EVALUATION QUALITY ASSESSMENT FRAMEWORK

for AE-Led Evaluations

Commissioner: Independent Evaluation Unit GREEN CLIMATE FUND
Author: DeftEdge
Date: June 2025

CONTENTS

INTRODUCTION **2**

Overview **2**

FRAMEWORK DEVELOPMENT **2**

Context **2**

GCF Evaluation Requirements **3**

EQA Processes of Other Agencies **7**

GCF/IEU EVALUATION ASSESSMENT SYSTEM **11**

EQA Tool **11**

Assessment Components **11**

Review Process **13**

Summary Report **13**

REVIEWER GUIDANCE **14**

ANNEXES **17**

Annex 1: Overall Structure of GCF's Integrated Results Management Framework (IRMF) **17**

Annex 2: GCF Evaluation Criteria Definitions **18**

Annex 3: Assessing Logical Framework and ToC Quality **20**

Annex 4: Key Resources for Conducting EQAs **21**

Annex 5: EQA Tool **22**

***** Unless otherwise noted, all tables and figures in this document are produced by DeftEdge.

INTRODUCTION

Overview

This Green Climate Fund (GCF) Independent Evaluation Unit (IEU) Evaluation Quality Assessment (EQA) Framework for evaluations led by Accredited Entities (AEs) is designed for both accountability and learning purposes. It reflects GCF's unique mandate in climate finance, its focus on catalyzing a paradigm shift towards low-emission and climate-resilient development pathways, and progress in respect to its outcome indicators.

Evaluations are individually and collectively assessed by independent reviewers on the extent to which they meet GCF Evaluation Standards[1] (2022) and relevant expectations set out in the GCF Evaluation Policy and other guidance documents[2], as well as United Nations Evaluation Group (UNEG) standards. Ratings are accompanied by narrative feedback on key strengths and areas for improvement for individual evaluations and the portfolio as a whole. The IEU will use the EQA results to further tailor its guidance and support to AEs.

This document is designed to orient GCF staff, AEs, evaluation managers and evaluators to GCF's quality assessment process. It has three sections. The first covers the relevant GCF policies and procedures as well as quality assessment systems used by other international organizations that informed the development of this framework. The next section describes GCF's assessment system, including the components of the EQA Tool used to rate evaluation reports and how reviews are conducted. The final section provides guidance for reviewers. The EQA Tool is provided in the annexes.

FRAMEWORK DEVELOPMENT

Context

As per the GCF Evaluation Policy, the IEU is requested to "perform quality assurance upon request by the board"[3] to ensure that evaluations led by AEs are effectively implementing the expectations of the Policy. As part of these efforts, IEU has introduced a process for the independent assessment of AE-led evaluations to enhance credibility, consistency, and overall quality of these evaluations.

The IEU is mandated by the Board of the GCF to inform its decision-making and in doing so, has an accountability and learning function. The quality assessment process is consistent with the IEU's principal aims which are to (a) undertake and deliver high quality evaluations, (b) build and deliver an evaluation-based learning, advisory and capacity strengthening program, and (c) engage strategically to learn, share and adopt best practices in the climate change evaluation space.

---

[1] Green Climate Fund (2022). Green Climate Fund Evaluation Standards. Songdo, South Korea https://ieu.greenclimate.fund/sites/default/files/document/220428-gcf-evalluation-standards-web.pdf

[2] Including the GCF Monitoring and Accountability Framework for Accredited Entities, the Evaluation Policy for the GCF (2021), the GCF Integrated Results Management Framework (2020), the GCF Results Handbook for IRMF (draft, 2022), and the GCF Evaluation Guidelines (2023).

[3] Green Climate Fund (2021). Evaluation Policy for the GCF, p. 7. Songdo, South Korea https://ieu.greenclimate.fund/sites/default/files/document/evaluation-policy-gcf.pdf pg 7.

Under the GCF's Evaluation Policy and the Monitoring and Accountability Framework (MAF), AEs must submit interim and final evaluations[4] for all funded activities. Before these evaluations are formally accepted, the GCF Secretariat conducts quality assurance of the interim and final evaluation reports and provides feedback to AEs in cases where improvements are needed to meet GCF evaluation standards. This internal evaluation **quality assurance process** is supplemented by **quality assessment** which is the purview of the IEU. The latter involves rating an evaluation report ex-post according to pre-determined parameters and is typically carried out by an independent firm[5].

The EQA framework and assessment tool were developed by the IEU and DeftEdge[6], a firm with significant experience in building and implementing evaluation quality assessment processes in the UN system[7].

GCF Evaluation Requirements

This section reviews expectations for evaluations that are stipulated in key GCF documents and highlights those that go beyond standard UNEG requirements.

The **GCF Monitoring and Accountability Framework for Accredited Entities** makes it clear that interim and final evaluations are required for each funded activity, and that these evaluations should also assess performance against the GCF investment framework criteria, including financial/economic performances as part of the efficiency and/or effective ness criterion (see Section 1.2(b)). Accordingly, the EQA Tool incorporates the need for financial management of projects to be assessed.



The **Evaluation Policy for the GCF (2021)** states that the theory of change for the evaluation function is based on the rationale that

> *"trusted, high-quality evidence produced from credible evaluations helps to inform GCF investments, policies, structure, performance, processes and strategies by informing and guiding the Fund for its day-to-day operations and providing strategic guidance to the Board, Secretariat, independent units and AEs. This, in turn, ensures GCF investments have greater impact and that they are expected to contribute to building a healthier planet"* (para 13, p.3).

The policy emphasizes that evaluations are to serve both accountability and learning purposes, and the need for

[4] According to the GCF Evaluation Policy the interim evaluation is the evaluation performed midway through the implementation of an intervention to assess progress towards and likelihood of achievement of outcomes and impacts. It usually has a strong formative focus. A final evaluation that is near or at the end of an intervention to provide evaluative evidence covering the entire intervention. It measures the overall impact, effectiveness, efficiency, sustainability, replicability and lessons learned of a project. (GCF Evaluation Policy, pg 18-19)

[5] As per UNFPA's Evaluation Quality Assurance and Assessment system, "Quality assurance takes place throughout each phase of the evaluation, and quality assessment occurs after an evaluation is completed. An external independent reviewer assesses the final evaluation report against established quality standards and criteria."

[6] DeftEdge is contracted under a Long-Term Agreement with GCF to provide evaluation capacity building services to support the IEU's accountability and learning function.

[7] DeftEdge has conducted evaluation quality assessment for UNFPA, UNICEF, UNPBF, UNODC and ILO.

upholding the principles of (a) impartiality, objectivity, and unbiased, (b) relevance, use and participation, (c) credibility and robustness (including use of mixed methods and benchmarking), and (d) measurability of performance. The issues are all reflected in the EQA Tool.

The Policy also identifies GCF's evaluation criteria, and these are shown in Table 1. Recognizing that the GCF requires evaluations to include more criteria than most other UN entities, the IEU has determined that not all need to be addressed individually and that it is acceptable for some to be integrated into other criteria. The EQA Tool considers that the level of emphasis on each should be determined by the type of project, evaluation scope, and type of evaluation, with only final evaluations having to address Innovativeness and Replicability & Scalability.

Table 1 GCF Evaluation Criteria

Relevance, Effectiveness, Efficiency, Impact and Sustainability of projects and programs
Coherence in climate finance delivery with other multilateral entities
Gender Equity
Country Ownership of projects and programs
Innovativeness in result areas – the extent to which interventions may lead to paradigm shift towards low-emission and climate-resilient development pathways
Replication & Scalability – the extent to which the activities can be scaled up in other locations within the country or replicated in other countries
Unexpected Results, both positive and negative

The EQA Tool also considers evaluation budgets as the Policy underscores the need for interim and final evaluations to be adequately resourced. Evaluation budgets "consistent with Global evaluation international best practices, should range from 2-5 percent of the project budget" (para 41, p.10).

The **GCF Evaluation Standards (2022)** document sets out 15 overarching standards for evaluations, which are shown in Table 2. It also sets out accompanying indicators. The stated objective of these Standards is to "support and enable the production of state-of-the- art evaluations with high-quality evidence and recommendations" (para 2, p.2). They "are intended to ensure that the Fund can inform its overall performance – including results, impacts and unintended consequences – in a credible and measurable manner" (para 6, p. 3).

Table 2 GCF Evaluation Standards

1. Independence	9. Confidentiality
2. Impartiality and Objectivity	10. Cost-effectiveness
3. Utility and Value Added	11. Ethics
4. Ownership and Participation	12. Integrity
5. Credibility and Rigour	13. Accountability
6. Transparency	14. Competence
7. Learning	15. Respect and Beneficence
8. Human Rights, Gender Equality and Environmental Considerations	

The Standards align closely with UNEG expectations for evaluations, including the UNEG Ethical Guidelines, and emphasize principles of good evaluation practice, such as the importance of participatory processes and the need to avoid conflicts of interests. However, some GCF Standards go further than UNEG - such as the need for evaluations to include a dissemination plan - and their reach extends beyond the conduct of evaluations to include how evaluations are planned for, commissioned and managed.

The EQA Framework incorporates a review of the extent to which evaluations meet GCF Standards. However, the tool does not provide a comprehensive assessment as it is limited to information that can reasonably be expected to be found in an evaluation report and its Terms of Reference (ToR), if attached. It does accept that the inclusion of the UNEG Pledge of Ethical Conduct in Evaluation signed by evaluators is sufficient for addressing relevant indicators.

The Standards document notes that "not every standard would be applicable to every evaluation [but that] the evaluator and those commissioning the evaluation will have to explain why a standard was not relevant for a particular evaluation" (para 10). This is particularly relevant for Standard 8, Human Rights, Gender Equality, and Environmental Considerations which covers also the rights of indigenous peoples. The EQA tool provides some flexibility in the breadth and depth to which these are covered although all evaluations must assess the application of Environment and Social Safeguards (ESS) and Gender Policy and Indigenous Peoples Policy.

The **GCF Integrated Results Management Framework Policy (2021)** is highly relevant for assessing evaluation quality as it defines the agency's top-level theory of change and eight results areas. The policy states that all GCF-funded projects/programs need to monitor and report for three levels of results:

1. **GCF impact level**: Paradigm shift potential for which progress is tracked on three dimensions (scale, replicability, sustainability) using a defined scorecard approach.
2. **GCF outcome level:**
    a. **Reduced Emissions and Increased Resilience** for which there are four quantitative indicators:
        i. GHG emissions reduced, avoided or removed/sequestered
        ii. Direct and indirect beneficiaries reached
        iii. Value of physical assets made more resilient to the effects of climate change and/or more able to reduce GHG emissions
        iv. Hectares of natural resource areas brought under improved low-emission and/or climate-resilient management practices.
    b. **Enabling Environment** for which there are also four indicators (qualitative and quantitative) that, using a scorecard approach, measure the degree to which GCF investments contribute to:
        i. Strengthening institutional and regulatory frameworks for low-emission climate-resilient development pathways in a country-driven manner
        ii. Technology deployment, dissemination, development or transfer and innovation
        iii. Market development/transformation at the sectoral, local or national level
        iv. Effective knowledge generation and learning processes, and use of good practices, methodologies and standards.
3. **Project/programme-level outcomes and outputs** that are defined in the approved logframe.

The latter must be addressed in all evaluations. Although all evaluations are also encouraged to provide an analysis of the higher-level GCF results, it is recognized that most AE projects are at a more modest scale that makes it difficult to show contribution. Therefore, the EQA Tool requires that only final evaluations need to address progress towards paradigm shift (Impact) and progress towards/achievement of Reduced Emission and Increased Resilience (GCF Outcomes as relevant to the project), with the expected depth of analysis scaled to the type of project. In addition, the use of paradigm-shift scorecards is only required for projects approved from 2022. As per the IRMF, "enabling environment is an outcome-level indicator evaluated using a three-point scorecard (low/medium/high) during interim and final evaluations of GCF-funded mitigation and adaptation projects/programmes".

Annex 1 includes diagrams of the overall structure of the IRMF and GCF's eight results areas.

The **GCF Evaluation Guidelines (2023)** provides more comprehensive expectations for how AEs should evaluate GCF-supported interventions to meet the requirement of GCF's Evaluation Policy. The document contains guidance on how to plan, scope, commission and manage AE-led evaluations. Relevant to quality assessment, it distinguishes the difference in focus of interim and final evaluations, noting that: *"Although all criteria need to be covered in all cases, the importance of each varies between interim and final evaluations. AE-led evaluations should apply proportionality and prioritize properly when identifying evaluation questions for each criterion to ensure that evaluation resources are used efficiently".* The Guidelines also define all GCF criteria and explain how the quality of logical frameworks and Theories of Change (ToCs) should be assessed – this information is attached as Annex 2 and 3 to this document.

The Guidelines include a grid outlining the GCF Secretariat's Evaluation Quality Assurance and Review Framework for AE-Led Evaluations. It has four sections, as bulleted below, and within each are multiple considerations.

- Impartiality, Objectivity and Bias Mitigation
- Relevance, Use and Participation
- Credibility and Robustness
- Structure and Clarity of the Evaluation

The grid maps to UNEG guidelines but is less comprehensive than the EQA Tool and does not capture all of the expectations covered by the Guidelines including processes specific to evaluating GCF's outcome-level results.

The **GCF Results Handbook for IRMF (draft 2022)** provides useful guidance for evaluations although as it has not been formally approved, its procedures are not yet formally required[8]. It notes that in addition to the GCF evaluation policy and the AE's own evaluative requirements, the Evaluation Manager is responsible for ensuring that the evaluation ToR incorporates all the IRMF requirements, namely:

- Scorecard assessment of progress towards paradigm shift
- Scorecard assessment of progress against enabling environment indicators
- Assurance/validation that agreed IRMF-related monitoring methodologies and processes are being applied, and are generating robust data

[8] As per the Board decision on the IRMF (decision B.29/01), the GCF Secretariat continues to gather comments and feedback from GCF stakeholders on the draft results handbook for improvement and before being considered for approval by the GCF Board.

The latter point is of relevance to the current EQA Framework as the Tool reflects the need for evaluators to validate reported results for GCF outcomes. It is insufficient to rely solely on project monitoring data. EQA reviewers look at whether this information has been triangulated – e.g. by a subject-matter expert, benchmarking against internationally accepted standards, or statistical analysis - to confirm its accuracy.

The Handbook also points out that, while the AE has given the evaluators considerable freedom to develop their own methodology, the approach should take into account – and ensure delivery of – the GCF evaluation policy and the IRMF requirements. It also has suggestions for how participatory processes, a key part of the GCF Evaluation Standards, can be used to assess progress on the dimensions set out in the paradigm shift and enabling environment scorecards.

EQA Processes of Other Agencies

An important part of developing the EQA Framework involved reviewing the evaluation quality assessment and evaluation report validation processes used by the evaluation functions of other international organizations including UNICEF, UNDP, UNFPA, UNODC, UNEP, FAO, World Bank (IEG), ADB, and GEF, as well as another fund, UN Peacebuilding Fund (UNPBF). Key insights from this review are shown in the following table.

Table 3 Approaches to Reviewing Evaluation Quality in the UN System

Consistency Across Organizations	Most international organizations, including UNFPA, FAO, World Bank, and UNDP, follow a structured EQA process that aligns with international standards such as UNEG and OECD-DAC. This goes some way towards having a consistent approach to evaluations that are credible and comparable across different sectors and regions. Like GCF, both GEF and UNEP have additional criteria beyond OECD-DAC that evaluators are required to incorporate into their analysis, and performance ratings are required for each.
Weighted Criteria Systems	Organizations including UNFPA, UNPBF and FAO emphasize weighted criteria in their EQA processes and place the most emphasis on findings/analysis and methodology. This ensures that evaluations are robust, with critical aspects like data reliability and analysis receiving appropriate emphasis. UNEP and GEF do not use a weighted approach.
Integration of Cross-Cutting Issues	Many organizations, especially UNICEF and UNFPA, have integrated cross-cutting issues such as gender equality, human rights, and environmental sustainability into their evaluation frameworks. This not only aligns with global priorities but also ensures that evaluations are holistic and consider broader impacts. Like GCF, both GEF and UNEP capture these aspects as 'Environmental and Social Safeguards, and Gender'. However, unlike most other UN entities, GCF does not assess AE-led evaluations according to UN-SWAP requirements for Gender Equality and the Empowerment of Women (GEEW), as these evaluations would have already undergone such an assessment by their respective commissioners. GCF does have its own Gender Policy and Indigenous Peoples Policy that are to be applied at the evaluation design stage.
Use of Peer Reviews and Expert Panels	Organizations such as FAO, ADB, and the World Bank place a strong emphasis on peer reviews and the involvement of subject-matter experts (who may be external). These processes are noted as enhancing the objectivity and technical accuracy of evaluations, ensuring that the findings are credible and actionable. They are typically led by the evaluation office and are sometimes applied at earlier stages of the evaluation cycle (for the terms of reference, inception and draft reports), in which case they would be considered quality assurance. FAO uses peer review for draft reports. The World Bank appears to use peer reviews, and occasionally expert panels, to a greater extent for research-related publications.

Evaluation Report Validation	Similar to the above, GEF IEO conducts a validation process for a sample of terminal evaluations that are either prepared or independently reviewed by evaluation units of the GEF Agencies, and validates all terminal evaluations prepared by the operational units of Agencies that do not undergo an independent review. The validation process entails: a review and synthesis of the evidence on performance of the relevant project or program; a validation of the performance ratings provided in the terminal evaluations; and an assessment of the quality of the terminal evaluation report. The process is conducted by a 3-person team including a peer feedback provider.
Environmental and Social Standards	GEF and UNFPA highlight the importance of integrating environmental and social standards into the evaluation process, particularly in the context of sustainability and climate-related evaluations. This approach ensures that evaluations consider long-term impacts on the environment and society.
Independent Quality Assessment	The UN entities reviewed all use independent external reviewers to conduct their EQA processes with the exception of UNEP whose Evaluation Office conduct the reviews.

The analysis also compared rating scales, criteria that were used to rate evaluations, the weights given to different sections of the EQA tool, the emphasis on quantitative scores versus narrative feedback, among other aspects to the extent this information was available[9].

Given that most of the entities considered in this review adhere to the UNEG norms and standards for evaluations, many of the criteria/types of questions in the EQA tools are similar, although most also have aspects tailored to their respective mandates (e.g., UNFPA assigns a heavier weighting to gender responsiveness and UNPBF includes questions on conflict analysis). More significant differences were seen in the UNDP template as it also rates the quality of the evaluation ToRs. Differences are also apparent in UNEP's template which has considerably more detail on how each section and each component of Findings should be rated, and the overall rating of the report is based on the mean score of each section.

Template sections and weightings: The following table shows the sections and respective weightings of the EQA templates of five other UN entities evaluation functions, and how these compare to those adopted by GCF/IEU. Like others, GCF/IEU's EQA Tool assigns the highest weightings to Findings and then Methodology, emphasizing the importance of well-conceptualized and rigorous process to assess climate-related initiatives.

Table 4 Comparison of Sections and Section Weightings across UN Entity Evaluation Functions

	Section Weightings (%)					
Section	UNICEF /EO	UNODC /IES	UNFPA /EO	UNDP /IEO	UNPBF /IEO	GCF /IEU
ToR				15		
Executive Summary	5	7			5	5
Structure/Exec Summary			7			
Background	5				5	10
Purpose/Scope	5				5	5
Context/Purpose		5				
Design/Methods/Data Reliability	20	20	24	30	25	25
Findings	25	25	40	40	25	30

[9] The current EQA templates of the World Bank (IEG) and ADB were not included as they are not publicly available.

Conclusions/Lessons	10	13	11	15	15	5
Recommendations	15	15	11	10	10	10
Structure/Presentation	5	10		**	5	5
Evaluation Principles (inclusion/safeguards)	5			15		5
GEEW (UN-SWAP)	5	5	7		5	n/a
Total	100	100	100	100	100	100

** UNDP includes Structure with Methodology

The GEF/IEU template is structured differently and has 14 sections (which are called 'criteria'). These map roughly to the sections of the templates used by the above UN agencies as shown in Table 5.

Table 5 Comparison of GEF sections with Other Agencies

GEF Section/Criteria	Extent of Alignment with Other Agency Templates
Timeliness	This is not covered by other templates.
General information	This is typically covered under Background or Context but does not have its own section.
Stakeholder involvement in evaluation	If assessed, this is typically covered under Methodology.
Theory of change	This is typically covered under Context or Findings but does not have its own section.
Methodology	All templates include this section, but typically a more detailed discussion is required.
Outcomes	This is covered under Findings.
Sustainability	This is covered under Findings.
Monitoring and evaluation	This is covered under Findings, typically under Effectiveness.
Finance	N/A for most entities. PBF has a lighter touch assessment under Context.
Implementation	This is covered under Findings (Effectiveness or Efficiency)
Environmental/social safeguards and gender	These issues are typically covered under Findings (in a section on human rights, gender, leave no one behind, disability inclusion, and sometimes environment)
Lessons and recommendations	These typically have their own sections or Lessons are combined with Conclusions.
Performance ratings	N/A except for UNEP.
Report presentation	This is also typically a section in other templates.

Assessing Evaluation Terms of Reference: UNDP is the only comparator agency that was found to rate the quality of the ToR, and the assessment counts for 15% of the overall EQA score. That section has five questions that ask about whether the document appropriately and clearly outlines the purpose, objectives, criteria, methodology and key questions for the evaluation following the UNDP Evaluation Guidelines. GCF/IEU recognizes that the quality of the ToR can significantly influence the quality and usefulness of the evaluation report. The EQA Framework takes the ToR into account in assessing alignment with GCF Evaluation Standards but there is not a rating specific to the ToR.

Table 6 Comparison of Overall Ratings

Overall Report Ratings	%
GCF/IEU	
Exceptional	96-100
Very Good	90-95
Good	75-89
Fair	50-74
Unsatisfactory	<50
UNICEF/EO	
Exceptional	96 - 100
Highly Satisfactory	87.5 - 95.99
Satisfactory	62.5 - 87.49
Fair	35 - 62.49
Poor	0 - 34.99
UNODC/IES	
Very Good	90+
Good	70 - 89
Fair	50 - 69
Unsatisfactory	< 50
UNFPA/EO	
Very Good	85 +
Good	65 - 84
Fair	50 - 64
Unsatisfactory	< 50
UNPBF/IEU	
Very Good	90+
Good	75-89
Fair	50-74
Unsatisfactory	<50

Overall Report Ratings: There is also variation in the overall rating categories for the assessed reports and how these are derived. Table 6 compares what was adopted for GCF with four other UN entities. Like UNICEF and UNFPA[10], GCF/IEU chose to have a narrow top band designed to capture those reports that go beyond meeting the standard assessment criteria.

Narrative Comments: There is a range in the extent of narrative feedback provided. While some templates are more of a checklist, reviewers for both UNICEF and UNDP provide detailed comments that include examples from the report to justify the rating given to each question. Some templates have comments for each section of the template with minimal or no comments at the question level. It is standard practice to have an overall summary that explains the strengths and weaknesses of the report. UNICEF takes this further be having a subsection that provides guidance directed to evaluation managers and evaluators for how improvements can be made. The GCF/IEU EQA Tool has narrative comments for each question and overall summary comments.

Question-level Rating Options: Approaches to this vary. Like UNICEF, GCF/IEU has chosen to use a 4-point scale to rate each question - Yes, Mostly, Partly, No - with the option of Not Rated. Some agencies use a 3-point scale – yes, somewhat, no – with the option of not applicable (UNODC, UNPBF). UNEP and GEF do not rate each question but use a 5-point scale to rate each section of their template (highly satisfactory, satisfactory, mostly satisfactory, unsatisfactory, highly unsatisfactory).

Number of Templates: Of the sampled agencies, only UNODC and UNEP appear to use different templates for rating interim/mid-term and final evaluation reports to accommodate additional expectations for the final evaluations. GCF/IEU uses one tool for all types of evaluations, and it provides an option for reviewers to use 'not rated' for questions that are not applicable, and this is factored into the scoring formula. However, the GCF/IEU tool includes two components.

[10] Note that UNFPA recently changed their categories to Excellent, Highly Satisfactory, Satisfactory, Fair and Unsatisfactory but their template does not show how these are calculated. Table 2 shows their system used prior to 2024.

GCF/IEU EVALUATION ASSESSMENT SYSTEM

EQA Tool

The GCF IEU Framework is distinct from the quality assessment processes of other agencies as it looks beyond generally accepted expectations for evaluation by also assessing the application of GCF's evaluation criteria and adherence to GCF Evaluation Standards. The main part of the Framework is the EQA Tool which was tested on representative, stratified sample of AE-led evaluation reports to confirm its applicability to the range of GCF investments.

The EQA Tool is a rubric for providing quantitative ratings for multiple aspects of each evaluation report as well as qualitative comments that are used to justify the ratings and explain strengths and areas for improvement. The first section has space for report details (e.g., name of AE, year of report, if ToR is included), classification information (e.g., region, project type and size, type of report, ESS category), and other factors (e.g., evaluation complexity, budget). This information enables the analysis of the cumulative results that is prepared for the Summary Report to be disaggregated in multiple ways.

Assessment Components

The assessment part of the Tool has three components:

1. **Adherence to UNEG standards and GCF-specific expectations** (such as coverage of environmental and social safeguards and reporting on climate adaptation and resilience results): This part of the tool has nine sections, each with a series of questions (50 in total) that are rated on a four-point ordinal scale: Yes=1, Mostly=.67, Partly=.33, No=0, with the option for N/A. As shown in Table 7, the sections have different weightings to reflect their relative importance with analysis of findings and methodology weighted the most heavily. The tallied scores provide an overall report rating that ranges from Exceptional to Unsatisfactory as shown below.

Table 7 GCF EQA Tool Sections and Weights

Section	Weight
Executive Summary	5%
Background	10%
Purpose/Scope	5%
Design/Methods	25%
Findings	30%
Conclusions/Lessons	5%
Recommendations	10%
Structure/Presentation	5%
ESS	5%
Total	100%

Table 8 Overall EQA Rating Categories

• • • • •	Exceptional (96-100%)	Exemplifies expectations for GCF evaluation quality
• • • • -	Very Good (90-95%)	Effectively upholds GCF standards for evaluation quality.
• • • - -	Good (75-89%)	Generally meets GCF standards for evaluation quality
• • - - -	Fair (50-74%)	Needs improvements to meet GCF standards for evaluation quality

• - - - -	Unsatisfactory (<50%)	Does not sufficiently meet the GCF standards for evaluation reports

2. **Coverage of GCF's evaluation criteria:** The assessment considers whether all GCF criteria are addressed, with the analysis supplemented by a checklist that provides a visually depiction of the extent to which each is covered. The checklist has four answer options: Yes, Partial, No, Not Applicable. 'Yes' is used when reviewers consider the level of analysis to be adequate and 'Partial' is used when the respective criterion is briefly addressed. "Not Applicable" is used for midterm reviews that do not address the Innovativeness and Replicability/Scalability criteria as these are only required for final evaluations.
3. **Alignment with GCF Evaluation Standards:** Another series of questions (45) address adherence to the 15 Standards. The questions were extracted from the GCF Evaluation Standards and Guidelines documents[11] and reflect the requirements that can reasonably be expected to be found in an evaluation report and/or evaluation ToR. As noted earlier, many elements within the Standards are principles which are not typically addressed in evaluation reports; hence this tool does not provide a comprehensive assessment of the extent to which GCF's Standards are being upheld by AEs. It is particularly limited when the ToR is not provided; in which case, Cannot Rate is used for multiple questions with the other rating options being Yes, Partial, No, and Not Applicable. Despite its limitations, this assessment is still considered useful for showing the relative level of importance that is placed on aspects covered by the Standards by AEs and evaluators, and for extracting good practices that can be shared.

In this case, the questions all carry the same weight. The tallied scores result in a descriptive rating as shown below. Although the Standards Assessment scores do not factor directly into the overall evaluation report score, an evaluation with a Standards score below 70% would not be eligible for an overall report rating of Very Good or Good.

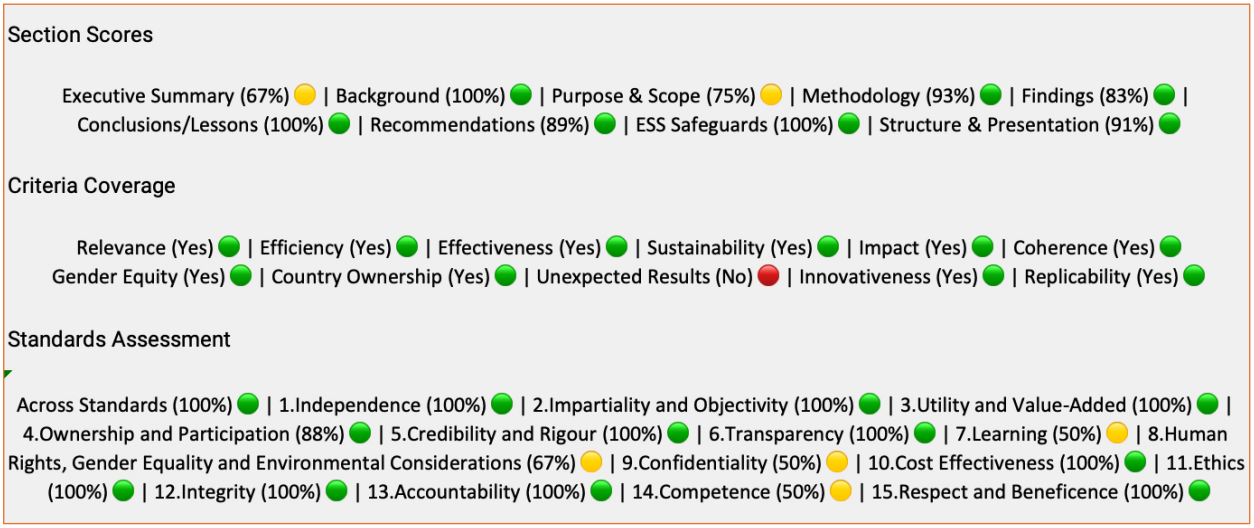
Table 9 Scoring Categories for Alignment with Standards

≥ 85%	Aligns with GCF Evaluation Standards
70–84%	Mostly aligns with GCF Evaluation Standards
50–69%	Needs substantial improvements to align with GCF Evaluation Standards
< 50%	Does not align with GCF Evaluation Standards

The content of the three dimensions of the EQA Tool is provided in Annex 5. However, the actual Tool is in the form of an excel spreadsheet which enables auto-calculation of section-level and overall scores, and the presentation of visual depictions of the section scores, criteria coverage and adherence to each of the 15 Evaluation Standards. The figure below shows the visual results from an assessed evaluation report.

[11] GCF Evaluation Standards https://ieu.greenclimate.fund/sites/default/files/document/220428-gcf-evalluation-standards-web.pdf and Evaluation operational procedures and guidelines for Accredited Entity-led evaluations https://www.greenclimate.fund/document/evaluation-operational-procedures-and-guidelines-accredited-entity-led-evaluations

Figure 1 Visual Depiction of Results in EQA Tool



Review Process

EQAs are undertaken by an independent review team comprised of experienced evaluators. A structured orientation process helps to ensure uniform interpretation of the assessment process. The orientation entails a systematic review of the EQA Tool, examining each main criteria and its corresponding questions and sub-questions in detail, and the alignment of the criteria with GCF Evaluation Policy, Evaluation Standards and Quality Assurance Framework. The session includes an analysis of evaluation report examples to help calibrate reviewer understanding and standardize quality judgments. This deliberate approach to each assessment criterion enables the team to have clear benchmarks and a shared understanding of quality thresholds.

Each evaluation report is carefully read and rated by a reviewer using the EQA Tool; the internal version of which has supplementary guidance that addresses minimum requirements for different levels of ratings. A senior expert conducts a second-level check of the EQA for quality assurance and to ensure reliable and consistent application of the framework. If unresolved issues remain, another full review may be undertaken.

The completed EQAs are then shared with GCF/IEU. Although the review process is independent, the IEU can request clarification or further justification of any of the scoring or comments.

Summary Report

To support the learning objective of the Framework, on a periodic basis, the results of the individual assessments are compiled and presented in an EQA Summary Report. The aggregated analysis examines both quantitative and qualitative dimensions for all reviewed reports. Quantitative analysis systematically assesses scoring patterns across evaluation criteria, with particular attention to GCF priorities including GCF outcomes, environmental and social safeguards, and adherence to Evaluation Standards. Qualitative analysis is drawn from the narrative comments made by the reviewers in the EQA forms.

The report presents overall quality trends across the portfolio, examining performance against key evaluation criteria, and regional and thematic patterns. There are synthesized findings on evaluation practice strengths and weaknesses. The recommendations section provides guidance for strengthening evaluation quality, improving climate-specific assessments, and supporting AE evaluation capacity. The report utilizes data visualization techniques to effectively communicate patterns and trends through various analytical representations.

REVIEWER GUIDANCE

Accountability and Learning Functions

The EQA process serves an important accountability function, helping to ensure that GCF-funded initiatives are thoroughly, credibly and appropriately evaluated. However, equally important is the role that EQAs can play in influencing improved evaluation practice. Reviewers are expected to provide a thorough assessment but in a constructive manner that suggests areas for improvement and also highlights where good practices are evident.

Reviewer Resources

In order to fairly apply the EQA Tool's ratings, in addition to being knowledgeable about UNEG evaluation standards, reviewers need to be familiar with GCF's unique mandate in climate finance, its IRMF and results areas, as well as its particular expectations for evaluation practice and standards. The annexes include a list of key resources, a diagram of the overall IMRF, definitions of all GCF evaluation criteria and standards, and agency-specific guidance on assessing theories of change and results frameworks.

Even with a solid understanding of the requirements, there is a level of subjectivity in scoring a significant number of the questions. The different type, size, and ESS category of projects also call for flexibility in how ratings should be applied. To help ensure inter-rater reliability and consistency, the reviewer-version of the EQA Tool includes guidance for rating each question/sub-question. A 'notes column' in the tool provides space for the reviewer to indicate where a second opinion is needed for any question.

Constructive Feedback

The narrative comments are key to acknowledging and promoting good evaluation practice. General principles for reviewers to keep in mind in structuring comments include the following:

- Write as a peer, putting oneself in the shoes of the evaluator. Aim to use words like "could", "may consider", is "advised to" and avoid using terms such as "failed to" and "should" unless this is warranted by the complete absence of a GCF standard (such as not including the ToR as an annex).
- Feedback should justify the ratings and briefly indicate how the report met or did not meet each question/indicator. It is helpful to cite illustrative examples.
- Go beyond the indicators - feedback must add value and avoid repeating the information from the indicators within the text.
- Be mindful of the project size and risk category and ensure that comments are proportionate to the scale and complexity of the evaluation.
- Acknowledge where the evaluation has clearly faced constraints because of the context, the complexity of the project, unclear ToR, etc. Let the reader know that these may have affected evaluation quality even if the standards require the report to be rated low.
- The summary boxes should provide a concise overview of how the evaluation performed according to the GCF/IEU and UNEG expectations. The first box, which address overall quality, should highlight the topline strengths and weaknesses of each section of the tool, note the methodology used, highlight best practices and any factors that contributed to the complexity of the evaluation process. The second box, which addresses GCF Evaluation Standards, should also highlight topline strengths and weakness, as well as comment on how extent of adherence to the Standards effected overall evaluation quality.

- Suggestions for improvement should be specific to the sections of the report where shortcomings were identified. Where weaknesses are significant, resources should be cited that may be useful for future evaluations. As relevant, suggestions should also be made for increasing methodological rigour.

Use of Key Words

To support the qualitative analysis process conducted for the Summary Report, aim to use the following key words/terms, where applicable, when highlighting report strengths, weaknesses, and examples of good practices that could be useful for organizational learning. It is particularly important to highlight specific strengths of reports rated as Exceptional or Very Good.

Key words: notable – exemplary -- good practice -- key strengths -- areas for improvement – shortcomings

Steps for Conducting an EQA

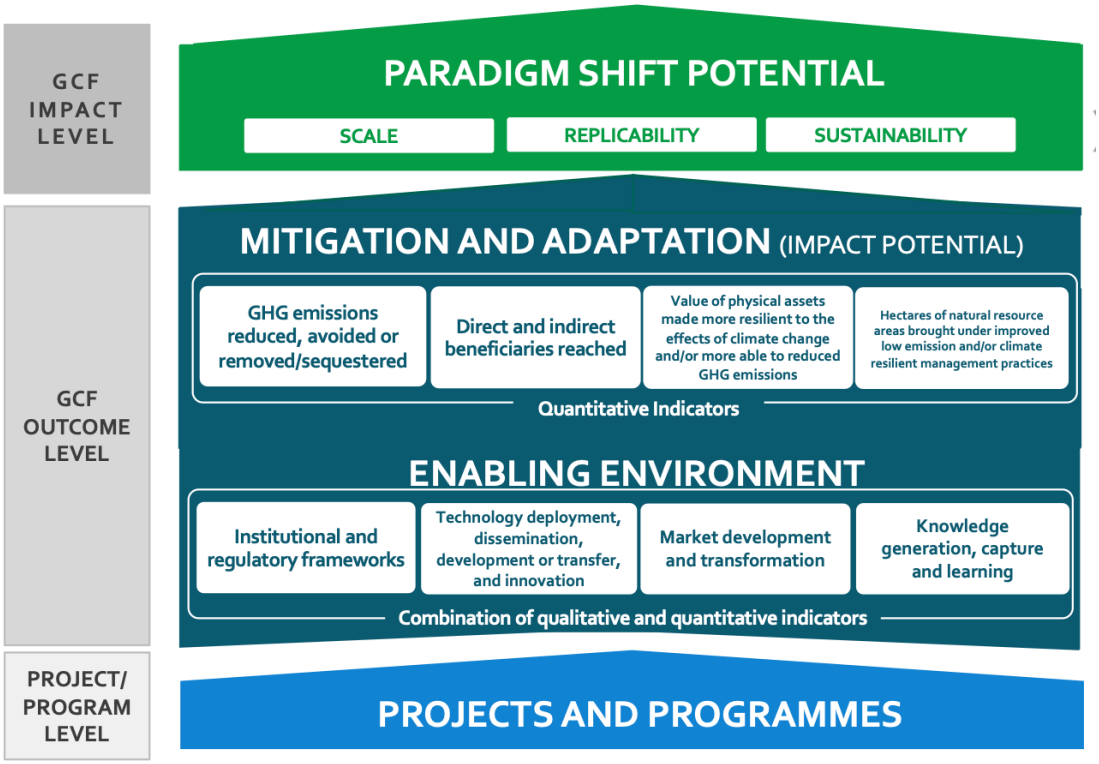
The following figure presents a suggested strategy for undertaking an EQA.

Figure 2 Suggested Steps for Conducting EQAs

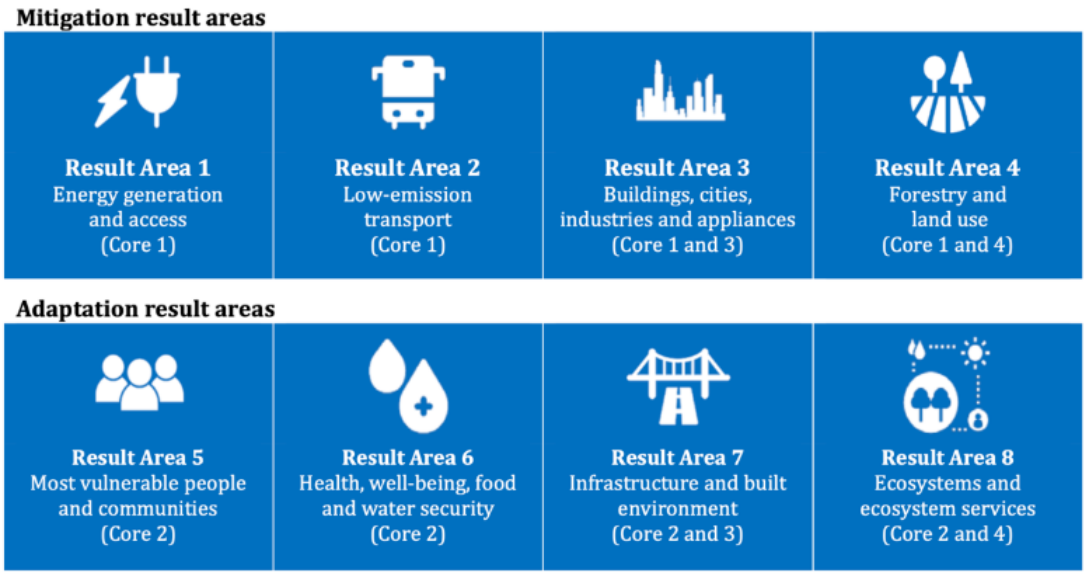


ANNEXES

Annex 1: Overall Structure of GCF's Integrated Results Management Framework (IRMF)



GCF Results Areas with Quantitative Indicators



Source: GCF Integrated Results Management Framework: Results Handbook, 2022.

Annex 2: GCF Evaluation Criteria Definitions

GCF EVALUATION CRITERIA	DEFINITION
RELEVANCE, EFFECTIVENESS, EFFICIENCY, IMPACT AND SUSTAINABILITY OF PROJECTS AND PROGRAMMES	The criteria are, as defined by OECD-DAC:[23] Relevance – the extent to which the project/programme objectives and design respond to beneficiaries', global, country, and partners'/institutions' needs, policies, and priorities, and continue to do so if circumstances change. For GCF-funded projects/programmes, it is important to assess the responsiveness of a project/programme to climate-related needs, and how direct and indirect beneficiaries were identified and targeted are important elements under this criterion. Effectiveness – the extent to which the project/programme achieved, or is expected to achieve, its objectives, and its results, including any differential results across groups. Efficiency – the extent to which the project delivers, or is likely to deliver, results in an economic and timely way. For GCF projects/programmes, the degree to which resources have been optimized to maximize the impact of resources spent on climate adaptation and mitigation results. Impact – the extent to which the project has generated or is expected to generate significant, positive or negative, intended or unintended, higher-level effects. Sustainability – the extent to which the net benefits of the project/programme continue or are likely to continue.
COHERENCE IN CLIMATE FINANCE DELIVERY WITH OTHER MULTILATERAL ENTITIES	• The extent to which the project/programme (i) is compatible with and/or complements (and not duplicate) other climate finance funding[24] and (ii) avoids creating parallel systems through creating synergies in funding mechanisms and interventions.
GENDER EQUITY	The extent to which the project/programme level gender action plans have been implemented, and its activities and outcomes are monitored as well as the extent to which key stakeholders have been consulted, and their inputs were considered and used to inform design and implementation. The extent to which the project/programme has addressed vulnerability to ensure equal participation of men, women, children and youth, indigenous people, people with disability and other marginalized groups as well as benefits from the GCF's investments. The degree to which the project/programme demonstrated understanding on how the impacts of climate change are differentiated by gender and other markers of difference, the ways

	that behavioural changes, gender and other markers of difference can play in delivering a paradigm shift, and the role that women, indigenous people and other groups play in responding to climate change challenges both as agents but also for accountability and decision-making.
COUNTRY OWNERSHIP OF PROJECTS AND PROGRAMMERS	The extent to which the project is consistent with the existing national climate strategy, policies, capacity to deliver of implementing entities, intermediaries or executing entities, and engagement with civil society organizations and other relevant stakeholders. This criterion is strongly related to the Relevance criterion.
INNOVATIVENESS IN RESULT AREAS	The extent to which the project/programme creates opportunities for targeting innovative solutions, new market segments, developing or adopting new technologies, business models, modal shifts and/or processes that are appropriate to a given context and which can bring about a paradigm shift.[25]
REPLICATION AND SCALABILITY	The extent to which the activities are replicated in other countries or other locations within the same country. The extent to which the project/programme has expanded its scale and impact without increasing the total costs of implementation.
UNEXPECTED RESULTS, BOTH POSITIVE AND NEGATIVE	The degree to which unforeseen and unintended results, both positive and negative, and which were not anticipated as part of the planning or articulated in the original theory of change, were captured and from which relevant lessons were identified and utilized to adaptively manage the project/programme and improve project/programme implementation and decision-making.

source: GCF Evaluation Guidelines, 2023

Annex 3: Assessing Logical Framework and ToC Quality

GCF Evaluation Operational Procedures and Guidelines for Accredited Entity-led Evaluations suggest that AEs and the commissioned evaluators should ensure that the following (non-exhaustive) lines of enquiries are considered when checking the quality of the logical framework and theory of change:

Logical Framework

- Are the outcome indicators consistent with the definition and methodology specified in the IRMF? Why/why not?
- Are the results specified at the right level (i.e., outputs do not measure activities, outcomes do not measure outputs, etc.), and co-benefits specified properly?
- If reporting core indicator 1, does the logframe disaggregate Greenhouse Gas (GHG) reporting by technology or result area?
- If reporting on core indicator 2, does the logframe clearly identify beneficiaries and provide disaggregation of the number of people supported by gender and other potentially important levels of disaggregation?
- Are the targets in relation to the baseline appropriate? If the project/programme has overachieved or is overachieving, is it because of efficiency and effectiveness or underestimation during the design stage?
- Is the AE undertaking (or has it undertaken) the means of verification specified in the FAA logframe? Did the expected data sources materialize?
- Are assumptions/risks identified and internalized within the logframe? Are they being monitored?
- Is there a logical progression in the results achieved, e.g., reporting states that outcomes have been overachieved when the outputs have not been met, and if not, has this been explained?

Theory of Change

- Does the theory of change logically present the pathway(s) to the changes that the project/ programme intends to achieve?
- Is it clear how the activities and outputs bring about the adaptation, mitigation or cross-cutting outcomes that relate to GCF result areas as well as the intended co-benefits, and how these outcomes could contribute to the desired contribution to the paradigm shift?
- Is the project/programme implementation consistent with the theory of change? Why/why not?
- What revisions to the theory of change may be required to make it more accurate and reflective of implementation realities?
- Which of the risks, barriers and assumptions hold/do not hold? How are they affecting the achievement of results?

Source: GCF: Evaluation Operational Procedures and Guidelines for Accredited Entity-led Evaluations, 2023

Annex 4: Key Resources for Conducting EQAs

1. GCF. (2020, November). Integrated Results Management Framework.
2. GCF. (2021). Evaluation Policy for the GCF.
3. GCF. (2022, April 28). GCF Evaluation Standards.
4. GCF. (2022, December). Integrated Results Management Framework: Results Handbook.
5. GCF. (2023, March). Evaluation Operational Procedures and Guidelines for Accredited Entity-led Evaluations.
6. GCF. (2023). Evaluation Guidelines.
7. GCF. (n.d.). Monitoring and Accountability Framework for Accredited Entities.
8. GCF. (2019). Gender Action Plan of the GCF, 2020-2023.
9. GCF. (2021). Revised Environmental and Social Policy.
10. GCF. (n.d.). Indigenous Peoples Policy

Annex 5: EQA Tool

See following pages.

Assessment Tool for Rating the Quality of AE-led Evaluations

The GCF evaluation quality assessment (EQA) framework is based on GCF's Evaluation Policy, GCF's Evaluations Standards and related guidance, as well as United Nations Evaluation Group (UNEG) standards. The assessments are conducted for mid-term reviews and final evaluations by external reviewers.

The EQA Tool has three components which address: overall evaluation quality, coverage of GCF criteria, and adherence to GCF Evaluations Standards. This version of the EQA Tool can be useful as a checklist to help guide evaluation planning, the content and structure of evaluation reports, and internal quality assurance processes. The version used by reviewers is a spreadsheet that enables auto-calculation of section-level and overall scores.



PART A: EVALUATION QUALITY ASSESSMENT

This part of the Tool includes a series of questions with a total of 50 sub-questions that are used to rate evaluation reports on the extent to which UNEG and GCF-specific expectations for evaluations are met. It is organized into nine sections that have different weightings to reflect their level of importance. The calculated scores lead to an overall report rating of Exceptional, Very Good, Good, Fair or Unsatisfactory. Although space for comments is provided here for each section, the reviewer version of the EQA Tool has space for comments for each sub-question that justify each of the ratings.

RATING SCALE: A four-point ordinal scale is applied to each sub-question: Yes=1, Mostly=.67, Partly=.33, No=0, with the option for N/A.
[Insert the appropriate score on the right side of each statement]

A	EXECUTIVE SUMMARY	Weight: 5%
1	Can the executive summary inform decision-making?	RATING
1.1	Is clearly presented and reflects an analytical summary of progress/results as per GCF's relevant thematic area.	
1.2	Serves as a standalone document to inform decision-making by including key elements (overview of project, evaluation purpose and objectives, intended audience, methodology, key conclusions on findings, lessons learned, key recommendations) reflected in the main body of the report. [Lessons only required for final evaluations]	
1.3	Is of relevant conciseness and depth for key users (maximum of 5 pages unless otherwise specified in ToR).	
	COMMENTS	TOTAL

B	BACKGROUND	Weight: 10%
2	Is the object of the evaluation clearly described?	RATING
2.1	Clear and relevant description of the intervention including general objective(s), location(s), timelines, cost/budget, and implementation status (including any extensions and rationale for why extension was granted).	
2.2	Clear and relevant description of intended rightsholders/beneficiaries. Their needs and interests in relation to climate mitigation and/or adaptation are apparent.	

3	Is the context of the intervention clearly described, including the climate-resilient pathways being addressed?	RATING
3.1	Clear and relevant description of the context of the intervention, as well as other key issues (i.e., policy, socio-economic, political, gender, cultural, institutional, international factors).	
3.2	Linkages drawn to NDCs and other relevant global, regional, and national climate plans and commitments.	
4	**Are key stakeholders, their relationships and contributions clearly identified?**	**RATING**
4.1	Identification of implementing agency(ies), national implementing partners, local communities, social institutions, and other stakeholders.	
4.2	Identification of the specific contributions and roles of key stakeholders including all implementing agencies and government. There is clarity on the linkages between stakeholders (e.g., stakeholder map), and on how the AE has worked in a coordinated way with other implementing agencies.	
5	**Is the theory of change and/or the logical framework well-articulated?**	**RATING**
5.1	Clear description of the intervention's intended objectives / main results.	
5.2	Causal relationship between outputs and outcomes is presented in narrative and/or graphic form (e.g., results chain, logical framework, theory of change).	
5.3	The theory of change is assessed and, if needed, is reformulated/improved by the evaluators with the changes being explained. Assumptions are clearly articulated and are assessed for validity by evaluators. [not required if project approved before 2020]	
	COMMENTS	TOTAL

C	EVALUATION PURPOSE, OBJECTIVES, SCOPE & CONDUCT	Weight: 5%
6	**Are the purpose and parameters clearly described?**	**RATING**
6.1	Purpose of evaluation is clearly defined, including key intended uses and users.	
6.2	Clear description of evaluation objectives, which should include assessment of the results related to climate mitigation and/or adaptation. Any changes made to the objectives included in the ToR are explained.	
6.3	Clear description of the scope of the evaluation: what will and will not be covered (thematically, chronologically, geographically), as well as, if applicable, the reasons for this scope.	

7	Is the management and undertaking of the evaluation process clear and ensuring of independence?	RATING
7.1	Clear description of the roles and responsibilities of stakeholders in the evaluation process (i.e., for commissioning AE, evaluation manager, NDA, EE, reference group)	
	COMMENTS	TOTAL

D	EVALUATION DESIGN AND METHODOLOGY	Weight: 25%
8	Does the evaluation systematically use questions and the relevant evaluation criteria that are appropriate for the purpose of the evaluation? GCF criteria include Relevance, Efficiency, Effectiveness, Sustainability, Impact, Coherence in Climate Finance Delivery, Gender Equity, Country Ownership, and Unexpected Results. Final evaluations to also cover Innovativeness and Replicability/Scalability.	RATING
8.1	Evaluation questions / sub-questions cover all criteria and are appropriate for meeting the objectives and purpose of the evaluation. They incorporate the assessment of project design and financial management.	
8.2	Questions address GCF's unique mandate in climate finance, its focus on catalyzing a paradigm shift (or its potential) towards low-emission and climate-resilient development pathways, and progress in respect to its outcome indicators. [required for final evaluations; n/a if not addressed in interim reviews]	
8.3	A thorough evaluation matrix is presented that, along with evaluation questions and sub questions, includes indicators (or other metrics to measure achievement), data sources and methods.	
9	Is the evaluation design and methodology sufficiently robust and enabling of credible findings?	RATING
9.1	Evaluation design/approach and set of methods are clearly explained, relevant and adequately robust for the evaluation's purpose, objectives, and scope; they enable causal connections to be drawn between outputs and expected outcomes.	
9.2	Data sources are appropriate and adequate - these would normally include qualitative and quantitative sources (unless otherwise specified in the ToR) - and are all clearly described. The number of respondents for each method is identified and it is clear how triangulation is achieved.	
9.3	Sampling strategy is provided - it addresses how diverse perspectives are captured (or, if not, provide reasons for this).	
9.4	Clear description of the methods of analysis. The type of analysis (such as descriptive &/or analytical statistics) and processes for analysing data (such as excel database for coding, AI, or any specific software used) are identified.	

9.5	Clear description of limitations and constraints faced, and of how these impacted the quality of the evaluation and were addressed by the evaluators to the extent feasible.	
9.6	Clear description of how the methodology was responsive to human rights, gender equality, indigenous peoples, disability, and other aspects of inclusion, including of how the participation/representation of the range of beneficiaries/rightsholders was ensured. Evaluation respondents are disaggregated by gender and by other group relevant to the intervention.	
10	**Are ethical issues and considerations described?**	**RATING**
10.1	Steps taken to adhere to ethical principles and ensure ethical conduct within the evaluation process are described (i.e., confidentiality, informed consent, respect for dignity and diversity, do no harm approach when consulting beneficiaries and particularly minors and vulnerable groups).	
	COMMENTS	TOTAL

E	**EVALUATION FINDINGS**	**Weight: 30%**
11	**Do the findings clearly address the evaluation objectives and questions?**	**RATING**
11.1	Findings marshal sufficient information to systematically address all the evaluation's questions related to the standard criteria of relevance, effectiveness, efficiency, impact and sustainability. There is explicit reference to, and use of, the intervention's results framework/ToC, and climate results.	
11.2	Findings marshal sufficient information to systematically address all the evaluation's questions related to the additional GCF criteria of Coherence in Climate Finance Delivery, Gender Equity, Country Ownership, and Unexpected Results, as well as project design and financial management.	
11.3	Findings marshal sufficient information to systematically address the criteria of Innovativeness and Replicability/Scalability. They also assess contribution to paradigm shift and enabling environment (using scorecards if project approved after May/22), and assess and verify outcomes related to reduced emissions/increased sequestration and/or adaptation. [Required for final evaluations, n/a if not covered in midterm reviews]	
12	**Are evaluation findings clearly presented? Are they derived from the conscientious, explicit and judicious use of the best available, objective, reliable and valid data and by accurate quantitative and qualitative analysis of evidence?**	**RATING**
12.1	Evaluation uses credible forms of qualitative and quantitative data. It presents both output and outcome-level data as relevant to the evaluation framework. Triangulation is evident through the use of multiple data sources.	

12.2	Findings are clearly presented. They are supported by, and respond to, the evidence presented, including both positive and negative aspects. Findings are based on clear performance indicators, standards, benchmarks, or other means of comparison as relevant for each question. There is an indepth level of analysis of outcome-level results in particular.	
12.3	Causal factors (contextual, institutional, managerial, etc.) leading to achievement or non-achievement of results are clearly identified in the analysis.	
13	**Does the evaluation address the intervention's Integrated Results Monitoring Framework (IRMF)?**	**RATING**
13.1	Adequacy of the intervention's design and monitoring system is assessed (i.e., completeness and appropriateness of results framework, monitoring tools and their usage) to support decision-making.	
13.2	Use and communication/sharing of monitoring data and reporting for learning and accountability purposes is assessed.	
	COMMENTS	TOTAL

F	**EVALUATION CONCLUSIONS & LESSONS LEARNED**	**Weight: 5%**
14	**Do the conclusions clearly present an objective overall assessment of the intervention?**	**RATING**
14.1	Conclusions are clearly formulated and reflect all of the criteria that were to be covered including gender dimensions and inclusivity. They are derived appropriately from the findings, clearly presenting the strengths and weaknesses of the intervention. They do not introduce new information.	
14.2	Conclusions add insight and analysis beyond the findings. They reflect the purpose and objectives of the evaluation and are sufficiently forward looking (if an interim review or if the implementation is expected to continue or have an additional phase).	
15	**Are logical and informative lessons learned identified?**	**RATING**
15.1	Identified lessons stem logically from the findings and have wider applicability and relevance beyond the object of the evaluation. They are clearly and concisely presented (with the topic identified) yet have sufficient detail to be useful for organizational learning. [only required for final evaluations]	
	COMMENTS	TOTAL

G	RECOMMENDATIONS	Weight: 10%
16	Are recommendations well-grounded in the evaluation and useful?	RATING
16.1	Recommendations align with the evaluation purpose and are logically derived from the findings and conclusions. They address any major weaknesses identified in the findings, including in addressing gender and ESS.	
16.2	Recommendations are useful and realistic. They are actionable for primary intended users and uses (specific and relevant to the intervention); guidance is given for implementation, as appropriate.	
17	Are recommendations clearly presented?	RATING
17.1	Recommendations are clearly stated with stakeholders responsible for action for each recommendation identified (including the lead if multiple actors responsible). [responsibility not required for small scale projects where target is clearly the AE]	
	COMMENTS	TOTAL

H	ENVIRONMENTAL & SOCIAL SAFEGUARDS	Weight: 5%
18	Are environmental and social safeguards (ESS) adequately addressed?	RATING
18.1	Analysis includes extent to which environmental, social and gender risks, and efforts to avoid, mitigate and manage these, were considered at project design stage (including use of screening and assessment processes) and reflected in Environmental and Social Management Plans.	
18.2	Analysis includes extent to which these risks and impacts have been monitored, managed and reported during project implementation, and stakeholder engagement carried out with communities most affected. This includes whether monitoring covers adverse impacts based on gender and age (including SEAH), and if the exercising of due diligence and use of participatory monitoring approaches is evident. [Depth of expected analysis is proportional to risk level]	
18.3	Conclusions and Recommendations reflect the analysis of ESS and address shortcomings found.	
	COMMENTS	TOTAL

I	REPORT STRUCTURE AND PRESENTATION	Weight: 5%
19	**Does the evaluation report include all relevant information?**	**RATING**
19.1	Opening pages include name and project number of evaluated object, timeframe of the evaluation, date of report, location of evaluated object, names and/or organization(s) of the evaluator(s), name of organization commissioning the evaluation, and table of contents.	
19.2	Annexes include at a minimum the terms of reference, evaluation matrix, list of stakeholder groups participating as informants, ToC and/or logical framework (unless included in report body), data collection instruments, and list of documentary evidence.	
20	**Is the report logically structured?**	**RATING**
20.1	Report is easy to navigate (for instance, with numbered sections, clear titles and sub-titles, well formatted).	
20.2	Structure follows an easily discernible logical flow. Context, purpose, and methodology would normally precede findings, which would normally be followed by conclusions, lessons learned and recommendations.	
21	**Is the report well presented?**	**RATING**
21.1	Report is of reasonable length; the main part of the report (excluding opening pages and annexes) does not exceed the number of pages specified in ToR, or if not specified, has a maximum of 60 pages (with some flexibility for size of project and complexity).	
21.2	Report is easy to understand (written in accessible way for intended audience) and generally free from grammar, spelling, and punctuation errors.	
21.3	Well-chosen visual aids (such as infographics, maps, tables, figures) are used to convey key information. These are clearly presented and labeled, and the source and connection to the text is discernable.	
	COMMENTS	

PART B: CRITERIA COVERAGE

This section addresses the extent to which each of the GCF evaluation criteria are addressed, recognizing that the level of emphasis on each should be determined by the type of project, evaluation scope, and type of evaluation. It is acceptable for some to be integrated into other criteria, and only final evaluations are required to address Innovativeness and Replicability & Scalability.

RATING OPTIONS: Yes, Partial, No or N/A
(Partial to be used when there is only very brief and insufficient coverage)

Criteria	Covered?
Relevance	
Efficiency	
Effectiveness	
Sustainability	
Impact	
Coherence in Climate Finance Delivery	
Gender Equity	
Country Ownership	
Unexpected Results	
Innovativeness	
Replicability/Scalability	

PART C: GCF EVALUATION STANDARDS ASSESSMENT

This section assesses the extent to which evaluations align with the 15 GCF Evaluation Standards through a series of 45 questions. The questions were extracted from the GCF Evaluation Standards and Guidelines documents[12] and reflect the requirements that can reasonably be expected to be found in an evaluation report and/or evaluation ToR. As such, this does not provide a comprehensive assessment of alignment with Standards, and it is even more limited if the evaluation ToR is not attached to the report.

A major focus of the Standards is on ethical evaluation practices. The inclusion in the evaluation report of the UNEG Pledge of Ethical Conduct in Evaluation signed by evaluators is considered sufficient for addressing relevant indicators.

RATING SCALE: A three-point ordinal scale is applied to each sub-question: Yes=1, Partly=0.5, No=0, N/A [Insert the appropriate score on the right side of each statement]

Across Standards	RATING
Does the ToR inform evaluators of the need to apply GCF evaluation standards and GCF evaluation criteria? (ToR)	
Does the ToR, at a minimum, contain a project/programme description, evaluation scope and objectives, evaluation questions, methodological guidance/preferences (while retaining flexibility), deliverables, team requirements, timeframe, and management structure of evaluation process? (ToR)	
Does the ToR set out clear and unified set of instructions for the evaluation? (ToR)	
1. Independence	**RATING**
Is it clear that the evaluation was conducted by external evaluators who were free of conflict of interest? (evaluation report)	
Are evaluators informed of reporting responsibilities and mechanisms for conflicts of interest? (ToR)	
2. Impartiality and Objectivity	**RATING**
Does the evaluation appear to have been implemented in accordance with the TOR? (evaluation report)	
Is it evident that methods and tools for data collection, site selection and key informants were based on objective criteria to ensure the absence of bias? (evaluation report)	

[12] GCF Evaluation Standards https://ieu.greenclimate.fund/sites/default/files/document/220428-gcf-evalluation-standards-web.pdf and Evaluation operational procedures and guidelines for Accredited Entity-led evaluations https://www.greenclimate.fund/document/evaluation-operational-procedures-and-guidelines-accredited-entity-led-evaluations

3. Utility and Value-Added	RATING
Are the intended audiences of the evaluation results identified? (ToR and evaluation report)	
Is there clarity about how evaluation results are to be used (ToR and evaluation report)	
Are evaluation findings and recommendations relevant and valuable to the work of the GCF and the work of the entity conducting the evaluation? (evaluation report)	
Are recommendations presented in a form that enables decision-makers to clearly identify their responsibility and track follow-up action? (evaluation report)	
4. Ownership and Participation	**RATING**
Does the TOR clearly set out the major stakeholders and their interests, both in the subject of the evaluation and the evaluation itself, along with a plan for their consultation and engagement during the evaluation? This includes their review and comment on draft deliverables. (ToR)	
Is it clear how stakeholders were engaged throughout the evaluation process in the evaluation approach or methodology beyond solely as subjects of data collection? (evaluation report)	
Do findings appropriately reflect the various perspectives and voices of the multiple stakeholders involved? (evaluation report)	
Is it evident that power imbalances were recognized and addressed, and that participatory and empowerment approaches were favored? (evaluation report)	
5. Credibility and Rigour	**RATING**
Did evaluators use credible and independent data and methodologies that measure and assess whether, what, how and how much the GCF contributes towards mitigation and adaptation to climate change? (evaluation report)	
Did the evaluation use mixed methods and draw on multiple sources of evidence, including quantitative and qualitative evidence derived from both primary and secondary data sources? (evaluation report)	
Did the approach enable for testing of development hypotheses or assumptions and ensure that causal pathways are well defined? (evaluation report)	
Are the questions for surveys, focus groups or interviews value neutral, culturally and age appropriate, and not likely to cause stress to participants? (evaluation report)	
6. Transparency	**RATING**
Are evaluators informed of the obligation to provide a clear and transparent accounting of the findings? (ToR)	
Were quality assurance processes required and established? (ToR and evaluation report)	

Are findings coherently anchored in the evidence base with sources on which they are based clearly stated? (evaluation report)	
Are conclusions coherently anchored in the analysis and do recommendations have their genesis in the conclusions? (evaluation report)	
Is the evaluation easily readable? (evaluation report)	
7. Learning	**RATING**
Does the evaluation contribute to organizational learning about the effectiveness of GCF's efforts to address the climate change crisis—specifically by increasing understanding of the causal relationships and impacts of its interventions?	
Does the evaluation include a learning plan that identifies the mechanisms to share findings widely? (evaluation report)	
8. Human Rights, Gender Equality and Environmental Considerations	**RATING**
Is it evident that data-collection methods were sensitive to the intersecting factors among stakeholders, such as ethnicity, age or socioeconomic class? (evaluation report)	
Does the analysis consider how vulnerable groups and other stakeholders may have been impacted in different ways by the intervention, i.e. present disaggregated results? (evaluation report)	
Is it clear how the evaluation minimized its carbon footprint, e.g. in travel arrangements and carbon offsetting of emissions? (evaluation report)	
9. Confidentiality	**RATING**
Were the identity and confidentiality of evaluation participants protected throughout the evaluation process, and is it evident participants were informed about how the evaluation would use their data? (evaluation report)	
Does the evaluation plan have clear protocols for the storage and destruction of data after the evaluation? (evaluation report)	
Is it evident that data collection teams (where used) were trained on the importance of confidentiality and their responsibility for protecting participants' privacy? (evaluation report)	
10. Cost Effectiveness	**RATING**
Is it apparent that the proposed approach to gathering evidence is the most cost-effective? (ToR)	
Are the time frame and resources (i.e., # of days) realistic for achieving the intended purpose and outcomes, including for engaging local stakeholders and communicating findings to different stakeholder groups? (ToR or evaluation report)	
11. Ethics	**RATING**
Is it clear that evaluators need to adhere to ethical requirements? (ToR)	

Is there an explanation for how evaluators adhered to high standards of ethics and professional conduct? (evaluation report)	
Is it evident that evaluators received free, prior and informed consent and permission from participants in advance of data collection? (evaluation report)	
Were formal or informal ethical review processes applied before conducting the evaluation in cases where primary data was collected from potentially vulnerable people or in sensitive contexts? (evaluation report)	
12. Integrity	**RATING**
Is it apparent that evaluators and those being evaluated were guided by a professional sense of integrity and conduct? (evaluation report)	
13. Accountability	**RATING**
Is there transparency regarding the evaluation's purpose? (ToR and report)	
It is apparent that the evaluation was conducted in a rigorous, fair and balanced manner with judgments based on sound and complete evidence? (evaluation report)	
14. Competence	**RATING**
Is there a requirement for the evaluation team to have a combination of expertise in leading evaluations, adaptation, sector (i.e., forestry/ecosystems/agriculture /infrastructure etc.), gender and social inclusion, and qualitative and quantitative approaches depending on the evaluation design? GHG accounting is expected for cross-cutting and mitigation interventions. (ToR)	
Is it evident that the evaluators, collectively, had the required qualifications, expertise and experience to conduct the evaluation competently? (evaluation report)	
15. Respect and Beneficence	**RATING**
Is it evident that the evaluators were sensitive to the beliefs, manners and customs of the social and cultural environments in which they worked? (evaluation report)	
Is it evident that the evaluation team was comprised of members with appropriate representation regarding sex and a broad mix of backgrounds, skills and perspectives, including national and international expertise? (evaluation report)	
TOTAL	

COMMENTS: