. In the first instance, bidders should refer to the DFID GEC website: <https://www.gov.uk/guidance/girls-education-challenge> for general information concerning the Girls' Education Challenge.

17. Bidders should refer to the following *GEC programme* documentation:

- Grant Recipient Handbook
- Evaluation Guidance
- Logframe and workplan guidance

18. Bidders should refer to the following *GEC project* documentation that includes:

- Project logframe;
- Project Full Application as included in the Accountable Grant Arrangement; and
- Projects MEL framework.

19. Bidders should also refer to relevant country data and information that is currently available, as required, to prepare the proposal.

### *Professional Skills and Qualifications*

20. Qualifications: bidders are required to clearly identify and provide CVs for all those proposed in the Evaluation Team, clearly stating their roles and responsibilities for this

evaluation. Please note that if the enumeration is to be sub-contracted, the evaluator will be ultimately responsible for the enumerators they are subcontracting to.

The proposed evaluation person / team should include the technical expertise and practical experience required to deliver the scope of work and evaluation outputs, in particular, with regards to:

- Evaluation design: the team should include skills and expertise required to design, plan and conduct mixed-method impact evaluation, potentially using experimental or quasi-experimental techniques;
- Skills in quantitative and qualitative data collection and analysis, drawing findings from multiple sources and handling potential contradictions between data sets.
- Relevant subject matter knowledge and experience: knowledge and experience required on conducting research with children, the education sector, disability and gender to ensure that the evaluation design and research methods are as relevant and meaningful as possible given the aims and objectives of the project and the context in which it is being delivered;
- Evaluation management: manage a potentially large-scale and complex evaluation and research process from end-to-end, including conducting and reporting a baseline study and final project evaluation report
- Primary research: gender-sensitive design, management and implementation of primary quantitative and qualitative research in potentially challenging project environments, such as fragile and conflict affected states – this could include the design of longitudinal household panel surveys, EGRA /EGMA tests, in-depth interviews, focus groups, etc.;
- Country experience: it is particularly important that the team has the appropriate country knowledge /experience and language proficiency required to conduct the research required;
- Information management: design and manage sex- and disability-disaggregated data and information systems capable of handling large datasets for MEL purposes;
- Statistical analysis: a range of statistical modelling and analysis of impact data; highly proficient user of: SPSS or STATA; and qualitative data analysis techniques, including the use of software e.g. ATLAS.ti, NVivo or equivalent where needed;
- VfM assessment of education projects: education economics expertise to conduct cost benefit analysis and cost effectiveness analysis as part of the assessment of the project's VfM; and
- Safety considerations: Ensuring the whole evaluation process adhere to best practice for research with children including the implementation of child protection policy and procedures to ensure safety of participants. Note that all bidders are expected to be able to show that they have a child protection policy in place to safeguard children that the research team would come into contact with through the research activities.

21. Day-to-day project management of the evaluation will be the responsibility of [Recipient to insert name and position of responsible person].

#### *Deliverables and Schedule*

22. Project deliverables: the main deliverables for this project are as follows:

- Inception report: setting out the design of the MEL strategy and plan and associated planning, logistics, quality assurance, child protection measures and risk management information including gender analysis.
- Baseline study report: design, conduct and submit a baseline study that describes the initial conditions (before the start of the project) against which progress can be



measured or comparisons made to show the effects and impacts of the project in the final project evaluation report. A final report structure will be provided by the FM.

- Midline project evaluation report: design, conduct and submit a midline evaluation report that assesses the effectiveness, impact and VfM of the project at the midline point.
- [If applicable] Third project evaluation report: design, conduct and submit a third project evaluation report that assesses the effectiveness, impact and VfM of the project.
- Final project evaluation report: design, conduct and submit a final project evaluation report that assesses the effectiveness, impact and VfM of the project.

23. Report requirements: all reports should be submitted in electronic form and should be submitted in English.

The Evaluation Team will be required to provide face-to-face presentations in-country of all deliverables as an integral part of the submission process.

The Evaluation Team will be expected to provide a fully 'cleaned-up' dataset in SPSS, Stata or SAS file format accompanied by the code used to carry out analysis and a variable codebook.

24. Detailed work plan: bidders are required to provide a detailed work plan incorporating all relevant tasks and milestones from start to finish of the evaluation study.

25. Project milestones: bidders are required to include in their detailed work plans the milestones set out below.

Typical project milestones /outputs for deliverables	Deadlines
Invitation to tender sent out to bidders	Recipient to insert date
Deadline for receipt of tenders	Recipient to insert date
Evaluation of tenders and shortlisting completed	Recipient to insert date
Interviews of shortlisted suppliers held	Recipient to insert date
Supplier appointed	Recipient to insert date
1. Inception Phase	Deadlines
Inception Meeting held	Recipient to insert date
Literature/document review & data gathering completed	Bidder to complete
Review of project's theory of change, impact logic and evaluability completed	Bidder to complete
Stakeholder consultation completed	Bidder to complete
Child protection framework developed	Bidder to complete
Sampling framework for primary research for baseline completed	Bidder to complete
Design of data collection strategy including cohort tracking design completed	Bidder to complete
Design of primary research instruments for baseline completed	Bidder to complete
Draft Inception Report (including design of baseline study) submitted for review and comments by Project Manager and Project Partners.	Bidder to complete
Presentation to Evaluation Steering Group	Bidder to complete
Review complete and comments returned to supplier	Bidder to complete
Final Inception Report submitted	Recipient to insert date
2. Baseline Study Phase	Deadlines
Tool development and piloting	Bidder to complete
Baseline research starts	Bidder to complete
Baseline research completed	Bidder to complete
Draft Baseline Study Report submitted for review	Bidder to complete



Presentation to Evaluation Steering Group	Bidder to complete
Review by Project Management and stakeholders completed	Bidder to complete
/comments provided to Supplier	
Supplier addresses comments and revises Baseline Study Report	Bidder to complete
Final Baseline Study Report submitted	Recipient to insert date
<b>3. Subsequent Project Evaluation Phases</b>	<b>Deadlines</b>
<b>3.1 Start of Design Review Phase</b>	<b>Bidder to complete</b>
Preliminary review of project information and data completed	Bidder to complete
Review of evaluation design and research methods completed	Bidder to complete
Revisions to evaluation design and research methods completed	Bidder to complete
Review of sampling framework for primary research completed	Bidder to complete
Review of primary research instruments for primary research completed	Bidder to complete
Draft Research Design Report reviewed by Project Manager, Evaluation Steering Group, etc completed and comments returned to supplier	Bidder to complete
Final Research Design Report submitted	Bidder to complete
<b>3.2 Start of Research Phase</b>	<b>Bidder to complete</b>
Analysis of financial and monitoring data completed	Bidder to complete
Analysis of cohort tracking data completed	Bidder to complete
Primary quantitative research starts (e.g. household surveys, school surveys)	Bidder to complete
Primary quantitative research ends	Bidder to complete
Primary qualitative research starts (e.g. EGRA, EGMA, focus groups, workshops, semi-structured interviews of stakeholders /partners)	Bidder to complete
Primary qualitative research ends	Bidder to complete
Data verification, cleaning and validation completed	Bidder to complete
<b>3.3 Start of Analysis Phase</b>	<b>Bidder to complete</b>
Start of analysis phase	Bidder to complete
Analysis of data and results completed	Bidder to complete
Draft Interim Report submitted	Bidder to complete
Presentation to Evaluation Steering Group	Bidder to complete
Draft Interim (Emerging Findings) Report reviewed by Project Manager, Evaluation Steering Group, etc and comments returned to supplier	Bidder to complete
Final Interim (Emerging Findings) Report submitted	Bidder to complete
<b>3.4 Start of Reporting Phase</b>	<b>Bidder to complete</b>
Draft Final Project Evaluation Report submitted	Bidder to complete
Presentation to Evaluation Steering Group	Bidder to complete
Draft Final Project Evaluation Report reviewed by Project Manager, Evaluation Steering Group, etc and comments returned to supplier	Bidder to complete
Final Project Evaluation Report submitted	Bidder to complete
Final Project Evaluation Report agreed	Recipient to insert date

### *Reporting and Contracting Arrangements*

26. The Evaluation Team will be expected to identify a Project Director and Project Manager for communication and reporting purposes. At the Inception meeting the Evaluation Team Project Manager will be expected to submit a full contact list of all those involved in the evaluation.
27. The Evaluation Team will be expected to attend report to the Evaluation Steering Group and attend all meetings as agreed with the Project Evaluation Manager. The Team will be required to submit to the Project Evaluation Manager bi-weekly progress reports (by email) during the study periods summarising activities /tasks completed to date (per cent achieved), time spent etc.

### *Budget*

28. The estimated budget for this work is [Recipient to specify]. This budget should cover the data collection, analysis and reporting for [x number] evaluation points. This budget is inclusive of all costs covering team member costs, travel, research costs and any other costs associated the completion of the work including where required costs for reasonable adjustment. Bidders are required to organise and fund their own duty of care arrangements as required.
29. Bidders are required to provide a fully costed proposal in the form of a price schedule that as a minimum should include:
  - Sub-total of fees for the delivery of any task or deliverable;
  - Sub-total for number of days per partner organisation (as applicable);
  - Expenses and overheads broken down by the project cost categories [Recipient to provide];
  - Reasonable adjustment costs; and
  - Total costs before and after any taxes that are applicable.

Bidders are required to provide a payment schedule on the basis of milestone payments for the successful delivery of each deliverable.

Recipient note: if costs are required to be submitted on a time and materials basis then the above minimum could also include:

- Study team inputs – broken down by the number of days for each individual study team member against each of the tasks set out in the detailed work plan;
- Day rates for each study team member;
- Total number of days per team member; and
- Total fees per team member.

GEC Recipient practice note: whilst omitting budget information could incentivise cost-effective ways to meet the requirements of the evaluation, the more likely scenario is that without a budget estimate, bidders will find it difficult to determine how much and what kinds of data should be collected for the evaluation. Therefore, we recommend that bidders for an evaluation are provided with an estimate of the budget available to undertake the work. To ensure VfM, consider requiring bidders to describe what they can expect to achieve within this envelope and to provide a breakdown of costs and expenses, including number of days and daily fee rate.



## **Appendix D: Sampling framework template**

Please refer to the first tab of the sampling framework template for details.

## Appendix E: Payment by results (PbR)

As in the first phase of the GEC programme, a PbR approach will be applied to the GEC-T phase, with some significant differences. The main changes include that PbR will now be linked to learning outcomes for **all projects** within the window, if they are able to measure learning outcome performance using comparison groups. In addition, PbR is to be restructured as a bonus mechanism, responding to findings around the effectiveness of the PbR 'downside' in the first phase of the GEC.

### Payment by Results (PbR): Core principles

**Any project that cannot for some reason conduct comparison groups and therefore cannot be on PbR needs to have this agreed with the Fund Manager at contracting stage. Any decisions to remove a project from the comparison group requirement during the implementation period will need a contractual amendment and agreement with the Fund Manager and DFID.**

1. PbR linked to learning outcomes will be applied to all projects in the GEC-T funding window unless the evaluation is unable to use a rigorous experimental method with comparison groups. Exemptions need to be agreed with the Fund Manager.
2. PbR will no longer involve quarterly financial retentions, and there will be therefore no PbR 'downside' associated with not achieving outcome targets. Quarterly payments following requests for funding will continue to be released on the basis of achievement of agreed milestones.
3. PbR decisions on bonus payments will be based on the average rate of literacy and numeracy improvement against agreed targets.
4. Literacy and numeracy targets will be 0.25 standard deviations (using the difference-in-difference methodology) per year of implementation. Performance against these targets will form the basis of PbR payments.
5. The optional third learning outcome, transition outcome, attendance outcome and other intermediate outcomes will not be components of PbR.
6. PbR payments will be paid for projects achieving above their target for the average of literacy and numeracy performance.
7. PbR payments will be paid only for statistically significant results.
8. Payments will be scaled linearly and applied to the PbR proportion of budgets linked to PbR.
9. Projects will have the equivalent of 10% of their expenditure linked to the PbR bonus, and this provides the upper-end of the bonus.
10. The performance cap linked to this expenditure will be 500% of literacy and numeracy targets.

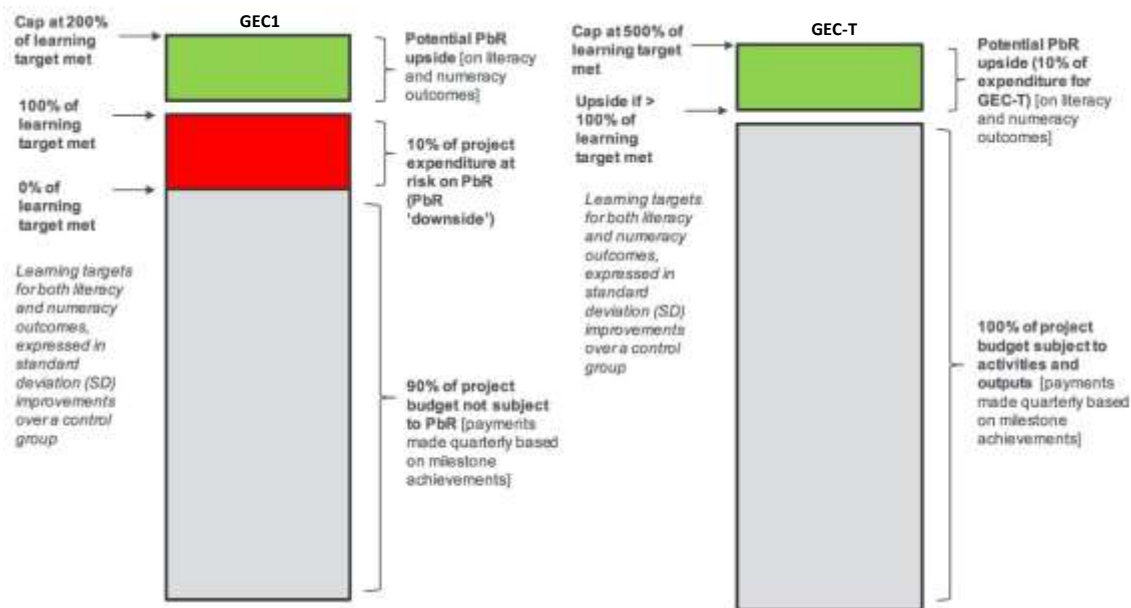
PbR payments will be aligned to the following principles:

- Payments are linked directly to the evaluation, and as such would be paid after the evaluation points beyond baseline.
- Payments to be made based on expenditure incurred up to the reporting quarter in which the evaluation report was submitted.

- The distribution of any PbR payments among consortium partners must be agreed by the Lead Organisation and their partners. It is advised that this is done at the start of the contract.

### **The PbR model for GEC-T**

The PbR approach for GEC-T is an evolution of the approach from the first phase of the GEC programme. As shown in the diagram below, the GEC approach in the first phase involved a downside for projects achieving between 0% and 100% of their learning targets, and an upside for projects achieving above 100% of their learning targets. The new approach shown on the right of the diagram loses the PbR 'downside', and only involves an upside for projects that over-achieve their targets.



The PbR approach will be compulsory for all Recipients, unless they do not use a comparison group for the evaluation of learning outcomes. Exemptions on comparison groups will need to be agreed with the Fund Manager. The exclusion of non-comparison group evaluations is to ensure that PbR payments can only be made on rigorous, statistically significant outcome performance. Details of the model are below.

### **GEC-T PbR model: 0.25 standard deviation target, 10% of expenditure as upside cap, 500% achievement cap**

- Maximum PbR bonus proportion = The equivalent of 10% of project expenditure in the period
- Learning target = 0.25 standard deviations
- Learning target cap = 1.25 standard deviations (500% achievement)
- Bonuses above learning target, i.e. >0.25 standard deviations (>100% achievement)

### **Outcome results indicators**

The assessment performed by the Recipient's external evaluator will be the basis for determining whether or not targeted outcome results are achieved. Within the project Monitoring, Evaluation and Learning (MEL) framework, the Recipient will need to identify how

the external evaluator will be able to provide the Fund Manager with assurance as to whether targets have been met. This will involve assessing a representative sample of the target beneficiary population and then scaling these results to the entire population. Recipients are required to include their sampling methodology and sampling framework within their overall MEL framework. Sample sizes should at a minimum be statistically significant on the achievement of PbR learning targets.

Results achieved at the second evaluation point (the baseline being the first evaluation point) will serve as the effective starting point for the third evaluation point to be assessed against. In exceptional circumstances, targets may be amended.

### ***Setting targets using outcomes spreadsheet***

The outcomes spreadsheet is the key reporting tool for all projects for their Learning and Transition outcomes. Using the outcomes spreadsheet, targets will be set on a project-by-project basis for the GEC outcome indicators.

The Fund Manager requires that targets for learning outcome indicators be set on the basis of a 0.25 standard deviation effect size per year of implementation. This effect size is over and above the performance of a comparison group, so it is the additional amount of learning achieved. The target for a particular group of girls in a grade should be derived using the distribution of test scores of the girls in the grades above (one year above for each year until the next evaluation point).

Recipients need to submit the outcomes spreadsheet after baseline data collection and agree all outcome indicator targets for subsequent evaluation points with the Fund Manager.

### ***PbR exceptions***

Note that for some projects in exceptional circumstances it can be agreed with the Fund Manager to not be on PbR. Where Recipients have such an agreement, they are still expected to fully comply with all evaluation principles as if they were on PbR.

### ***Conditions for overpayments due to overachievement of results***

Any funds paid to the Recipient for overachievement of results are governed by the following restrictions.

- Funds must be spent in line with the International Development Act 2002.<sup>33</sup> The International Development Act 2002 is the legal authority for DFID expenditure and gives the authority to spend money through a number of different "powers", including:
  - Provision of development assistance which contributes to poverty reduction (the so-called "core" power)
  - Provision of development assistance to the UK Overseas Territories
  - Provision of humanitarian assistance
  - Contributions to multilateral development banks.
- Funds must be spent within three years from the Recipient's GEC project end date.
- The use of funds must be disclosed to DFID within six months of expenditure of all funds. Disclosure requirements will be communicated to the Recipient should their organisation receive funds for the overachievement of results.

---

<sup>33</sup> <http://www.legislation.gov.uk/ukpga/2002/1/contents>



It is strongly encouraged that Recipients reinvest any funds received for the overachievement of results against which payment is made into their GEC project. If this is not possible, it is strongly encouraged that the funds are spent in a country where GEC is operational for the betterment of marginalised girls and their communities.

# Appendix F: Gender Analysis Guidance and Framework

***The enclosed gender analysis template must be completed and submitted to the FM along with your V2 logframe and V1 MEL framework, by 14 July 2017 as it should be used to inform both project design and the development of projects' MEL frameworks.***

## Background

One of the contracting requirements for GEC-T is that all GEC-T Recipients must conduct a gender analysis to inform their project design in the changeover period.

This Guidance Note includes provides further information on this requirement including:

1. The GEC's Minimum Standards on Gender Equality and Social Inclusion.
2. Description of gender analyses, the rationale behind the need for a discrete gender analysis and how it will be used in the GEC.
3. A template with a set of guiding questions, along with suggested additional frameworks for conducting a gender analysis.

## Why is gender equality important for GEC-T?

The GEC was designed to provide girls with an opportunity to transform their lives through access to quality education, acknowledging that gender inequality can be a driver for the challenges faced by millions of school-aged girls. While some issues such as lack of schools and poorly skilled teachers affect all children's learning in a given context, other barriers to education apply *specifically* to girls. In order for girls' learning to improve, these gender-specific barriers must be understood and this understanding must be fed into the design and delivery of projects. Without this analysis, projects not only risk being less effective<sup>34</sup>, they also risk doing harm.

The contribution of education projects to improving the lives of girls is limited when projects work with girls in isolation. Simply including girls or focusing on girls in project interventions does not mean gender inequality is automatically addressed - girls are firmly located within a social context and are part of peer groups, families, schools and communities. An explicit analysis of these dimensions of girls' lives is critically important for project design and implementation.

The GEC has developed a set of Minimum Standards around Gender Equality and Social Inclusion (GESI) that all GEC Recipients must adhere to. These Minimum Standards have been developed to ensure that GEC-T projects design and deliver interventions that target highly marginalised girls and ensure their inclusion in education provision. In order to do so, the gendered factors that lead to girls' educational marginalisation must be understood.

## Gender Equality and Social Inclusion: The GEC Minimum Standards

---

<sup>34</sup> Projects need to be able to respond to the challenge of improving both the conditions and environments for learning.

1. A gender analysis of the context is conducted and used to inform the project's final design and Theory of Change.
2. The logframe includes gender-sensitive and disability focused quantitative and qualitative indicators.
3. Bi-annual reporting includes reflections on i) progress towards meeting gender transformative standards (further guidance forthcoming), ii) to what extent activities identified and addressed barriers to inclusion and opportunities for participation for people with disabilities.
4. Monitoring and evaluation processes include and differentiate girls from a variety of sub groups, including those with disabilities, from the start of the project. This data should track girls' experiences and whether interventions are responding to their needs.
5. A retention strategy that captures the reasons for girls' drop-out from school and provides appropriate support to re-engage girls in response to the common issues is articulated in project activities.
6. Do no Harm, Child Protection and risk analyses are informed by a gender equality and social inclusion lens.
7. Sex, age and disability disaggregated data is collected and analysed at baseline, midline and endline.
8. Disability data differentiates between the type and severity of disability of beneficiaries.
9. The project is resourced with staff, partners and contractors who have appropriate gender and social inclusion expertise.
10. Lesson learning and sharing of best practice captures achievement towards i) gender equitable and transformative outcomes and ii) the inclusion and participation in planning, implementation and M&E of people with disabilities.

### What is a gender analysis?

Gender analysis is a methodology designed for examining the differences in roles, responsibilities and norms for women and men, girls and boys. It looks at the different levels of power that people hold as well as the differences in their needs, recognising how these may vary from region to region. A gender analysis is essential when gender is a key component of a project's activities or outcomes.

Understanding how gendered barriers act to affect girls' and boys' access to education and the quality of education delivered to them is critical for good project design, and ultimately to improve a project's ability to achieve results.

A gender analysis should help project teams to understand the differences in power, and the use of power, including violence, to uphold gendered privileges. In the education context, these dynamics ultimately impact on boys' and girls' learning and life outcomes. Discriminatory gender practices and lower value placed on girls in many contexts contribute to lower expectations, aspirations and, consequently, lower outcomes for girls in education.

This kind of analysis helps to ensure that project design and planning, implementation, monitoring and evaluation take into account existing gender disparities, and that the interventions designed to remove or reduce them are the right ones for the project context.

### How should the gender analysis be used?

A gender analysis is a prerequisite for good project design, and is critical for a robust Theory of Change. By identifying differences of gender, a gender analysis allows project teams to

anticipate the ways in which their interventions will impact on gender roles, relations or responsibilities. It will also highlight areas where improvements can be made, allowing the team to deliver the project in a way that responds to the needs of their particular context. ***The findings from the gender analysis should therefore be used by projects to inform and refine final project design, make adjustments to the Theory of Change, complete the project logframe and draft the MEL framework.*** GEC-T projects' gender analysis will also inform project monitoring and will be used by the FM to guide discussions in GEC-T's Annual Review meetings, designed to assess project progress on an annual basis.

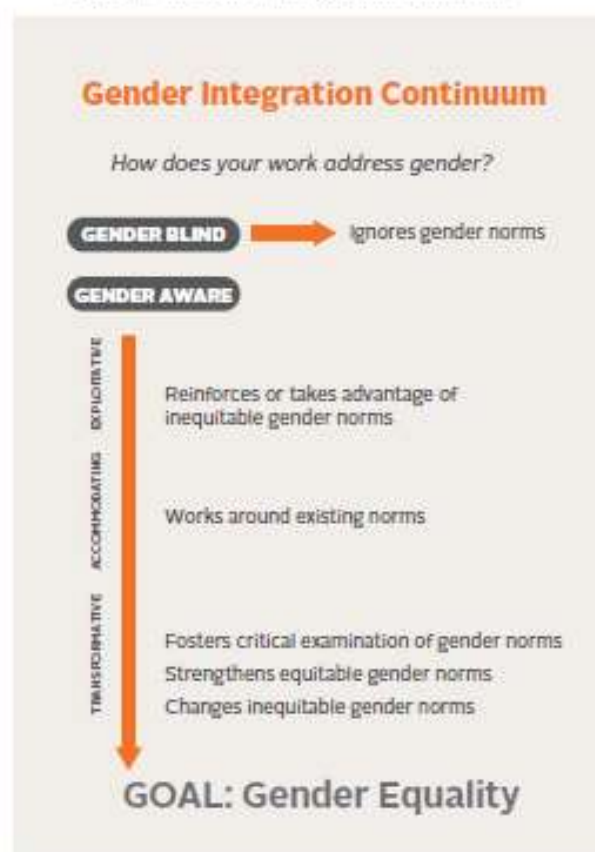
### How should the GEC-T's gender analysis be developed?

The gender analysis template should be completed using secondary data on the country-context in which the project will be implemented. An in-house gender specialist should undertake the development of the gender analysis. Where internal capacity is not available, projects will need to contract a consultant experienced in conducting gender analyses, audits and/or strategies; however, please note there will not be additional funds to do so.

As a minimum, projects should complete the GEC gender analysis template using the core questions presented, as these draw from various analytical frameworks in use, as well as the GEC's own framework on educational marginalisation.

Projects can also enhance their gender analysis by engaging more deeply in one or more of the commonly known frameworks in use for their gender analysis. Some of the most commonly known ones are the Harvard Analytical Framework, Moser Framework, Gender Analysis Matrix, Capacities and Vulnerabilities Analysis Framework, and Social Relations Approach. The Human Rights Based Approach is used for education equity analysis. More recent work to develop analysis around girls' rights (see the work of the Population Council and Plan's Because I am a Girl report series and its references), as well as intersectionality analysis – understanding how intersecting forms of inequality include disability status, age/life stage, ethnicity - could also provide useful references.

Figure 1. The Gender Integration Continuum<sup>14</sup>



Projects are also encouraged to consider their analysis in relation to the Gender Integration Continuum developed by FHI360 (illustrated in Figure 1 below) and to include some reflection of their design in relation to this framework. The gender analysis should include some detail on where the project feels it currently sits on the Gender Integration Continuum, and its ambition in terms of project progress. ***All projects should aim for a gender transformative approach where safe and appropriate to do so.***

In this approach, gender stereotypes and norms are challenged and the project seeks to transform unequal power relations between boys and girls through changes in roles, status and through the redistribution of resources. The response is more likely to focus on girls' strategic needs.

Gender Integration Continuum (FHI360)	Description
Gender blind	The issue of gender is <b>not addressed</b> . Sex disaggregated data is not used.
Gender aware	Sex disaggregated data is used in the analysis but the <b>response doesn't address the differential gendered needs</b> of girls and boys.
Gender accommodating	Gender issues are acknowledged and specific responses designed to address the needs and concerns of girls and boys are included in the project activities and outcomes. The response is more likely to focus on <b>girls' practical needs</b> .
Gender transformative	Gender stereotypes and norms are challenged and the project seeks to transform unequal power relations between boys and girls through changes in roles, status and through the redistribution of resources. The response is more likely to focus on <b>girls' strategic needs</b> .

### What does the GEC expect from gender transformative approach?

A transformative approach to education programming involves improving marginalised girls' access to quality education, as well as supporting them, their families, schools, and communities to understand and challenge the social and gender norms that perpetuate the inequalities that affect marginalised girls. This ambition of progress towards gender equality moves beyond access to school/learning, and requires that girls and boys also experience the same levels of quality and outcomes of education.

Inequalities between girls and boys affect their educational opportunities and outcomes at every level. Traditional expectations and norms around girls' choices and behaviour can determine whether girls get access to the classroom in the first place, limit the time they have available for learning, undermine the confidence they have in certain subjects, and the degree to which they participate in lessons, and ultimately shape their future aspirations. They also influence the perceived value of girls' education among others. As girls get older, the gendered norms they are under pressure to conform to become more pronounced and the opportunities they have to learn often contract. Girls may be expected instead to get married, have children, take on greater caring and domestic responsibilities, and contribute to family income.

Without a clear approach to strategically address these issues, attendance and learning outcomes for girls will decline through lower and upper secondary and drop-out rates will increase.

---

## GEC-T Projects Gender Analysis

May 2017

**Purpose:** Gender analysis is a methodology designed for examining the differences in roles, responsibilities and norms for women and men, girls and boys. It examines the different levels of power that people hold as well as the differences in their needs. The gender analysis supports projects to ensure that project design and planning, implementation, monitoring and evaluation take into account existing gender disparities, and that the interventions designed to remove or reduce them are the right ones for the project context.

**Project Name:**

**Lead Partner:**

**Completed by (name and role):**

**Date:**



<b>1. Individual/Household/Community Level – Attitudes, Beliefs and Norms</b>	<i>Potential data sources: contextual reports on gender-specific (girls') education from academics/NGOs; national government statistics agencies, UN agencies – particularly UNICEF, UNFPA and UN Women; your project team's reports; your project's GEC evaluations; other projects' GEC evaluations (including qualitative data)</i>
What are the socio-cultural norms (beliefs and perceptions) and the practices (behaviours) that may be contributing to girls' ability to enrol in school, stay in school, and perform?	
<ul style="list-style-type: none"> <li>Do parents and other adults have different <b>expectations of girls' and boys' academic performance</b>? What are girls' and boys' own aspirations?</li> </ul>	
<ul style="list-style-type: none"> <li>What is the <b>household division of labour</b> like in the context in which you are working? How does this affect girls' time use and their ability to enrol in, attend school regularly and use their out-of-school time for activities like homework and extracurricular activities?</li> </ul>	
<ul style="list-style-type: none"> <li>What are the <b>aspirations and expectations</b> of girls and adolescents? What are the expectations around marriage as girls</li> </ul>	

get older? How are these different to those for adolescent boys? How are these norms around girls' behaviours and <b>life choices</b> likely to affect girls' ability to enrol in and attend school regularly?	
<b>2. School Level – Teaching and Learning</b>	<i>Potential data sources: contextual reports on gender-specific (girls') education from academics/NGOs; your project team's reports; your project's GEC evaluations; other projects' GEC evaluations</i>
How well does the school environment support girls' and boys' enrolment, attendance, retention and performance? Does it support girls and boys differently?	
○ Are there differences in girls' and boys' <b>enrolment, attendance and retention rates</b> ?	
○ What is the <b>quality of teaching</b> like in the context in which you work? How does the quality of teaching affect girls and boys differently?	
○ What are teachers' perceptions of girls' and boys' <b>academic potential and performance</b> ?	

<ul style="list-style-type: none"> <li>○ Is the school environment safe and conducive to learning? <ul style="list-style-type: none"> <li>▪ What are the perceptions of girls and boys of the school environment – is it considered to be a <b>safe space in which to learn</b>? Is there a difference in their perceptions?</li> <li>▪ Do girls and boys <b>experience violence</b> in and around schools differently?</li> <li>▪ Do the <b>pedagogical approaches</b> in use engender a learning environment that is conducive to girls' learning?</li> <li>▪ Do girls and boys <b>participate equally</b> in the classroom?</li> <li>▪ Do teachers use pedagogical approaches that encourage more <b>student participation</b>?</li> </ul> </li> </ul>	
<p><b>System Level – Laws and Policies</b></p>	<p><i>Potential data sources: EMIS; Gender Parity Index; SDG 4 Reporting (from 2017 onwards)</i></p>
<p>What is the policy environment like for girls' education in the context in which you are working?</p>	
<ul style="list-style-type: none"> <li>○ Does the government have a supportive policy framework for progress towards <b>gender equality in education</b>, and for promoting girls' education?</li> </ul>	

<ul style="list-style-type: none"> <li>○ What is the government policy around <b>mothers and pregnant girls</b>/young women attending school?</li> </ul>	
<ul style="list-style-type: none"> <li>○ Does the government have a supportive policy framework in place, where the lack of <b>female teachers</b> has been identified as a barrier to learning?</li> </ul>	

Based on your analysis above, what are the key project interventions which seek to address girls' **practical** and **strategic** needs to achieve education in this context? (Potential data sources: *contextual reports on gender-specific (girls') education from academics/NGOs; your project team's reports; your project's GEC evaluations; other projects' GEC evaluations*)

<b>Practical Needs-</b> <i>Short-term, immediate, practical assistance for girls according to perceived need to assist their learning progress</i>

<b>Strategic needs</b> – Interventions that will challenge existing gender roles with the goal of achieving gender equality. They seek to transform girls' status and role in the home, community and school