

GEF IEO Annual Performance Report 2023

(Prepared by the Independent Evaluation Office of the GEF)

- Draft Approach Paper -

Point of contact:

Neeraj Negi, Senior Evaluation Officer, nnegi1@thegef.org

Contents

1. Background and Context.....	1
2. Evaluation Questions, Coverage and Methodology	1
2.1 Performance of Completed Projects and Programs	1
2.2 Special APR Theme: Behavior Change for Sustainability	3
2.3 Management Action Record 2023	6
2.4 Peer and Stakeholder Feedback	6
3. Resources and Schedule	7
5.1 Evaluation Team.....	7
5.2 Schedule of Work and Resources	7
References	8
Annex A: terminal evaluations submitted	9
Annex B. Performance Criteria and Rating Scales	10
Annex C. Terminal Evaluation Report Validation Guidelines	17

Abbreviations

ADB	Asian Development Bank
AFDB	African Development Bank
APR	GEF Annual Performance Report
EBRD	European Bank for Reconstruction and Development
FAO	Food and Agricultural Organization of the United Nations
GEF	Global Environment Facility
GEF IEO/IEO	Global Environment Facility's Independent Evaluation Office
IADB	Inter-American Development Bank
IFAD	International Fund for Agricultural Development
MAR	Management Action Record
STAP	Scientific and Technical Advisory Panel
STAR	System for Transparent Allocation of Resources
UNDP	United Nations Development Programme
UNEP	United Nations Environment Programme
UNIDO	United Nations Industrial Development Organization
WB	World Bank
WWF-US	United States World Wildlife Fund

1. Background and Context

The Independent Evaluation Office (IEO) of the Global Environment Facility (GEF) prepares the Annual Performance Report (APR) to provide an overview of the performance of GEF activities and processes, the key factors that affect performance, quality of Monitoring and Evaluations systems (M&E), and the Management Action Record (MAR). In addition, APR2023 will include a special theme on the use of behavior change approaches in GEF activities.

This paper discusses the key questions, methodology, arrangements for stakeholder involvement, evaluation products, and calendar of activities for APR2023. Two peer feedback providers will provide quality assurance. The emerging findings of APR2023 will be shared across the GEF Partnership through an interagency meeting to facilitate discussion and feedback on the draft report. The findings of APR2023 will be presented at the GEF Council's meeting in June 2023.

2. Evaluation Questions, Coverage and Methodology

APR 2023 will present an analysis on various themes drawing on different datasets and methodological approaches.

2.1 Performance of Completed Projects and Programs

Evaluation Questions

Assessment of, and reporting on, the performance of completed projects and programs is a regular feature of the APR. The report provides an assessment of project outcomes, sustainability, implementation, and M&E. In addition, the report includes a discussion of the issues related to the quality of terminal evaluations. The reporting is on a cumulative basis with special attention to projects for which terminal evaluations were submitted after the previous APR. APR2023 will seek to answer the following questions:

- To what extent have projects achieved their expected results?
- How well are the projects implemented and executed?
- What is the quality of project M&E?
- What is the quality of the terminal evaluations?

To what extent have GEF projects achieved their expected results? The assessment will determine the extent to which GEF projects delivered their expected results based on the criteria of relevance, coherence, effectiveness, efficiency and sustainability. It will also analyze the trends in outcome and sustainability ratings of completed projects and assess the factors, such as country context, stakeholder engagement, and co-financing, that may affect achievement of results.

How well are GEF-supported activities implemented and executed? The assessment will examine how GEF-supported activities are implemented and executed on the ground. This includes the performance of the GEF Agencies in designing and implementing projects, including their supervision and support of executing partners, and the performance of the executing partners in executing planned activities. It will examine challenges faced in implementation and execution, and how these are addressed by the GEF Agencies and their partners. It includes trends in implementation and execution ratings.

What is the quality of project M&E? The assessment will determine the quality of project M&E at the point of CEO Approval/Endorsement and during project implementation, including trends in quality of M&E ratings, and the factors that influence these ratings. For completed projects from GEF-6 onwards, the assessment will determine the extent to which a) indicators specified in the M&E plan adequately cover the project objectives, b) specific targets have been provided for each of the indicators, and c) achievement of targets is reported on.

What is the quality of terminal evaluations submitted by the GEF Agencies? The report includes an analysis of the quality of terminal evaluations based on: timeliness of conduct and submission; completeness of general information on evaluated projects; stakeholder involvement in report preparation; quality of reporting on the project theory of change; description of evaluation methodology; clarity and candor in accounts of project outcomes; quality of assessment of sustainability; appraisal of project M&E; reporting on implementation and execution; coverage of implementation of safeguards; recommendations and lessons; consistency of ratings; and report presentation. It will assess trends in the quality of terminal evaluation ratings and note strengths and weaknesses of the terminal evaluations submitted by the Agencies.

Coverage

APR2023 will report on completed projects for which GEF Agencies have submitted terminal evaluations through September 2022. The last APR in 2021 covered 1806 completed GEF projects. In addition to these, about 400 terminal evaluations that were submitted after closing for APR2021 will be covered in APR2023. Thus, in all, about 2200 completed projects would underpin the analysis for APR2023 (Annex A).

Methodology

Analysis of Performance Ratings of Completed Projects

Rating scales. Project performance will be rated in the following areas: outcomes, risks to sustainability, implementation, execution, M&E design, M&E implementation, and quality of terminal evaluation reports. Criteria for assessing performance on these has mostly remained the same as in APR2021 – the major change for APR2023 is the update to the criteria for assessing the quality of terminal evaluation reports. The quality of terminal evaluation reports will now be assessed on 14 criteria and 62 sub-criteria, instead of the six criteria that were assessed earlier. The Report on the Review of the Terminal Evaluation Validation Process (GEF IEO 2020) had called for a more in-depth assessment of the quality of terminal evaluation reports. The change in criteria to assess the quality of terminal evaluations

addresses the findings of the report. The 14 criteria covered in the revised approach also include the six that were used earlier, which will facilitate comparison across periods. Rating scales used are described in detail in Annex B.

Validation of Terminal evaluations. Terminal evaluations of projects included in the APR are reviewed by the GEF IEO or the Agency evaluation units. The validation process ascertains whether the ratings provided in a terminal evaluation are well-substantiated and consistent. Where necessary, the reviewers may revise ratings to ensure that these are consistent with the evidence.

The GEF IEO accepts the terminal evaluation review ratings provided by the World Bank's Independent Evaluation Group, UNDP Independent Evaluation Office, UNEP Evaluation Office, and IFAD Independent Office of Evaluation, if these ratings are available within two years of completion of a terminal evaluation.¹ Acceptance of ratings provided by these evaluation units reduces duplication of effort. To track whether the ratings provided by these Agency evaluation units continue to be consistent with the GEF IEO ratings, the Office validates a sample of 10 percent of the terminal evaluations submitted by these Agencies. Overall, there is little difference in ratings provided by the GEF IEO and evaluation offices of these Agencies. Terminal evaluations submitted by other Agencies are validated by the GEF IEO.

The GEF IEO process for terminal evaluation validation is standardized. This process is described in detail in Annex C.

Tracking achievement of project objectives through results indicators

Previous APRs (GEF EO 2006, 2009) have presented quality-at-entry assessments of M&E arrangements, as well as trends in ratings of the quality of M&E design and implementation. The analysis in APR 2023 will assess whether M&E plans of completed GEF-6 projects adequately include results indicators to track progress in achievement of project objectives. It will examine whether specific targets are provided for the specified indicators, and if so, whether the terminal evaluations provide information on the extent to which these targets were achieved. This analysis will be based on the desk review of the project documents and terminal evaluations.

2.2 Special APR Theme: Behavior Change for Sustainability

¹ Report on the Review of the GEF Terminal Evaluation Validation Process (GEF IEO, 2020) concluded that UNIDO, FAO, and IFAD, were well positioned to graduate from the GEF IEO validations. The GEF IEO agreed with the conclusion in principle. It has graduated the terminal evaluations submitted by IFAD from the validation process and will accept the validation reports prepared by the IFAD's Independent Office of Evaluation. The Office had planned to extend the graduation from validation of terminal evaluations submitted by FAO but had to defer because its Office of Evaluation shifted the responsibility of conducting the terminal evaluations to the operations. The Office is in contact with the office of Evaluation and Internal Oversight at UNIDO to explore how its terminal evaluations may be graduated from the validation process.

Over the last 30 years, the GEF has funded a range of technologies and practices to generate global environmental benefits. Previous GEF IEO evaluations have found that for these environmental outcomes to be sustained and/or scaled up beyond project completion, key actors have to be motivated to change their behavior, i.e. to adopt and continue using these technologies and practices (GEF IEO 2021a, 2022a). Similarly, institutions have to change their behavior in terms of providing support for the adoption and continued use of these technologies and practices (e.g. GEF IEO 2018, 2019, 2021a, 2021b, 2022a, 2022b, 2022c).

The GEF's Scientific and Technical Advisory Panel (STAP) has observed that while many GEF projects require behavior change to achieve their targets, behavior change is often an implicit objective. That is, while the behavioral outcomes sought are clear, the mechanisms to achieve them often are not (Metternicht et al 2020).

This study aims to assess if and how GEF-supported interventions contribute to behavior change at the project level. It will take an in-depth look at the mechanisms used by projects to catalyze behavior change that contributes to global environmental benefits being sustained beyond project completion. In this context, behavior change is narrowly defined as the adoption and continued use of technologies and practices--by the relevant stakeholders--that a GEF project has introduced or promoted to generate global environmental benefits. The study will focus on identifying project activities that facilitate and foster such behavior change, how effectively these activities have been implemented, and to what extent they have contributed to observable behavior change. The findings will provide lessons for the design of new GEF projects, as well as to improve the design of ongoing projects at or prior to midterm. More specifically the study will answer the following questions:

- To what extent have completed GEF-supported projects achieved their intended behavior change outcomes?
- What project activities and factors have promoted behavior change in successful projects, and prevented behavior change from happening in similar but unsuccessful projects?
- To what extent have completed GEF-supported projects supported conditions that increase the likelihood of behavior change being sustained?

Coverage

The ex-post review will cover about 70 completed projects that were CEO-endorsed in GEF-6 onwards, focusing on the GEF's latest approaches. In addition, projects included in the study must have as an objective the adoption of at least one technology or practice that can directly generate environmental benefits. The identification of the final portfolio will be done in conjunction with the TE validation exercise.

Methodology

The study is based on the premise that 1) behavior change among individual and institutional actors is central to a project's environmental outcomes being achieved and sustained, and 2) project activities can catalyze changes in behavior. The ex-post review will break down each project's theory of change

into three key components: which behavior needs to change to generate environmental benefits, whose behavior needs to change, and which project activities aimed to catalyze this behavior change.

This study will primarily use [STAP guidance](#) (Metternicht et al 2020) to assess project theories of change, as this guidance is based on a synthesis of various behavior change frameworks. A standardized instrument will be developed to extract information from terminal evaluations on the types of practices and technologies promoted, types of activities implemented to support adoption and continued use, extent of behavior change and environmental benefits achieved, factors affecting differences in behavior outcomes, and likelihood of behavior change being sustained.

To what extent have completed GEF-supported projects achieved their intended behavior change outcomes? The ex-post review will assess the degree of success in achieving behavior change and global environmental benefits in the covered portfolio, as reported in terminal evaluations. The types of behavior change, corresponding actors whose behavior needs to change, and associated project activities will be identified. A rating scale will be developed to characterize the extent of behavior change achieved. These ratings will be compared across projects that promoted similar technologies and practices.

What project activities and factors have promoted behavior change in successful projects, and prevented behavior change from happening in similar but unsuccessful projects? Based on ratings generated by the review of terminal evaluations (see previous paragraph), projects supporting similar interventions will be grouped as successful or less successful. These will be analyzed for activities, characteristics and conditions that may be associated with their success or lack thereof.

In general, two types of project activities and factors help catalyze behavior change: those that make it easier to adopt the desired behavior, and those that make it difficult to continue undesirable behavior or return to the status quo. Examples of project activities that support behavior change are education and training on how to use the technologies and practices, widely accessible financing, and infrastructure such as help centers that assist in troubleshooting problems. Examples of factors that affect extent of behavior change are the types of stakeholder groups targeted, degree of cultural acceptance, existing incentives and disincentives to adopt practices, presence or absence of supporting institutions, ease or complexity of technology use in the particular context, etc. Project activities and factors that contributed to differences in behavior outcomes will be identified by combining qualitative and quantitative tools for analysis, such as NVivo, EvalC3 and statistical correlations and regressions.

To what extent have completed projects supported conditions that increase the likelihood of behavior change being sustained? The study will assess the likelihood of the behavior change being sustained, building on the GEF IEO's previous findings on [how impacts are scaled up in the GEF](#) (GEF IEO 2021a). Enabling conditions that increase the likelihood of adoption and sustained use of practices and technologies include, among others: sustainable financing, the appropriate policy framework and operating guidelines, individual & institutional training, multistakeholder interactions & partnerships, systematic learning mechanisms, incentives & disincentives, knowledge & information dissemination, and participatory mechanisms. This assessment may build on the rating scale used to assess over-all likelihood of project sustainability in terminal evaluations.

2.3 Management Action Record 2023

The GEF Management Action Record (MAR) system was established in 2005 to track the level of adoption of GEF Council decisions and underlying GEF IEO recommendations by the GEF Management. In June 2020, as a response to the findings and recommendations of the Third Professional Peer Review of the Independent Evaluation Function of the Global Environment Facility, the GEF IEO shifted to tracking the implementation of the action plan included in the Management Response to an evaluation. MAR 2023 will, for the very first time, present progress on the implementation of action plans.

GEF IEO has recently undertaken a separate review of the lessons from the GEF MAR process, which takes stock of good practice used by other multilateral organizations. The review was presented to the Council in December 2022. The lessons from the review will be incorporated in MAR2023.

2.4 Peer and Stakeholder Feedback

In addition to quality control mechanisms within the terminal evaluation review process, APR2023 will engage two peer reviewers to provide feedback during the evaluation process. The reviewers are presently being identified.

The GEF IEO publishes the TE validation reports on the GEF IEO website on a rolling basis. In addition, the GEF IEO also shares validation reports with the respective Agencies before an APR is prepared. This ensures that reporting through the APR is accurate, and that the validation process is clear and transparent.

The preliminary findings of the APR2023 will be shared with the key stakeholders through an interagency meeting. The meeting will provide an opportunity for key stakeholders, such as the GEF Agencies (operations and evaluation units), the Secretariat, STAP, and the CSO Network, to provide feedback on the emerging findings of the APR.

Expected Outputs, Outreach and Tracking

The APR is primarily intended for the GEF Council and a GEF corporate audience, including the GEF Secretariat, the GEF Partner Agencies, STAP, and the CSO Network. The report will be delivered at the June 2023 Council meeting. The report will be published on the GEF IEO website and will also be available as a Council document on the GEF website. Data on ratings can also be accessed at the Data & Ratings web page at the GEF IEO website (<https://www.gefio.org/data-ratings>). Other outputs include technical papers on the behavior change outcomes of GEF-supported initiatives.

3. Resources and Schedule

5.1 Evaluation Team

APR2023 will be completed by a team including Neeraj Kumar Negi, Senior Evaluation Officer; Jeneen Garcia, Evaluation Officer; and other GEF IEO staff and consultants.

5.2 Schedule of Work and Resources

The draft report will be ready by March 21st, 2023, and the final report will be ready by May 15th, 2023, in time for the spring meeting of the GEF Council. Table 1 shows the schedule of work activities for completion and presentation of the findings of APR2023.

Table 1. Schedule of work activities for completion and presentation of APR 2023

Project milestone	Work period or completion date
Approach paper	November 30 th , 2022
Review of terminal evaluations	November 1 st , 2022 – January 31 st , 2023
Analysis of terminal evaluation data	Feb 1 st , 2023 – Feb 28 th , 2023
Review of behavior change for sustainability	January 1 st 2023 to Feb 28 th , 2023
Draft report	March 21 st , 2023
Complete report for sharing with Secretariat	April 15 th 2023
Council information document of APR2023 uploaded	May 15 th , 2023
Publication of the finalized report of APR2023	June – September 2023

References

Global Environment Facility Evaluation Office (GEF EO). 2006. *GEF Annual Performance Report 2005*, Evaluation Report No. 31, Washington, DC.

Global Environment Facility Evaluation Office (GEF EO). 2009. *GEF Annual Performance Report 2008*, Evaluation Report No. 49, Washington, DC.

Global Environment Facility Independent Evaluation Office (GEF IEO). 2018. *Evaluation of GEF Support for Transformational Change*, Evaluation Report No. 122, Washington, DC: GEF IEO.

Global Environment Facility Independent Evaluation Office (GEF IEO). 2019. *Strategic Country Cluster Evaluation: Small Island Developing States (SIDS)*, Unedited Report, Washington, DC: GEF IEO.

Global Environment Facility Independent Evaluation Office (GEF IEO). 2021a. *GEF Support to Scaling Up Impact*, Evaluation Report No. 138, Washington, DC: GEF IEO.

Global Environment Facility Independent Evaluation Office (GEF IEO). 2021b. *GEF Support to innovation: Findings and Lessons*, Unedited Report, Washington, DC: GEF IEO.

Global Environment Facility Independent Evaluation Office (GEF IEO). 2022a. *GEF Engagement with Micro, Small, and Medium Enterprises*, Evaluation Report No. 144, Washington, DC: GEF IEO.

Global Environment Facility Independent Evaluation Office (GEF IEO). 2022b. *Strategic Country Cluster Evaluation of Sahel and Sudan-Guinea Savanna Biomes*, Evaluation Report No. 141, Washington, DC: GEF IEO.

Global Environment Facility Independent Evaluation Office (GEF IEO). 2022c. *Strategic Country Cluster Evaluation of the Least Developed Countries*, Evaluation Report No. 142, Washington, DC: GEF IEO.

Metternicht, G., Carr, E., Stafford Smith, M. 2020. *Why behavioral change matters to the GEF and what to do about it. A STAP Advisory Document*. Scientific and Technical Advisory Panel to the Global Environment Facility. Washington, D.C.

Toth, F., 2018. *Innovation and the GEF*. Scientific and Technical Advisory Panel to the Global Environment Facility, Washington, DC.

Annex A: terminal evaluations submitted

Terminal evaluation covered in APR – by Agency

GEF Agency	Up to APR2021	New submissions for APR 2023 ²	Cumulative total for APR 2023
ADB	18	8	26
AfDB	2	0	2
CI	2	10	12
EBRD	1	0	1
FAO	36	25	61
IDB	22	5	27
IFAD	29	0	29
IUCN	1	2	3
Joint – multi-Agency	71	0	71
UNDP	799	270	1069
UNEP	231	34	265
UNIDO	64	8	72
WB	528	78	606
WWF	2	1	3
All Agencies	1806	441	2247

² As of September 30, 2022.

Annex B. Performance Criteria and Rating Scales

The evaluators will rate project performance on the following criteria: outcomes, sustainability, implementation, execution, M&E design, and M&E implementation. The rated dimensions are described along with a description of the level of performance for a specific rating. In most instances, actual performance may not fully correspond to any of the rating descriptions. Therefore, a rating will be based on the description that provides the best fit based on the evidence. Where available evidence is insufficient to provide rate performance, the performance will be rated as unable to assess.

Outcome Rating

The overall rating of the project outcome will be based on the following criteria:

- a. **Relevance and coherence:** The evaluators will assess the extent to which the project outcomes aligned with the GEF focal areas/operational program strategies, country priorities, needs of the beneficiaries, and mandates of the Agencies. They will assess the extent to which the project is compatible with other relevant projects and programs being undertaken in the recipient country. The evaluators will assess if the project is well-targeted and the extent to which the project design is appropriate for delivering the expected outcomes. They will assess internal coherence by determining the extent to which there is alignment among the project's theory of change, governance structure, activities, and M&E system.
- b. **Effectiveness:** The evaluators will consider the extent to which project outcome achievements were commensurate with the *ex-ante* targets. They will weigh the extent to which the project made the expected level of contributions to global environmental benefits. They will also consider the overall progress in achieving the long-term objectives. They will also consider the unintended consequences of the project and the extent to which they add to, or negate, project benefits.
- c. **Efficiency:** The criterion is focused on the extent project was cost-effective in delivering its intended results. The evaluators will consider the project's cost/time versus output/outcomes equation, and, where feasible, compare it to alternatives. They will also consider the extent to which project activities were completed in a timely manner.

Project outcome rating will be based on the extent to which the expected outcomes were achieved, and the extent to which it was relevant and cost effective. A six-point rating scale is used to assess outcome. The top three ratings comprise the satisfactory range and the bottom three (excluding unable to assess) the unsatisfactory range:

- Highly satisfactory (HS): The outcomes exceed targets, and they are highly relevant and cost effective.
- Satisfactory (S): Level of outcomes achieved meets targets. The outcomes are relevant and cost effective.

- **Moderately Satisfactory (MS):** Level of outcomes achieved was generally close to the targets. Majority of the targets were met or almost met but some were not. The outcomes are generally relevant and cost effective.
- **Moderately Unsatisfactory (MU):** Overall, the level of outcomes achieved is lower than targets, although some outcomes were substantially achieved. The outcomes are generally relevant but not sufficient given the costs or alternatively generally cost-effective but not adequately relevant.
- **Unsatisfactory (U):** The expected outcomes were not achieved, or achievement was substantially lower than expected, and/or the achieved outcomes are not relevant. Alternatively, the outcome was cost ineffective compared to alternatives.
- **Highly Unsatisfactory (HU):** Negligible level of outcomes were achieved and/or the project had substantial negative consequences, that outweigh its benefits.
- **Unable to Assess (UA):** The available information does not allow an assessment of the level of outcome achievement

Sustainability Rating

The rating for likelihood of sustainability will be based on the probability of occurrence of a risk and the magnitude/severity of its effects on continuation of net benefits when it materializes. The assessment also considers resilience of the project benefit stream to the likely risks. The assessment will assess likelihood of continuation over a time frame reasonable for the given project. At the time of the evaluation, a project may not face the consequences of the risk materializing, or the risk may be just beginning to materialize. The assessment should be based on the evidence of risks, available at the time of evaluation. Most risks may be categorized as financial, sociopolitical, institutional, and environmental risks.

- **Financial resources.** The evaluators will assess the likelihood that financial resources will be available to continue the activities that sustain project benefits and risks associated to its availability. For example, support for income-generating activities that support environmentally friendly behavior, regular government budget allocations for the activities supported by the GEF project, and trends that suggest that in the future adequate financial resources for sustaining the project outcome will be available or conversely unavailable.
- **Sociopolitical.** The evaluators will assess the extent to which social or political risks may undermine the longevity of project outcomes. They will assess the extent to which the level of stakeholder ownership is insufficient to allow for project outcomes/benefits to be sustained. They will assess the extent to which the interests of key stakeholders are aligned to support continuation of the project benefit flow. They will assess the extent to which there is sufficient public/stakeholder awareness in support of the long-term objectives of the project.
- **Institutional framework and governance.** The evaluators will assess if the legal framework, policies, governance structures and processes pose any threat to the continuation of project benefits. While assessing these risks, the evaluators will consider if the required systems for accountability and transparency, and the required technical and institutional know-how, are in place.
- **Environmental.** The evaluators will assess if there are any environmental risks that can undermine the future flow of project benefits. The evaluators should assess whether certain activities in the project area will pose a threat to the sustainability of project outcomes. For example, project

outcome may be especially vulnerable to climate change risks. Similarly, biodiversity-related gains made by a project targeting marine protected areas may be affected by an increase in pollutant accumulation.

In providing an overall sustainability rating, other risks that are important but do not fall in these categories also need to be considered. Considering the probability of incidence of all relevant risks, and magnitude of effect/severity, the reviewer will provide a rating for the overall likelihood of sustainability using the following four-point scale:

- Likely (L). Either there is negligible risk to continuation of benefits or there are some risks, but the magnitude of their effect is too small and/or the probability that they will materialize is too small. Overall, it is likely that the net benefits of the project will continue.
- Moderately Likely (ML). There are some risks to sustainability, and they may have some effect on continuation of benefits if they materialize. However, probability of materialization of these risks is low. Net benefits are more likely to continue than abate.
- Moderately Unlikely (MU). There are significant risks to sustainability. The effect on continuation of benefits would be substantial if these risks materialize and the probability of materialization of these risks is significant. Overall, net benefits of the project are likely to abate.
- Unlikely (U). There are severe risks to sustainability. These risks have either already materialized and halted accrual of net benefits or have high probability of materialization and will halt accrual of net benefits when they materialize. Therefore, overall, it is unlikely that net benefits will continue to accrue, and the long term intended impacts of the project will be achieved.
- Unable to Assess (UA). Unable to assess the expected incidence and magnitude of risks to sustainability.

Implementation and Execution Ratings

The performance of the GEF Agency and of executing agency will be considered separately. A GEF Agency that implements a project is responsible for activities related to a project's identification, concept preparation, preparation of detailed proposal, project start-up, oversight and supervision, completion, and evaluation. The Agency is also, overall, responsible for efficient utilization of project inputs and delivery of project outputs. The performance of the GEF Agency will be considered to rate the quality of implementation.

GEF activities are executed on the ground by the executing agencies. The executing agencies are involved in the management and administration of the project's day-to-day activities under the overall oversight and supervision of a GEF Agency. The executing agencies are responsible for the appropriate use of funds, as well as the procurement and contracting of goods and services following the regulations of the GEF Agency. The performance of the project's executing agency/agencies will be considered to rate the quality of execution.

Scale for rating Implementation and Execution

Ratings	Implementation (GEF Agency)	Execution (executing agency/agencies)
Highly satisfactory (HS)	Performance of the GEF Agency was exemplary. Project preparation and	Performance of the executing agency/agencies was exemplary. The

	implementation were robust. The Agency ensured that the relevant GEF policies were applied in project preparation and implementation. Project supervision was strong – the Agency identified and addressed emerging concerns in a timely manner. The GEF Agency ensured that project implementation stayed on track and was completed in time.	execution of project activities was timely and of high quality. Relevant GEF policies and requirements were adhered to. Guidance from the GEF Agency was followed and corrective actions, if required, were taken promptly. The executing agency also undertook measures to mitigate risks to sustainability and is taking steps to support follow-up to the project. Completed project activities in time.
Satisfactory (S)	Performance of the GEF Agency met expectations and did not have any salient weakness. Project preparation and implementation were robust, and relevant GEF policies were applied. The GEF Agency supervised the project well – it identified and addressed emerging concerns in a timely manner. The GEF Agency ensured that project implementation was on track.	Performance of the executing agency met the expectations and was without any salient weakness. The execution of project activities was timely and of good quality. Relevant GEF policies and requirements were applied. Guidance from the GEF Agency was followed. The executing agency also undertook measures to mitigate risks to sustainability of project outcomes.
Moderately Satisfactory (MS):	Overall, the performance of the GEF Agency met expectations. Project preparation and implementation were adequate and relevant GEF policies were applied although there are some weak areas. The GEF Agency supervised the project adequately – it identified and addressed emerging concerns although some concerns may be inadequately addressed. Project implementation had minor delays and may have had a few dropped activities.	Performance of the executing agency had some weaknesses but, overall, it met the expectations. The execution of project activities was generally timely but with some instances of delay. Relevant GEF policies and requirements were applied although some minor slip-ups may also have been observed. Guidance from the GEF Agency was followed and problems were fixed. There are some areas where the performance of the executing agency was below par, although overall the executing agency's performance was adequate.
Moderately Unsatisfactory (MU):	Overall, the GEF Agency did not meet expectations although there were some areas of solid performance. Project preparation and implementation had weaknesses although these were not too severe. Project supervision was somewhat weak. Although most emerging concerns were identified, many remained unaddressed or inadequately addressed. Project implementation was delayed, and a few activities were	While there were some areas of solid performance, the overall performance of the executing agency did not meet expectations. The execution of project activities was delayed. The observed capacities of the executing agency were a limitation of project execution. Several slip ups in application of GEF policies and requirements were observed. Guidance from the GEF Agency was generally followed and problems were fixed but usually such actions were not timely.

	dropped or reduced in scale because of issues that were largely under the control of the GEF Agency.	There are several areas for improvement in execution.
Unsatisfactory (U):	The GEF Agency did not meet the expected level of performance. Project preparation and implementation were weak. Emerging concerns were not identified by the GEF Agency in time and remained unaddressed or inadequately addressed. M&E implementation was weak – activities were not implemented in time or were not undertaken. Project implementation was delayed, and several activities were dropped or were reduced in scale.	The executing agency did not meet expectations. Execution of project activities was delayed and at least some activities were dropped due to factors largely under the control of the executing agency. Many slip-ups were observed in application of GEF policies and requirements. Guidance from the GEF Agency was not put into practice or was applied with considerable delay.
Highly Unsatisfactory (HU):	There were severe shortcomings in the quality of implementation. The GEF Agency mismanaged project implementation and its supervision was poor. Emerging concerns were not identified in time, including those that should have been obvious. Although instances of mismanagement were discovered, corrective actions were not undertaken. Project activities were poorly implemented, and several had to be dropped.	There were severe shortcomings in project execution. There were several instances of mismanagement. Emerging concerns were not addressed in time, including those that should have been obvious. Most activities were very poorly executed, experienced delays, and had activities dropped. GEF policies and requirements were not applied.
Unable to assess (UA)	The available information is not sufficient to allow rating of performance.	The available information is not sufficient to allow rating of performance.

Project M&E Ratings

The M&E arrangements will be rated at the project level. This will include both M&E arrangements vested in the coordinating project, and arrangements at the child project level to contribute to project M&E. The quality of project M&E will be assessed in terms of:

- **Design:** The review will assess quality of the M&E plan at CEO endorsement/approval. It will consider the extent to which the M&E plan was practical and well-thought through. It will assess the extent to which the M&E plan addresses the project's theory of change, GEF M&E requirements, incorporates applicable core indicators and tracking tools, and provides baseline information. It will discuss whether the indicators specified to track environmental, gender, socio-

economic, and other results, are appropriate (SMART³). For child projects and coordinating projects under a programmatic framework, the review will assess how the well M&E plan aligns with and is likely to contribute to the program M&E plan.

- **Implementation:** The review will assess the extent to which the M&E system operated as planned. Where applicable, it will consider if weaknesses in the M&E plan were addressed in time. It will consider if data on specified indicators was gathered systematically and as per schedule. It will consider the extent to which data on relevant GEF core indicators / corporate results indicators and/or tracking tools was analyzed and reported. It will consider the extent to which the methodological approaches used to analyze data were appropriate. It will consider the extent to which resources allocated for M&E were sufficient. It will also consider the extent to which the information from M&E system was used to improve project implementation and effectiveness. For child projects (including coordinating child projects) under a programmatic framework, the review will assess how the well M&E activities of the project aligned with and contributed to the program M&E.

Quality of M&E on these two dimensions will be assessed separately on a six-point scale:

Ratings	M&E Plan	M&E Implementation
Highly satisfactory (HS)	The project M&E plan is a good practice and did not have any weaknesses – its alignment with the project theory of change is robust. Complete baseline data has been provided. The specified indicators were appropriate, and arrangements for the M&E plan implementation were adequate. Overall, the M&E plan exceeds expectations and is exemplary.	The M&E plan implementation was excellent. Weaknesses in the M&E plan, if present, were addressed promptly. M&E activities were conducted in a timely manner, and data from M&E was used to improve project implementation. Overall, M&E implementation exceeded expectations and was exemplary.
Satisfactory (S)	The project M&E plan was robust and did not have any or had only minor weaknesses – the alignment with the project theory of change is robust. Baseline data provided or its collection is planned at project start. The specified indicators were appropriate, and arrangements for M&E plan implementation were adequate. The plan meets expectations.	The M&E plan implementation was generally as per the plan. Weaknesses in M&E were addressed in a timely manner. M&E activities were conducted in a timely manner, and data from M&E was used in improving project implementation. Overall, M&E implementation meets expectations.
Moderately Satisfactory (MS):	On balance, the project M&E plan was solid. The specified indicators were generally appropriate, and	The M&E plan implementation was generally as per the plan. Weaknesses in M&E were generally addressed

³ SMART: Specific, Measurable, Achievable/Attributable, Relevant/Realistic, and Time-bound, Timely, Trackable and Targeted.

	arrangements for M&E plan implementation were adequate. The alignment of the M&E plan with the project theory of change is solid. There were areas where the M&E plan could be strengthened but, overall, the plan was adequate.	although some weaknesses remained. Some M&E activities were delayed. M&E data was used for reporting but had little use in improving project implementation. Overall, M&E implementation meets expectations with some areas of low performance.
Moderately Unsatisfactory (MU):	Overall, a weak M&E plan although it had strengths in some areas. The specified indicators were generally appropriate but additional indicators were required to adequately capture project results and/or arrangements to gather data on indicators were not adequate. The alignment with the project theory of change is somewhat weak. The plan needs several improvements to meet expectations.	The M&E plan implementation was weak and/or did not address the weaknesses in the M&E plan. Most M&E activities were completed – some of them were either dropped or delayed. M&E data was not reported in a timely manner – there is little evidence to suggest that the data was used to improve project implementation. Overall, M&E implementation does not meet expectations although there are some areas where the performance is adequate.
Unsatisfactory (U):	The M&E plan had severe shortcomings. The alignment with the project theory of change is weak. No baseline data was provided nor any indication that it would be collected at project start. Indicators do not adequately address project outcomes and other results; for several results, relevant indicators have not been specified. There are gaps in arrangements for M&E plan implementation – no budget or an inadequate budget was provided for M&E.	The M&E plan implementation was flawed and/or did not address severe weaknesses of the M&E plan. Several M&E activities were either dropped or were incomplete. The data collection methodology was not sound. M&E data was not reported in a timely manner – there is little evidence to suggest that the data was used to improve project implementation. M&E implementation does not meet expectations.
Highly Unsatisfactory (HU):	No M&E plan was prepared.	No, or negligible, M&E activity was implemented other than conduct of the project evaluation.
Unable to assess (UA)	Unable to assess because project documents are not available.	Un Unable to assess as the terminal evaluation does not cover M&E implementation adequately.

Annex C. Terminal Evaluation Report Validation Guidelines

Introduction

1. The GEF Monitoring and Evaluation Policy (2019) requires that the GEF Agencies will conduct terminal evaluation of GEF funded projects at the end of implementation. Terminal evaluations are expected to provide a comprehensive and systematic account of the performance of a completed project. These evaluations assess a project's design, implementation, results, M&E arrangements, compliance with GEF policies, and lessons. The GEF Agencies submit terminal evaluations to the GEF IEO through the GEF Portal. The projects covered include stand alone projects and projects approved within the framework of a program. These include full size projects, medium size projects and enabling activities that were approved using non-expedited approaches.
2. The GEF IEO validates some of the terminal evaluations to ensure consistency in ratings used in portfolio analysis, and to provide feedback to Agencies on quality of terminal evaluations⁴. 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. These guidelines have been prepared for internal use by the GEF IEO for validation of the terminal evaluations.
3. These guidelines will be used by primary reviewers, peer feedback providers, and supervisors involved in conduct of validations. A standardized validation report template that is updated annually will be used to prepare the validation report. Finalized terminal evaluation validation reports and dataset based on these reports will be publicly available at the GEF IEO website.

Validation Process

4. A validation report reflects work of a three-member team, with each member having different roles in the validation process. A primary reviewer of a terminal evaluation prepares its validation report, including draft versions of the validation report. A peer feedback provider reviews drafts of the validation report and provides feedback to the primary reviewer. The primary reviewer addresses the feedback provided by the peer reviewer on various versions of the draft of the validation report. A supervisor maintains oversight of the validation process and, where necessary, help in resolving the differences between the assessments of the primary reviewer and the peer feedback provider. The supervisor also reviews a draft validation report when a rating provided by the primary reviewer differs with the corresponding rating provided by the GEF Agency by two points/grades. Once there is agreement among the primary reviewer, peer reviewer, and the supervisor, on the analysis and ratings presented in the validation report, it is considered final.

⁴ GEF IEO validates only a sample of terminal evaluations that are either prepared or independently reviewed by evaluation units of the GEF Agencies. GEF IEO validates all terminal evaluations prepared by the operational units of Agencies that do not undergo an independent review.

Sources of Information

5. Terminal evaluation validations will be primarily based on the information presented in terminal evaluation reports. However, other documents such as project implementation reports (PIRs), mid-term reviews (MTRs), and documents submitted during a project's appraisal, will also be reviewed to find corroborating evidence and – in some cases – to fill the information gaps. Where available, a reviewer will also consider other independent sources of information including peer review papers that may provide additional information relevant to assessment of a project's performance. At the minimum, the reviewer will search for such papers using search engines for scholarly publications and conduct a regular web search to identify relevant publications that may be useful. Where a project is expected to produce knowledge products such as publicly available reports, reviews, plans, websites, and datasets, reviewers will access these websites and documents as these may be an additional source of information. Different sources of information may be most reliable for different data. The reviewer will use the most reliable source for any given data point, and where possible triangulate the information. The reviewer will indicate the information sources used (including documents reviewed) in the appropriate section of the validation report.

Contents of the validation report

6. The sections covered in the terminal evaluation validation template includes the following: Project data; summary of project ratings; project objectives and theory of change; outcome and sustainability; project impacts; assessment of processes and factors affecting attainment of project outcomes; project M&E; implementation and execution; lessons and recommendations; and quality of terminal evaluation. The validation report template is updated annually. Nonetheless, the main topics covered by the validation report, including rated criteria, will remain the same to maintain continuity and facilitate comparisons among validations carried out across different periods.

Validated performance ratings

7. Primary reviewers will rate project performance on following criteria: outcomes, sustainability, implementation, execution, M&E design, and M&E implementation. The performance will be rated using an approach identical to that described in the terminal evaluation guidelines for full size projects. In addition to the rating project performance, primary reviewers will also rate quality of terminal evaluation reports. Ratings will be provided only in instances where there is sufficient evidence to allow an assessment of the level of performance.

Quality of Terminal Evaluation Reports

8. The quality of a terminal evaluation report will be assessed using 14 quality criteria⁵. Table 1 describes these criteria and presents sub-criteria that need to be considered along with other relevant information that may be available to help assess performance on the given criteria. The performance on

⁵ Up to 2021 GEF IEO has used a different set of criteria for assessing quality of terminal evaluation reports. The revised criteria are inclusive of the criteria used earlier. This will allow for comparison with past ratings of quality of terminal evaluations.

each of the 14 criteria will be rated on a six-point scale (Highly Satisfactory to Highly Unsatisfactory). Overall quality of the report will be rated based on average of the ratings on the 14 quality criteria.

Table 1. Terminal Evaluation Quality Criteria

Criteria/indicators of terminal evaluation quality
1. Timeliness: terminal evaluation was carried out on schedule and its report submitted on time. • Terminal evaluation conducted within six months before or after project completion • Terminal evaluation report submitted at the GEF Portal within 12 months of project completion
2. General information: Provides general information on the project and evaluation. • Provides GEF project ID • Lists evaluators that conducted the terminal evaluation • Lists the executing agencies • Specifies key project milestones (start date, first disbursement date, completion date) • Lists GEF environmental objectives • For projects under a program - identifies parent program
3. Stakeholder involvement in evaluation: Participation of key stakeholders sought and their feedback addressed. • Key stakeholders of the project were identified in the report • Feedback of key stakeholders was sought on the draft report • Feedback of key stakeholders was incorporated in finalization of the evaluation report • If national project, OFP Feedback was sought on the draft report of the evaluation • If national project, OFP feedback was incorporated in finalization of the report
4. Theory of change: provides solid account of the project's theory of change. • Discusses causal links/mechanisms to achieve intended impact • Presents the key assumptions of the theory of change • Discusses whether the key assumptions remain valid
5. Methodology: Provides an informative and transparent account of the methodology. • Discusses information sources for the evaluation • Provides information on who was interviewed • Provides information on project sites/activities covered for verification • Tools and methods used for the evaluation are described • Identifies limitations of the evaluation
6. Outcomes: Provides a clear and candid account of the achievement of project outcomes. • Assesses relevance to GEF priorities • Assesses relevance to country priorities • Assesses relevance of project design • Reports performance on all outcome targets • Discusses factors that affect outcome achievement at sufficient depth • Reports on timeliness of activities • Assesses efficiency in using project resources • Discusses factors that affected efficiency in use of resources

7. Sustainability: Presents realistic assessment of sustainability. • Identifies risks that may affect sustainability • Indicates likelihood of key risks materializing • Indicates the likely effects if key risks materialize • Indicates overall likelihood of sustainability
8. Monitoring and Evaluation: Presents sound assessment of the quality of the project M&E system. • Analyzes quality of M&E design at entry • Analyzes quality of M&E during implementation • Discusses use of information from the M&E system for project management
9. Finance: Reports on utilization of GEF funding and materialization of co-financing. • Reports on utilization of GEF resources • Provides data on materialized cofinancing • Provides data on sources of materialized cofinancing • Provides data on types of cofinancing (cash, in-kind; loan, grant, equity, etc) • Discusses reasons for excess or deficient materialization of co-financing • Discusses contributions of cofinancing to project results, including effects of excess or deficient materialization of co-financing
10. Implementation: Presents a candid account of project implementation and Agency performance. • Provides account of the GEF Agency performance • Provides account of the performance of executing agency • Discusses factors that affected implementation and execution • Discusses how implementation and execution related challenges were addressed
11. Environmental and Social Safeguards, and Gender: Discusses application of safeguards and gender analysis. • Reports on implementation of social and environmental safeguards • Reports on conduct of gender analysis • Reports on implementation of actions specified in gender analysis
12. Lessons and recommendations: based on project experience and relevant to future work. • Presents lessons • Lessons are based on project experience • Discusses applicability of lessons • Presents recommendations • Recommendations specify clearly what needs to be done • Specifies action taker for recommendations
13. Performance Ratings: Ratings are well substantiated by evidence, and are realistic and credible. • Ratings are supported with sufficient evidence • Evidence provided in support is credible
14. Report presentation: The report was well written, logically organized, and consistent. • Report is written in English (as required by the terminal evaluation guidelines) • Report is easy to read • Report is well-organized • Report is consistent

- Report makes good use of tools that make information accessible (graphs/charts/tables)

Overall Quality of the report: The 14 terminal evaluation quality criteria will be rated on a six-point scale (HS=6, S=5, MS=4, MU=3, U=2, HU=1). The overall quality will be determined by calculating the average the ratings on the 14 criteria and rounding off to the nearest digit. If the average is 5.5, 4.5, 3.5 and so on, it will be rounded off upwards.