

59th GEF Council
December 8–10, 2020
Virtual Meeting

**REPORT ON THE REVIEW OF THE GEF TERMINAL EVALUATION
VALIDATION PROCESS**

(Prepared by the Independent Evaluation Office of the GEF)

LIST OF ABBREVIATIONS

ADB	Asian Development Bank
AfDB	African Development Bank
APR	Annual Performance Report
BOAD	West African Development Bank (Banque Ouest Africaine de Developpement)
CAF	Development Bank of Latin America (Corporación Andina de Fomento)
EBRD	European Bank for Reconstruction and Development
FAO	Food and Agriculture Organization
GEF	Global Environment Facility
GEF-IEO	Global Environment Facility Independent Evaluation Office
ICR	Implementation Completion and Results Report (World Bank)
IDB	Inter-American Development Bank
IEG	Independent Evaluation Group (World Bank)
IFAD	International Fund for Agricultural Development
KII	key informant interview
M&E	monitoring and evaluation
MDB	multilateral development bank
PIR	Project Implementation Review
QA	quality assurance
TE	terminal evaluation
ToR	terms of reference
UNDP	United Nations Development Programme
UNDP-IEO	Independent Evaluation Office (UNDP)
UNEG	United Nations Evaluation Group
UNEP	United Nations Environment Programme
UNIDO	United Nations Industrial Development Organization

TABLE OF CONTENTS

List of abbreviation	i
Executive Summary.....	iv
Methodology.....	iv
The GEF-IEO TE Reviews and Validations	iv
Findings	v
I. Introduction	1
II. Goal And Scope of the Review	1
III. Methodology.....	2
IV. The GEF-IEO Terminal Evaluation Validation Process	4
V. Key Findings	6
VI. Final Remarks.....	28
Annex 1: Suggested areas for GEF-IEO future interventions.....	31
Annex 2: Assessment Matrix.....	34
Annex 3: List of TE Validations Reviewed	36
Annex 4: List of Key Informants	37
Annex 5: List of Key Documents and Websites Reviewed.....	38
Annex 6: Terminal Evaluation Report Review Guidelines	40
Annex 7: Terminal Evaluation Review form – Section “Quality of the Terminal Evaluation Report”	47
Annex 8: Interview Guide – for Agencies.....	48
Annex 9: Results of the GEF-IEO Validations and Reviews for the 12 Projects Analyzed	49
Annex 10: UNEP Matrix (example of one criteria).....	53
Annex 11: Terms of Reference.....	56
Short-Term Consultant for the review of the Terminal Evaluation Review Process.....	56

LIST OF TABLES AND FIGURES

TABLES

Table 1: Sample TE Validation Reports reviewed per Agency and Region	3
Table 2: Source of APR ratings (2017–20).....	10
Table 3: GEF-IEO TE Reports validated per year	11
Table 4: Comparison between TE and IEO Review Ratings (selected TEs)	15

FIGURES

Figure 1: Evolution of GEF Partnerships 1990s–2020.....	5
Figure 2: Results of TE Report Validation by the IEO (2017–20)	7
Figure 3: Quality of UNIDO, FAO, and IFAD Evaluations as rated by the GEF-IEO in the 2017–20 period .	12
Figure 4: Quality of remaining TEs as rated by the GEF-IEO in the 2017–20 review period	14

EXECUTIVE SUMMARY

1. This document presents the results of the independent Review of the Global Environment Facility (GEF) Terminal Evaluation (TE) Validation Process, conducted from July to September 2020. Currently, the GEF Independent Evaluation Office (GEF-IEO) validates all TEs submitted by nine GEF Agencies and a sample of TEs from the World Bank, the United Nations Development Programme (UNDP), and the United Nations Environment Programme (UNEP). To conduct the validations that are the object of this review, consultants and GEF-IEO staff assess the TEs and provide ratings against a detailed guideline and form on the project outcome, sustainability, implementation, and monitoring and evaluation, through a desk review of several other available related documents, including the Project Identification Form, , the Mid-Term Review, the Implementation/Project Completion Report, and the Agency’s own validation report when this exists. A TE validation review report is produced for each TE report.

Methodology

2. The methodology for the present review consisted of desk reviews, including the review of a sample of 12 GEF-IEO TE Validation Reports, and key informant interviews with selected GEF Agencies. The Assessment Matrix in Annex 2 contains details about the lines of evidence and indicators used. The analysis focused on GEF-IEO TE validation activities and on validated reports of projects completed since October 2017, which were reported as a cohort in the 2020 Annual Performance Report (APR). The focus was on answering the questions in the Terms of Reference (Annex 11):

- (a) To what extent is the terminal evaluation validation process transparent?
- (b) To what extent does the validation process enhance the credibility of ratings based on the evidence provided in the terminal evaluations?
- (c) How does the GEF IEO’s terminal evaluation validation process compare with that of other international development agencies?
- (d) What are the good practices and lessons that may be incorporated in the GEF-IEO’s validation process?

The GEF-IEO TE Reviews and Validations

3. The GEF-IEO TE Review and Validation Process is a well-established central element of the GEF’s APRs, which are a key mechanism of GEF accountability to the GEF Council members since 2004. The Reviews and Validations done by the GEF Independent Evaluation Office (GEF-IEO) ensure that project performance ratings across all GEF Agencies are comparable, consistent, and evidence based, for the purposes of APR reporting.

4. GEF Agencies which have been participating in the GEF-IEO TE Validations and Reviews have been evolving and increasing in number in the past few years. From an initial pool of 3 key partner Agencies (the World Bank, the UNDP, and the United Nations Environment Programme

(UNEP), the group evolved to about 5 Agencies with the addition of the Asian Development Bank (ADB) and the United Nations Industrial Development Organization (UNIDO) in 2006–11, and more recently to about 12 Agencies (with the addition of another 7 Agencies) which now have projects completed and TEs validated out of the current total of 18 institutions accredited as GEF Agencies.

5. The focus of this Review is on TE Validation reports prepared for TEs submitted in 2017–20 by 10 Agencies¹; these reports were used as the source of the reported ratings in the 2020 APR. Some findings also relate to the other three Agencies who currently do not undergo the GEF-IEO validation process: the UNDP, UNEP, and the World Bank.

Findings

6. The establishment of minimum evaluation requirements and the work undertaken in the early years of the GEF contributed to improvements in evaluation quality. According to key informant interviews (KIIs), the GEF-IEO TE guidelines and the parameters initially established to validate the TE ratings made important contributions to the quality of evaluations undertaken by the GEF Agencies participating at that time. Overall, during 2017–20 period, the large majority of the evaluations (92 percent) were rated Satisfactory and/or Moderately Satisfactory by the GEF-IEO TE Review and Validation process. The Guidelines for GEF Agencies in Conducting Terminal Evaluation for Full-sized Projects (2017 Guidelines) and other guidance provided for the conduct of evaluations and the preparation of these TE reports have been useful for the Agencies.

7. Well-established processes exist to ensure that the TE reports are submitted by the GEF Agencies and the validated ratings enable statistical analysis in the APR, and they inform other GEF-IEO thematic evaluation and resource allocation work. A total of 134 completed projects for which TEs were submitted during the 2019 calendar year were reported for the first time in the 2020 APR. Because the volume of added data from additional cohorts is relatively small, the impact of this new data on the overall dataset produces limited deviation from previous years. Every year, the overall APR results change little because the number of new TEs in is small in comparison with the overall database, and as such, the APR has been also incorporating more qualitative analysis in recent years.

8. The APR dataset is important to ensure availability and comparability of statistical data on the GEF performance. The data set is considered critical, but the need for GEF-IEO reviews and validations of TEs has been decreasing with improvements in quality. Initially, the GEF-IEO TE Validations were responsible for ensuring consistency across all GEF Agencies; they contributed to reducing the data gaps supplying it when it had not been provided in the TE reports prepared by the various implementing Agencies. As the TE data provided by the three partner agencies (UNDP, UNEP and World Bank) improved and the gaps in information related to GEF evaluation criteria reduced, their TEs were considered suitable to meet GEF information

¹ The African Development Bank, the Food and Agriculture Organization, the International Fund for Agricultural Development, Inter-American Development Bank, Asian Development Bank, the World Wildlife Fund, Conservation International, European Bank for Reconstruction and Development, the United Nations Industrial Development Organization, and the World Bank.

needs and are now only validated for quality assurance purposes and to ensure consistency across the GEF Agencies. Currently, only a smaller portion of the validated data for the 2020 APR datasets —about 24 percent of the total—is based on the IEO TE validation process.

9. The large majority of the TE ratings validated by the GEF-IEO are for projects from three GEF Agencies with established Evaluation units or GEF Agencies which may be well positioned to “graduate” soon. About 17 percent of the total GEF-IEO–validated reports in the period analysed were submitted by three Agencies—UNIDO, the Food and Agriculture Organization (FAO), and the International Fund for Agricultural Development (IFAD) —and considered high quality TEs; only 4 percent of these reports were rated by the GEF-IEO as Moderately Unsatisfactory, and the remaining 96 percent were rated Satisfactory, Highly Satisfactory, or Moderately Satisfactory. The KIIs confirmed that these evaluations are produced, or the quality assured by evaluation offices of these GEF Agencies, which were engaged in the TE preparation processes and final reports. The evaluation offices of all three Agencies are independent from management, according to KIIs and the Agencies’ documentation.

10. A relatively small share of the TEs validated in the evaluation period are for projects implemented by the multilateral development banks and nongovernmental organizations (NGOs), and capacity development and guidance are likely to help reduce existing data gaps and improve TE quality. About 10 percent of the total TEs validated in the period were submitted by five Agencies (the Asian Development Bank, Conservation International, the Inter-American Development Bank, the World Bank, and the World Wildlife Fund) and one was a Joint Evaluation conducted by more than one Agency. The quality of these evaluations, as rated by the GEF-IEO, is also good; about 89 percent of them were rated Moderately Satisfactory or Satisfactory and only 11 percent were rated Unsatisfactory. The majority of the evaluation data gaps currently found in the sample of the TEs reviewed are from multilateral development banks (Asian Development Bank, Inter-American Development Bank, World Bank, and European Development Bank to a lesser extent), mostly associated with criteria such as “Monitoring and Evaluation Design,” “Monitoring and Evaluation Implementation,” “Implementation Quality,” and “Execution Quality.” The validation process was therefore instrumental to filling this gap and providing the ratings in these categories for those evaluations missing this information. This suggests that more guidance and better forms or ToRs provided to multilateral development bank consultants prior to the conduct of the evaluations could help to reduce the data gaps.

11. As for the validation process, except for UNDP, UNEP and the World Bank, there is limited awareness among the GEF Agencies’ staff interviewed about the validations. Information about the validation process is not easily accessible or complete. Only a few GEF Agencies interviewed were aware of the existence of the TE validations, primarily through the GEF-IEO websites and exposure to the APR, via the web, or by participation in the GEF Council. The large majority of Agencies who have had their TE reports validated in the period analyzed, were not only unaware of the existence of the datasets in the GEF-IEO website; they also were unaware that the data related to their TEs had been validated and the validation results were publicly released in the APR datasets. Most Agencies were also unaware of the existence of TE

Validation Reports and had never seen them; these reports apparently have been posted in the PMIS but may not yet have been migrated to the new GEF Portal.

12. The TE Reviews are well substantiated and aligned with criteria, but the variations in the evidence presents challenges to ensuring evaluative judgments are consistent across reviewers. The GEF-IEO desk validations use analysis of the TE reports, the Project Implementation Reviews, and Mid-Term Reviews, and the project approval document as the basis for the validation assessments. In the 12 TE Reviews analyzed, the types, numbers, quality, and depth of these documents provided to the reviewers were quite varied and not consistent across all TE validations. The content of these documents also may not provide updated and unbiased data upon which to conduct validations, because the Project Implementation Reviews are descriptive documents prepared by project managers (team leaders) and the Mid-Term Reviews are usually produced two to three years prior to the TEs.

13. The quality of the TE is key to demonstrating the credibility of the process. Currently the quality of the large majority of the TEs validated by the GEF-IEO is assessed as Moderately Satisfactory or higher, which suggests that the evidence, though limited, is credible for assessments. The discrepancy between IEO validated ratings and Agency ratings on TEs is about 6 percent.

14. The criteria and ratings scales used by the GEF-IEO for validation are aligned with good practices, but there is scope for enhancing the assessments and using more qualitative information. There are key differences in the GEF-IEO assessment of the quality of the TEs, which take a bird's eye view as compared with the in-depth quality assessments used by other Agencies. This is explained by the fact that the IEO does a second level of validation in most cases. Most international Agencies use similar criteria in their assessments (e.g., Outcomes, Relevance, Effectiveness, and Sustainability), and the ratings systems are also similar.

15. The GEF Agencies have developed several good practices that could be shared and disseminated across all GEF Agencies, including quality assessment checklists to be used prior to GEF submission of TEs, better data collection and reporting for GEF projects, ToRs, and templates to ensure that GEF submissions are complete.

16. Validation work similar to the GEF-IEO validations is undertaken by other international agencies, and practices used in the World Bank Independent Evaluation Group (IEG) Reviews of the Implementation Completion and Results Report Reviews (ICRs) and the UNDP-IEO GEF-TE Validations could be considered for potential application in the GEF-IEO validations to elaborate on GEF-specific implications. These include adding mechanisms to desk reviews to enhance their credibility, such as broadening the lines of evidence used, tightening the quality assurance process, and enhancing consistency of judgments through more in-depth guidance on assessment criteria. Structured feedback could be provided to reviewers on the results of their validation or QA processes with the goal of enhancing learning, promoting better evaluation work, and improving the consistency of evaluative judgments across TE assessments.

17. International Agencies are enhancing systems to gather more qualitative data on their evaluations, investing in searchable databases that can quickly identify and retrieve data on substantive evaluation findings, recommendations, and lessons for mining purposes. They are also making this information widely available. In addition to quantitative data, the write-ups (qualitative data) of the World Bank ICR Reviews are stored in a searchable database within IEG and, for all projects that closed from fiscal year (FY)2011 onward, are posted on IEG's external website. The data are often useful as a starting point for IEG's ICR reviewers as a quick way to identify projects of different types, or with specific objectives or activities, when preparing to undertake larger country, sector, or thematic evaluations.

18. Independent evaluation quality assurance systems in which UN Agencies are formally engaged can be a good source of documentation for GEF-IEO to monitor performance of some of the GEF Agencies' evaluation units. This includes external assessment of UN entities undertaken by the UN-Secretariat Inspection and Evaluation Division of the Office of Internal Oversight Services and the UN Evaluation Dashboard. Because evaluation is one of the indicators against which the UN agencies report against the United Nations System-Wide Action Plan,² several UN Agencies also undertake independent assessments of the quality of their evaluations,³ Ongoing independent assessments of evaluation quality are being used by various agencies as a way to benchmark their evaluation functions and show adherence to highest evaluation standards.

Final Remarks

19. Overall, this Review considers that the GEF-IEO TE validations and reviews are currently at a turning point. Improvements have been made over the years with the building of an important dataset upon which to report on the GEF performance results. Over the past few cycles, the statistical analysis presented in each APR has been adding little interesting insight because the newly added cohort is so small in relation to the overall portfolio and possibly because there is little variation in performance over time.

20. More recently, following the graduation of the three larger Agencies (UNEP, UNDP, and the World Bank), improvements have also been made to TEs by UNIDO, FAO, and IFAD, which are now well positioned to graduate. The validation efforts could focus on a sample of their Agency portfolios for quality assurance purposes, as is done for the three larger Agencies. This would free up resources at the GEF-IEO with no impact on data quality for the purposes of the APR.

21. The analysis in the present Review suggests that about 3 percent of the missing data is from the remaining 10 percent of the portfolio (approximately), which mostly includes the multilateral development banks and the NGOs⁴; consequently, if the missing data could be

² For more details see <https://www.unwomen.org/en/how-we-work/un-system-coordination/promoting-un-accountability>

³ The first phase of UN-SWAP implementation focused on gender mainstreaming and planning; the second phase is designed to focus on results and includes monitoring activities and outcomes for gender-related Sustainable Development Goal results.

⁴ This is a rough estimate, based on the fact that 4 out of 12 terminal evaluations analyzed in this Review (about 30 percent) had a large portion of ratings missing. Considering that these evaluations were conducted by the Asian Development Bank,

simply be noted as “blank” or “missing,” this is likely to have limited overall impact on the APR cohort analysis, especially considering that new data in each APR (relative to the overall dataset) contain limited deviations from previous years. Validation efforts could then also perhaps only target a sample of the portfolio for quality assurance purposes, freeing up resources even further.

22. Because data gaps are relatively small and the evaluation quality is generally good, there are opportunities to address gaps and ensure more completed TEs through more targeted guidance to this group of Agencies. Opportunities may also exist for targeted capacity building to the newer Agencies (which have not yet submitted TEs⁵). It is important to ensure that from the outset, guidance is provided for the development of evaluation ToRs and the GEF-IEO expectations as to the quality of the TEs, which will likely result in fewer gaps in the data and better quality of the evaluation products. The validation reports could be more consistently shared and discussed with the TE Agencies as a way to provide feedback for future TE improvements.

23. The GEF-IEO could also support evaluation functions in the Agencies through the building of communities of practice or groups to provide targeted capacity building where issues exist, as well as formalizing informal relationships with the GEF Agencies.

24. The GEF-IEO could also seriously consider investing in and further developing a database of qualitative experiences, with searching capabilities. The existence of more qualitative data related to findings, recommendations, and lessons learned that are included in evaluations would likely enable the GEF-IEO also move forward (as other agencies are doing) toward mining qualitative data, which could perhaps be more valuable for GEF-IEO thematic, sectoral, and country-level evaluations.

25. As the validation and quality assurance processes evolve and data are used for GEF-IEO Evaluations (e.g., country evaluations, thematic, or strategic country clusters), opportunities may exist for GEF-IEO TE validation teams to participate in these other evaluation missions, cross-checking TE and validation information as well as getting feedback from country missions on the validated TEs and enhancing synergies across the work of various GEF-IEO teams.

Inter-American Development Bank, and the World Bank, they make up about 4 percent of the overall terminal evaluations rated in the 2017–20 period.

⁵ The size of their portfolio has not been confirmed by this review.

I. INTRODUCTION

1. The GEF TE review and validation process was instituted in 2004 as a mechanism to assure the quality of the evaluations undertaken by accredited Agencies who partner with the GEF in the implementation of a large portfolio of projects and programs.
2. When the implementation of a project is complete, GEF Agencies prepare a terminal evaluation to report on implementation experience and project performance. The GEF Independent Evaluation Office (IEO) validates all the TEs from 15 Agencies and a sample of TEs from the three original agencies, United Nations Development Programme (UNDP), the United Nations Environmental Programme (UNEP), and the World Bank. For the reviews or validations that are the focus of this Review, consultants and GEF-IEO staff assess the TEs and assign ratings against a detailed guideline and form, on the project's outcome, sustainability, implementation, and monitoring and evaluation (M&E), through a desk review of several other available related documents, including the Project Document, Mid-Term Review, the Implementation/Project Completion Report, and the Agency's own validation report when this exists. A TE Review (TER) is produced for each TE report.
3. This document presents the results of the Review of the Global Environment Facility (GEF) Terminal Evaluation (TE) Review Process, conducted from July to September 2020. The report contains six sections, including this Introduction. Section 2 briefly describes the goal and scope of this Review. Section 3 contains a quick overview of the methodology used. Section 4 describes the GEF-IEO TE Review Process and section 5 highlights the key findings of this Review. Section 6 and Annex 1 contain some suggestions for GEF-IEO consideration regarding the future of the GEF-IEO Review Process.
4. The validation data are then included in a dataset containing the ratings obtained by all TEs (including the ratings provided by UNDP, UNEP, and the World Bank) and the GEF-IEO undertakes aggregate analysis to identify performance trends, correlations, and causal relationships. The results of the analysis and qualitative information from the terminal evaluations are presented annually in the GEF IEO's Annual Performance Report (APR).⁶

II. GOAL AND SCOPE OF THE REVIEW

5. The goal of this Review is to identify gaps in the GEF-IEO's validation process and measures that may be used to strengthen the process, along with the cost implications of these measures. The Review seeks to answer the following questions from the Terms of Reference (ToRs):
 - (a) To what extent is the terminal evaluation validation process transparent? ^[]_[SEP]
 - (b) To what extent does the validation process enhance the credibility of ratings ^[]_[SEP] based on the evidence provided in the terminal evaluations? ^[]_[SEP]

⁶ Some analysis is presented on an annual basis; detailed analysis is presented in a four-year cycle aligned with the GEF replenishment.

- (c) How does the GEF IEO's terminal evaluation validation process compare with that of other international development agencies?
- (d) What are the good practices and lessons that may be incorporated in the GEF IEO's validation process?

6. The GEF-IEO is currently undertaking an Evaluation of Agency Self-Evaluation Systems. This Review will contribute to the Evaluation by highlighting aspects and identifying adjustments where needed in the GEF-IEO TE Review Process.⁷ This Review is primarily intended for the GEF-IEO, but it is also likely to be useful for the GEF Council and the GEF partners including the GEF Secretariat, the GEF Partner Agencies, the Scientific and Technical Advisory Panel, and the GEF-Civil Society Organization Network.

III. METHODOLOGY

7. The methodology to be used for this Review was designed to meet the requirements and expectations set up in the ToR, allowing for data gathering to answer the questions in the ToR. Data were analyzed using qualitative methods, subjective assessments, informed judgment, and expert opinion. Quantitative methods were not used.

8. The analysis focused on GEF-IEO TE validation activities and validated reports of projects completed since October 2017, six months after the introduction of "Guidelines for GEF Agencies in Conducting Terminal Evaluation for Full-sized Projects."⁸ The work involved answering the questions in the Assessment Matrix (Annex 2), which includes all ToR questions.

9. The universe of the analysis included the guidelines, templates, and protocols produced by the GEF-IEO as well as guidance provided through workshops, courses, websites, and other capacity building activities where documentation was available. The universe also included 12 TE Validation reports and Terminal Evaluation Reports produced by the GEF and/or the Agencies, as well as other reporting produced by the GEF-IEO and the Agencies, which refer to or use quality assurance and/or TE validation information.

10. Data analysis involved the triangulation and validation of the information gathered primarily through the following methods:

⁷ Agencies' assessments of monitoring and evaluation systems are beyond the scope of this Review and are likely to be covered by the Evaluation of the Agency Self-Assessment Systems, which will also review other GEF-IEO aspects related to policy for terminal evaluations (TEs), the harmonization of ratings, and the use of knowledge from the TEs which are all relevant to enhancing the quality of the evaluations produced.

⁸ This is consistent with the cohort of the evaluations considered in the 2020 APR.

Table 1: Sample TE Validation Reports reviewed per Agency and Region

Agency	AFR	Asia	ECA	Global	LAC	Grand Total
ADB		1				1
Conservation International				1		1
EBRD			1			1
EBRD/UNIDO			1			1
FAO	2					2
					2	2
IFAD	1					1
UNIDO		1				1
World Bank				1		1
WWF-US					1	1
Grand Total	3	2	2	2	3	12

Note: ADB = Asian Development Bank; AFR = Africa; ECA = Europe and Central Asia; EBRD = European Bank for Reconstruction and Development; FAO = Food and Agriculture Organization; IDB = Inter-American Development Bank; IFAD = International Fund for Agricultural Development; LAC = Latin America and the Caribbean. UNIDO = United Nations Industrial Development Organization; WWF-US = World Wildlife Fund (United States).

11. **Desk Review:** A list of documents reviewed is included in Annex 5. In addition to GEF-IEO documentation, protocols, Guidelines, TE validation data, and the Agencies’ websites, GEF Agencies’ evaluation reports were also reviewed. A sample of validated TE reports was selected from a total of 68 reports validated by the GEF-IEO for projects completed between 2017 and 2020.⁹ The list of TE Validations reviewed is presented in Annex 2 and includes 12 reports submitted by all Agencies (one report per Agency) representing all regions (two or three per region).¹⁰ Because the large majority of the 68 reports (from the 2017–20 review period) in the portfolio had their overall quality rated Moderately Satisfactory (MS) (40 percent) and Satisfactory (S) (51 percent), most of the reports selected in the sample were rated MS (five reports) and five reports were rated S. In addition, two reports rated Unsatisfactory (U) were added to the sample.

12. **Key Informant Interviews:** A list of people interviewed is included in Annex 4. Twenty-five people were selected for interviews which were a key line of evidence for this Review. These stakeholders were considered to be representative of all key people involved in the GEF-IEO validation process and those GEF-Agencies’ representatives who were external to the

⁹ This included one Joint Evaluation (ERBD/UNIDO). Not all Agencies had completed projects (and submitted reports) during this period. Out of the 18 Agencies, only 9 were represented in the 2017–20 period of the dataset available in the GEF-IEO website.

¹⁰ This sample included reports not typically validated by the GEF-IEO, such as reports from the World Bank. No UNDP or UNEP reports had been validated in the review period.

process but who could provide insights and information about their views of the validation processes as well as other processes used in other international Agencies for the purposes of comparison.

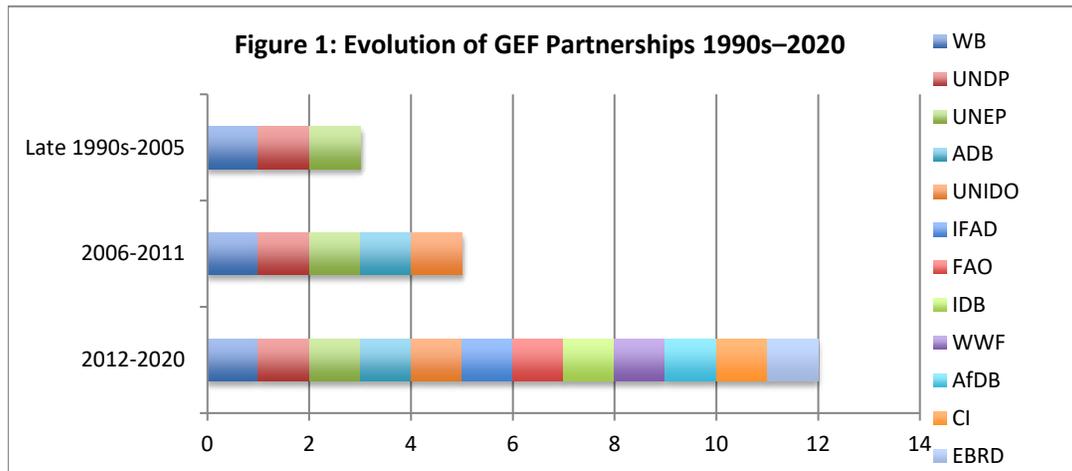
IV. THE GEF-IEO TERMINAL EVALUATION VALIDATION PROCESS

13. The GEF-IEO TE Review and Validation Process is a well-established central element of the GEF's APRs, a key mechanism of GEF accountability to the GEF Council members. Presented in 2004 as the "first step in the Global Environment Facility (GEF) to an independent assessment of the performance of the GEF's portfolio," the APRs have since been providing updates on the performance of the GEF's portfolio of completed projects. The 2020 APR presents an assessment of the cumulative results (related to outcomes, sustainability, quality of implementation, and M&E) of 1,706 completed GEF projects. The assessment is primarily based on statistical analysis of the ratings generated by the validated TEs; these populate a dataset which has expanded over the years as more GEF projects have been completed and TEs prepared and validated. The reviews and validations done by the GEF-IEO ensure that the project performance ratings across all GEF Agencies are comparable, consistent, and evidence based, for the purposes of APR reporting.

14. The work for the TE Review and Validation Process begins with the tracking of TEs to be prepared and submitted by the Agencies, gathering documentation needed for the validation of the ratings provided by the TEs, hiring consultants to undertake the reviews, and establishing and maintaining the databases to store all of this information, including the datasets.

15. GEF Agencies which have been participating in the GEF-IEO TE Validations and Reviews, as can be seen in figure 1, have been evolving and growing in number over the past few years. The initial pool of 3 key partner Agencies (UNDP, the UNEP, and the World Bank,) has evolved to about 5 Agencies with the addition of the ADB and the UNIDO in the 2006–11 period, and more recently to about 12 Agencies (with the addition of 7 more Agencies) which now have projects completed and TEs validated.

Figure 1: Evolution of GEF Partnerships 1990s–2020



Source: APR 2020 Dataset.

Note: ADB = Asian Development Bank; AfDB = African Development Bank; CI = Conservation International; EBRD = European Bank for Reconstruction and Development; FAO = Food and Agriculture Organization; IDB = Inter-American Development Bank; IFAD = International Fund for Agricultural Development; UNDP = United Nations Development Programme; UNEP = United Nations Environmental Programme; UNIDO = United Nations Industrial Development Organization; WB = World Bank; WWF = World Wildlife Fund.

16. Currently 18 institutions are accredited as GEF Agencies; for the purposes of this Review, they can be categorized in three different groups in relation to the TE-IEO validation process:

- (a) Agencies who have not yet submitted a TE and therefore have not yet participated in *the validation process*. This group includes six Agencies: Foreign Economic Cooperation Office, Ministry of Environmental Protection of China, the International Union for Conservation of Nature, as well as Development Bank of Latin America (CAF), Development Bank of Southern Africa, Brazilian Biodiversity Fund, West African Development Bank have not had evaluation reports validated by the IEO. These Agencies' projects are newer and as such, have not prepared or submitted TEs to the GEF-IEO for review and validation.¹¹
- (b) *Agencies who do not currently undergo the GEF-IEO validation process for the purposes of the preparation of the APR*. This group includes three Agencies—UNDP, UNEP, and the World Bank—which were the key GEF partners between the late 1990s and 2010. In the past, they underwent GEF-IEO TE validations and currently only have a sample of reports validated by the GEF-IEO for quality assurance purposes. The GEF-IEO uses their ratings for the APR purposes. The IEO

¹¹ The International Union for Conservation of Nature, Development Bank of Latin America (CAF), Development Bank of South Africa, Brazilian Biodiversity Fund, and the West African Development Bank were not included in the list of stakeholders provided to this consultant.

has deemed that these Agencies' review processes meet GEF information needs and accepts the reviews and the ratings verified by the independent evaluation unit of the respective Agency.¹² The GEF-IEO also validates medium-size projects (up to \$2.0 million in GEF funding) because the World Bank Independent Evaluation Group does not review the implementation completion memorandums for projects of this size as they do not meet the Banks' internal funding threshold requirement.¹³

- (c) *Agencies who had TE reports validated by the GEF-IEO in the 2017–20 period of analysis.* This group comprises nine GEF Agencies—African Development Bank (AfDB), FAO, IFAD, IDB, ADB, the World Wildlife Fund, Conservation International, European Bank for Reconstruction and Development (EBRD), and UNIDO—whose partnerships with the GEF are more recent and who had projects completed and TE reports validated starting in 2006. In the 2017–20 period, the GEF-IEO validated 68 TE reports produced by eight out of these nine Agencies, which were 24 percent of all terminal evaluations of GEF programs and projects included in the APR during this period.¹⁴

V. KEY FINDINGS

17. The key findings are based on the analysis and triangulation of the data gathered through the desk review and the key informant interviews. The Matrix in Annex 2 summarizes the key questions to be answered by this Review, the lines of evidence, and the sources of information used. The analysis focuses on TE validation reports prepared for TEs submitted in 2017–20 by 10 Agencies; these were used as the source of the APR 2020 Reported ratings (table 1). Most comments related to the GEF Agencies (unless clearly specified) refer to these Agencies who currently have TEs validated. To avoid repetition, good practices that could be considered by the GEF-IEO are highlighted throughout this report

Finding 1: The establishment of minimum evaluation requirements and work undertaken in the early years contributed to improvements in evaluation quality standards and to the good results in the quality of the evaluations validated by the GEF-IEO in the review period.

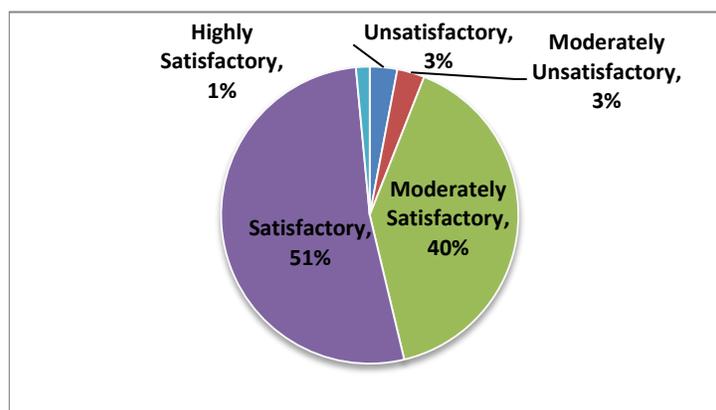
¹² In such cases, the GEF Evaluation Office will periodically assess the extent to which the independent review process continues to meet GEF concerns.

¹³ The other situation in which the GEF-IEO validates ratings of World Bank projects is when the World Bank/IEG has not reviewed and posted the Implementation Completion and Results Report validation and its ratings within two years of project completion.

¹⁴ The African Development Bank project was completed in 2016 and was not included in the current universe of analysis. 2020 APR dataset extracted mid-July 2020.

18. According to KIIs, the GEF-IEO TE guidelines and the parameters initially established to validate the TE ratings made important contributions enhancing the quality of evaluations undertaken by the GEF Agencies participating at that time. According to the 2020 APR, the overall quality of TE reports improved substantially from 1997 to 2019¹⁵ on various quality dimensions, especially reporting on project financing and M&E. Moreover, the 2020

Figure 2: Results of TE Report Validation by the IEO (2017–20)



Source: 2020 APR.

APR highlights—and interviews confirmed—for example, that TEs prepared by UNEP show improvements in quality: the proportion of TEs rated in the satisfactory range for quality increased from 81 percent for those prepared through 2009, to 97 percent for those prepared from 2010 onwards. Overall, during the 2017–20 period (figure 2) the large majority of the evaluations (92 percent) were rated Satisfactory and/or Moderately Satisfactory by the GEF-IEO TE Review and Validation process.

19. The [Guidelines for GEF Agencies in Conducting Terminal Evaluation for Full-Sized Projects](#) (2017 Guidelines) and other guidance provided for the conduct of evaluations and the preparation of these TE reports has been useful for the Agencies, according to KIIs, who point to the need to continue to push for evaluation quality and provide guidance, especially to those new Agencies which have not yet reached the stage of TE Review and Validation.

20. This is in line with the *Third Professional Peer Review of the IE Function of the GEF* (Peer Review), which found that a majority of the GEF Agencies responding to the Peer Review e-survey agreed that the 2017 Guidelines were useful and clear.¹⁶ For the present Review, KIIs confirmed that the 2017 Guidelines have initially contributed to improving the quality of TEs and to developing some related evaluation capacity among the units responsible for commissioning or conducting the TEs. Also, according to the KIIs, the 2017 Guidelines helped to establish the benchmark against which evaluations were assessed, contributing to improved quality.

21. Existing channels such as the GEF Extended Constituency Workshops, GEF-CSO Network meetings, and the GEF orientation sessions for new Agencies have been useful for communicating the evaluation requirements, providing guidance, and ensuring alignment of expectations.

¹⁵ Based on the analysis of data from 1,677 terminal evaluations for assessment of overall quality and other analysis of various quality dimensions.

¹⁶ EN_GEF.E_C58_inf_04_Third_Professional_Peer_Review_of_the_IE_Function_of_the_GEF.pdf

Finding 2: Well-established processes exist to ensure that the TE reports are submitted by the GEF Agencies and the validated ratings can be used for the APR statistical analysis and for informing other GEF-IEO evaluation work.

22. The work for the TE Review and Validation Process begins with the tracking of TEs to be submitted by the Agencies, which is conducted within a six-month period of project completion. To ensure that there are no gaps in the submissions, the GEF-IEO maintains ongoing communications with the Agencies on the status of these submissions.¹⁷ The GEF-IEO also receives Implementation Completion and Results Reports (ICRs) and retrieves the ICR Reviews available on the World Bank IEG website. The GEF-IEO also maintains Terminal Evaluation Review Forms (to be used for the undertaking of the TE Validation).

23. Established processes also exist for the completion of the TE reviews and validation of ratings, which are undertaken by GEF-IEO staff and/or independent reviewers hired by the GEF-IEO. These consultants assess the ratings provided in the TEs and complete a Terminal Evaluation Review form with validated ratings, following the GEF-IEO Terminal Evaluation Review Guidelines. The draft reviews are reviewed by a peer reviewer (GEF-IEO staff or consultant) and comments and feedback are addressed prior to finalizing the TE Review. Peer review is considered an important part of the validation process. At the end of the Review and Validation process, the GEF-IEO consultants prepare TE Review and Validation reports, and the ratings are used in the APR datasets. The TE Review and Validation reports are Microsoft Word documents stored in two repositories: an IEO internal shared drive (the “M” drive) and the GEF’s project management database (previously known as the Project Management Information System [PMIS]), now “the Portal.” The “M” drive is accessible to internal staff. The Portal is accessible to the GEF Agencies and contains a specific section in which the validation report documents are uploaded.

24. The APR datasets are updated annually adding the ratings and project information from the specific cohort year to the existing APR database. Some project information (Agency ID, country, list of participating countries, etc.) is extracted from the GEF Portal, other data and ratings come from Terminal Evaluation Review Forms. Some indicators are taken from other sources.¹⁸

25. The APR dataset enables the APR analysis of (1) Project outcomes and sustainability; (2) factors affecting attainment of project results; (3) M&E; and (4) Quality of Terminal Evaluation Report, among others. The data set is available on the GEF-IEO website and enables the identification and performance of all projects implemented in a certain country by focal area, as percentage of projects with outcomes or sustainability rated moderately satisfactory or higher, by GEF phase. It is one of the few sources of performance data that has been consistently

¹⁷ Terminal Evaluations not reviewed or covered in the Annual Performance Report include: (1) GEF activities under the Small Grants Programme (SGP); b) Enabling activities with funding below \$0.5 million and; c) Projects cancelled with less than \$0.5 million disbursed.

¹⁸ For example, the World Bank/UN definition of least developed countries (LDCs) is used to classify LDCs, and the World Bank list of fragile states is used to classify fragile states.

gathered since 1997; newer data retrieved through the review and validation process has been added annually since 2004.

26. The 2020 cohort includes TE reports of 134 completed projects for which TEs were submitted during the 2019 calendar year and reported for the first time in the 2020 APR. Because the volume of added data from additional cohorts is relatively small, its impact on the overall dataset shows limited deviations from previous years. The APR has been also incorporating more qualitative analysis in recent years.¹⁹

27. Finding 3: The APR dataset is important to ensure availability and comparability of statistical data on GEF performance.

28. The APRs are considered to be the tool under which the evidence from all the TEs is brought together and made available to the entire Partnership.²⁰ The APR dataset is also considered key for the APRs' statistical analysis of the performance of the completed GEF projects and is currently the only source of comparable data available. The current weight and contribution of the GEF-IEO TE Review and Validation process to the update, however, is relatively small.

29. In the initial APRs, the GEF-IEO review and validation of TEs was responsible for ensuring consistency across all GEF Agencies and contributed to reducing the data gaps by collecting and incorporating data when the implementing Agencies had not included it in their reports. Over the years, according to KIIs, the TE data provided by the three partner agencies UNDP, UNEP, and the World Bank improved, the gaps in information related to the GEF evaluation criteria reduced, and these Agencies' data were considered adequate to meet GEF information needs. The GEF-IEO review and validation process still plays a role in quality assurance of these TEs by assessing a sample of their TE reports but uses their ratings as submitted, for APR.²¹ As can be seen in table 2, in the period 2017–20, most of the data for the 2020 APR comes from about 76 percent of the evaluations produced and validated by Independent Evaluation Offices of UNDP (49 percent), the World Bank (33 percent), and UNEP (14 percent).²² 25 percent of the validated data for 2020 is based on the IEO TER validation process.

¹⁹ Annual Performance Reviews (APRs) focus on a specific theme in any given year. The theme is addressed in detail. The themes covered in the past three APRs include: Sustainability (APR 2017); Transportation (APR 2019); Quality of terminal evaluation reports (APR 2020). By focusing on one theme the GEF-IEO avoids repetitiveness in the APRs.

²⁰ GEF-IEO. APR 2020.

²¹ To address the possible inconsistencies in ratings across Agencies, the GEF-IEO conducts a terminal evaluation (TE) review for a random sample of the terminal evaluations prepared by the World Bank, UNDP, and UNEP. Usually, 10 percent to 15 percent of the TEs prepared by these three Agencies are selected. According to the APR, so far, 196 projects have been rated by the GEF IEO for quality assurance purposes. On a six-point scale, the average difference in the outcome rating from the two sources is 0.05, which is not statistically significant. Binary-scale on the net ratings match exactly for 94 percent of the projects (184 projects). Thus, there is no substantial difference in the validation of ratings by the IEO or by the Agency.

²² According to the GEF Guidelines for Agencies Conducting Terminal Evaluations, where "GEF Agencies have an established process in place whereby an independent evaluation group reviews and validates terminal evaluations and assesses the quality of terminal evaluation reports, the GEF-IEO will assess the extent to which the independent review process meets GEF concerns and provides the required information for GEF oversight and reporting."

Finding 4: The large majority of the TE ratings validated by the GEF-IEO are for projects implemented by three GEF Agencies with established by Evaluation units or from GEF Agencies which may be well positioned to “graduate” soon.

30. IEO validated 68 TEs over the 2017-2020 period. Most GEF-IEO validated reports (72 percent of the total) in the period analysed were submitted by three Agencies: UNIDO, FAO, and IFAD (with more than 10 reports from each agency validated). These represent only about 17 percent of the total TEs (288) validated in the 2017–20 period.

31. Figure 3 shows that the quality of the 49 TEs produced by these Agencies in the period analyzed is high. Only 4 percent of them were rated by the GEF-IEO as Moderately Unsatisfactory, and the remaining 96 percent were rated Satisfactory, Highly Satisfactory, or Moderately Satisfactory. The KIIs confirmed that these evaluations are quality assured by evaluation offices of these GEF Agencies, which were engaged in the TE preparation processes and final reports.²³

Table 2: Source of APR ratings (2017–20)

Agency	#	% of total
GEF-IEO rated	68	24%
UNIDO	21	7%
FAO	17	6%
IFAD	11	4%
IDB	6	2%
ADB	3	1%
JOINT	3	1%
WB	3	1%
WWF	2	1%
CI	1	0%
EBRD	1	0%
Independent Evaluation units	220	76%
WB	31	11%
UNEP	40	14%
UNDP	143	50%
JOINT	6	2%
Grand Total	288	100%

Source: APR 2020.

Note: ADB = Asian Development Bank; CI = Conservation International; EBRD = European Bank for Reconstruction and Development; FAO = Food and Agriculture Organization; IFAD = International Fund for Agricultural Development; UNDP = UN Development Programme; UNEP = UN Environmental Programme; UNIDO = UN Industrial Development Organization; WB = World Bank; WWF = World Wildlife Fund.

²³ The UN Industrial Development Organization (UNIDO) Independent Evaluation Division is established in the Office of Evaluation and Internal Oversight (ODG/EIO) which, according to the website, “is responsible for providing the Director General, and management at large, with independent, objective assurance, advice and evaluation designed to add value to and improve the efficiency and effectiveness of UNIDO’s operations, internal control framework, risk management, RBM and governance processes.” The Charter of the OEIO states that the work of the EIO “shall remain free from undue influence of interference by management.” See paragraph 12 under “Independence and Objectivity” at https://www.unido.org/sites/default/files/files/2019-04/DGB-2019-07_DG_Bulletin_Charter_of_the_Evaluation_and_Internal_Oversight.pdf

Table 3: GEF-IEO TE Reports validated per year

GEF Agency	Year of Terminal Evaluation Completion					
	2017	2018	2019	2020	Grand Total	%
UNIDO	4	12	5		21	31%
FAO	2	8	6	1	17	25%
IFAD	6	3	1	1	11	16%
IDB	1	5			6	9%
WB	2		1		3	4%
JOINT	1	2			3	4%
ADB			2	1	3	4%
WWF	1		1		2	3%
CI		1			1	1%
EBRD		1			1	1%
Grand Total	17	32	16	2	68	100%

Source: APR 2020.

Note: UNIDO = UN Industrial Development Organization; FAO = Food and Agriculture Organization; IFAD = International Fund for Agricultural Development; IDB = Inter-American Development Bank; WB = World Bank; ADB = Asian Development Bank; WWF = World Wildlife Fund; CI = Conservation International; EBRD = European Bank for Reconstruction and Development.

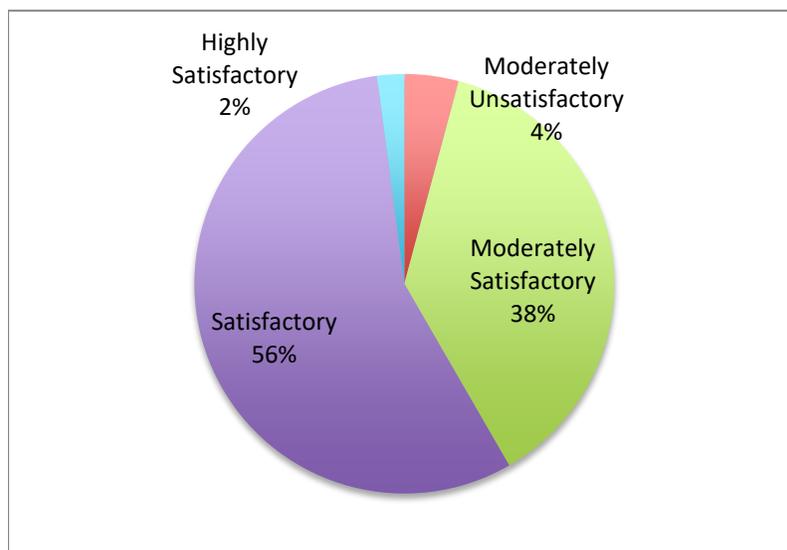
32. The evaluation reports from these 3 agencies are quality assured by the Agencies' evaluation units. For example, the UNIDO evaluation reviewed explains, "This evaluation was managed by the responsible UNIDO Project Manager with quality assurance by the Independent Evaluation Division."²⁴ According to the ToR of the TE report, "all UNIDO evaluations are subject to quality assessments by the UNIDO Office for Independent Evaluation. Quality assurance and control is exercised in different ways throughout the evaluation process (briefing of consultants on methodology and process of UNIDO's Office for Independent Evaluation, providing inputs regarding findings, lessons learned and recommendations from other UNIDO evaluations, review of inception report and evaluation report by the Office for Independent Evaluation)." The Office of Evaluation and Internal Oversight includes the three oversight functions: evaluation, investigation and audit) is organizationally located as an independent office under the Office of the Director General of UNIDO.²⁵

²⁴ According to the UNIDO Evaluation Manual, for independent project or program evaluations the Independent Evaluation Division plays an advisory, clearance, and Quality Assurance (QA) function. Currently, project evaluation administration is delegated to Project Managers. The Independent Evaluation Division is responsible for identifying and selecting evaluation consultants and clearing the evaluation Terms of Reference (ToR), inception report, draft, and final report. The evaluation requirements of specific donors (such as GEF) are considered as needed.

²⁵ Regarding the independence of the Office from Management, Decision 3(e) was adopted at the 44th session of the UNIDO Industrial Development Board and relates to the actual establishment of "...an effective and functionally and operationally independent Office of Internal Oversight" that requested the Director General "to ensure independent, timely and effective oversight (evaluation, investigation and audit) in line with standards and norms of the United Nations and reflecting the implementation of respective General Assembly resolutions." This decision is the basis for the preparation of the Charter of the Office of Evaluation and Internal Oversight (ref. DGB/2019/07, 26 March 2019). The independence of the evaluation function is

33. The TEs submitted by FAO are managed by FAO’s Office of Evaluation and produced by independent consultants hired externally.²⁶ The Office of Evaluation is independent from Management, reporting to the Director General and to the Council through the Programme Committee.²⁷ The Office of Evaluation reviews the reports and provides comments. The TE reports also undergo a clearance process, and both reports and management responses are public documents, available on the FAO website. As at other Agencies who have larger GEF portfolios, a FAO-GEF Coordination Unit is in place, in addition to Project Management units, and is tasked with ensuring that the Agency meets its fiduciary obligations, among which is the requirement to undertake TEs at the end of the project cycle. FAO manages a considerable portfolio of 191 projects (\$864 million), most of which are likely to be undergoing a TE in the next few years.²⁸

Figure 3: Quality of UNIDO, FAO, and IFAD Evaluations as rated by the GEF-IEO in the 2017–20 period



Source: APR 2020

34. At IFAD, the situation is slightly different regarding the involvement of its Independent Evaluation Office. As in the case of other international finance institutions, GEF projects are components of much larger IFAD projects. In line with the GEF-IFAD agreements, the GEF components are evaluated by independent consultants hired by the Environment, Climate,

also reflected in the UNIDO Evaluation policy (ref. DGB/2018/08, 01 June 2018) and in the Charter of the Office of Evaluation and Internal Oversight, under item IV—Independence and Objectivity. It should be noted that the Independent Audit Advisory Committee recommended further changes to the Charter to increase the independence of the Office and approval the Charter by the Industrial Development Board. In line with this recommendation, a revision of the Charter was prepared in consultation with the Independent Audit Advisory Committee and will be submitted to the next meeting (in November 2020) of the Industrial Development Board for its consideration and approval. Once approved, the revised Charter will be published on the UNIDO website.

²⁶ According to the FAO website, the following measures are in place to ensure that the FAO evaluation function conforms to UN Evaluation Group norms and standards: (1) peer review of major evaluation reports; (2) biennial review by a small group of independent peers for conformity of evaluation work to best practice and standards; (3) independent evaluation of the evaluation function every six years.

²⁷ According to the Charter of the Office of Evaluation, “Independence should be protected throughout the evaluation process: policy, institutional framework, management of the evaluation function, conduct of evaluations and follow-up. The evaluation function must be located in the Organization outside the line management that it is mandated to evaluate and must have a direct line of reporting to the governing bodies and the Director General.”

²⁸ FAO is among the top 4 of 18 GEF agencies, according to its website, which describes GEF as “an increasingly important partner for FAO not only because of the financing available, but because FAO and GEF share key priorities such as biodiversity mainstreaming, climate change adaptation, land degradation neutrality, and transboundary water and fisheries management..”

Gender and Social Inclusion Division Program Unit, which collaborates with the Independent Evaluation Office on the evaluation reviews. The IFAD Independent Evaluation Office undertakes evaluations of the IFAD projects, which may or not have specific GEF components; it is not formally engaged in the conduct or management of the evaluations of the GEF projects but could play a more formal role in the production and/or quality assurance of TEs.

35. The IEO review process is undertaken to verify ratings and assess the evaluation quality “where a GEF Agency does not have an independent evaluation unit or lacks an independent review process.” . Thus, several GEF Agencies are interested in being considered for graduation, and like UNDP, UNEP, and the World Bank, have the GEF-IEO consider their TEs suitable and of good quality to meet GEF reporting purposes.

Finding 5: A relatively small portion of the TEs validated in the review period are for projects implemented by the multilateral development banks (MDBs) and NGOs; capacity development and guidance from GEF-IEO could help reduce existing data gaps and improve TE quality.

Ten percent (19) of the total TEs in the 3 year period analyzed²⁹ were produced by five Agencies (IDB, World Bank,³⁰ ADB, the World Wildlife Fund, and Conservation International, and also includes evaluations of projects conducted jointly by more than one Agency.³¹ They represent about 28 percent of TEs validated by the GEF-IEO with the following breakdown: IDB with about 9 percent of reports (six TEs validated), the World Bank (three TEs), ADB (three TEs) and the World Wildlife Fund (two TEs), with about 4 percent of reports each, and the EBRD and Conservation International with about 1 percent each (one TE each).

²⁹ These represent 28 percent of the total 68 reviews undertaken by the GEF-IEO.

³⁰ The three World Bank projects included in the dataset were medium-size projects (up to \$2.0 million in GEF funding). The World Bank Independent Evaluation Group does not review the Implementation Completion Memorandums for projects of this size as they do not meet the Banks’ internal funding threshold requirement. The other situation in which the GEF-IEO validates ratings of World Bank projects is when the World Bank/IEG has not reviewed and posted the ICR validation and its ratings within two years of project completion.

³¹ During the timeframe covered by this Review, a total of three evaluations of projects implemented jointly by UNDP/IDB, EBRD/UNIDO and World Bank/IDB were validated by the GEF-IEO.

36. As shown in figure 4, the quality of these 19 TEs produced in the period analyzed is good. About 47 percent of them were rated Moderately Satisfactory by the GEF-IEO, and the remaining 42 percent were rated Satisfactory. Eleven percent of the TEs were rated Unsatisfactory. Many of these evaluations were produced or quality assured by independent evaluation offices of these GEF Agencies and/or managed by the Agencies' GEF Coordinating Unit.³²

Figure 4: Quality of remaining TEs as rated by the GEF-IEO in the 2017–20 review period



Source: APR 2020.

37. This review's analysis of the 12 TE validated reports compared the TE ratings and the IEO Review ratings on the key GEF evaluation criteria, identifying gaps in the TE ratings. The results are presented in table 4, which shows that most TEs that are missing ratings are in documents from the MDBs (ADB, IDB, World Bank, and EBRD to a lesser extent), which together have a small percentage of all TEs validated by the GEF-IEO in the 2017–20 period (table 2). This may be partially because the MDBs have their own internal validation and reporting processes which are different than those of the GEF. According to KIIs, these internal processes are more difficult to adjust, especially considering the small size of the GEF portfolio in comparison with the multilateral development banks' overall portfolios. Several of the data gaps were associated with criteria such as "M&E Design," "M&E Implementation," "Implementation Quality," and "Execution Quality."

38. Overall, when data gaps associated with the MDBs and other TEs are combined, out of the total 72 rubrics related to the six key criteria, 20 rubrics (27 percent) were not rated (and/or unable to rate) and were completed by the validation reviewer. The validation process was therefore instrumental to filling this gap and providing the ratings in these categories for those evaluations missing this information. The analysis is aligned with the APR findings that the "implementation" rubric is often not rated by the TEs.

³² Agencies with a GEF Coordination Unit: FAO; IFAD, IDB, UNIDO, UNDP; Agencies with Evaluation Unit: FAO, IFAD, UNIDO, UNDP, World Bank, ADB, EBRD.

Table 4: Comparison between TE and IEO Review Ratings (selected TEs)

Ratings obtained by Terminal Evaluations and Review and Validation per Evaluation Criteria													
Criteria	Outcomes		Sustainability		M&E Design		M&E Implementation		Implementation Quality		Execution Quality		Overall TE Quality
Agency	TE	IEO Review	TE	IEO Review	TE	IEO Review	TE	IEO Review	TE	IEO Review	TE	IEO Review	IEO Review
IFAD (3362)	S	MS	MU	MU	MU	MS	MU	MU	S	S	NR	MS	S
EBRD (3535)	NR	S	S	ML	NR	U	S	MU	HS	MS	S	MS	MS
ADB (3589)	NR	S	L	L	S	S	S	S	S	S	S	MS	S
EBRD/UNIDO (3593)	S	S	L	ML	HS	MS	S	MS	S	MS	S	S	MS
FAO (3777)	MU	U	ML	ML	MU	MS	U	U	MU	MU	MU	U	MS
UNIDO (3908)	S	S	ML	ML	S	S	S	S	S	S	MS	MS	S
FAO (3984)	MS	MS	ML	ML	MS	MS	MS	MS	S	MS	MS	S	MU
IADB (4113)	MS	MS	ML	ML	NR	S	NR	MS	NR	UA	NR	MS	MS
World Bank (5439)	S	S	NR	UA	NR	MS	NR	UA	NR	S	NR	S	MS
WWF (5771)	S- HS	S	ML	ML	UA	S	S	S	S	S	S	S	S
IADB (5838)	MS	MS	ML	UA	NR	S	NR	MS	NR	UA	NR	MU	U
CI (9163)	HS	S	ML	ML	HS	S	HS	S	HS	S	HS	S	S

Ratings	# of instances
Increased	3
Decreased	18
Same	31
Not rated by TE	20
Total	72

Source: Completed Summary of Project Ratings Section in The Terminal Evaluation Review form, GEF Independent Evaluation Office, APR 2020 (See Annex 9)

39. In addition to the gaps in M&E, Implementation, and Execution Quality) there were also gaps associated the Sustainability criteria. For example, one evaluation only contained assessments related to overall sustainability and did not contain assessments related to financial, socioeconomic, institutional, and environmental sustainability even when the ToR

contained references to the GEF evaluation policy and required that the evaluation apply the GEF standards and minimum requirements.

40. The use of different rating scales was also another issue, present in at least 2 out of the 12 evaluations. As described in the APR, there are gaps in compliance with TE guidelines across all agencies. These illustrate the need to ensure the Agencies submit TEs that are complete with all required GEF information, to avoid gaps for the purposes of the APRs. In many cases, to complete missing data in particular, the IEO reviewers relied on information provided in the narrative TE, showing that in certain situations the TEs contained the information required to complete this rubric but did not specifically assess or assign a rating. This suggests that more targeted guidance and better forms or ToRs provided to consultants prior to the conduct of the evaluations could help to reduce data gaps.

41. Even in cases (e.g., EBRD, IDB 4113, the World Bank) where evaluations were missing most ratings, the evaluations were still considered to be Moderately Satisfactory in terms of their quality by the GEF-IEO TE reviewer, a function of the specific criteria used for evaluation quality ratings (see Finding 8). This is an important element, considering that some judgments are made based on the TE narrative. Only one of evaluation (IDB 5838) was rated Unsatisfactory by the reviewer.

42. In addition to completing missing ratings, the GEF IEO reviewers confirmed, lowered or raised the ratings provided by the TEs. In the sample analyzed, out of the 52 rubrics provided, most of the ratings (43 percent) were maintained or confirmed by the reviewer. About 25 percent of the ratings were lowered (many of them from Highly Satisfactory [HS] to Satisfactory [S]), and only 4 percent of the total rubrics were raised. In most cases, adjustments lowered of the ratings by only one point. Adjustments in the ratings were well substantiated and properly justified, often based on the TE narrative because of the limited data from other sources, as is explained in more detail in paragraph 97.

43. The GEF-IEO reviewers reviewed the TEs and the related project PIRs, midterm reviews, and project documents and rated the performance of the completed projects based on evidence available in these documents. A TE review and validation report was then prepared in each case, with justifications for the ratings adjustments.

To what extent is the terminal evaluation validation process transparent? ^[L]_[SEP]

Transparency. An essential feature at all stages of the evaluation process, transparency involves clear communication concerning decisions for the program of work and areas for evaluation, the purpose of the evaluation, the criteria applied, the evaluation approach and methods, and the intended use of the findings. Documentation related to evaluations must be freely available, easily accessible, and readable for transparency and legitimacy. (GEF-IEO Evaluation Policy 2019)

Finding 6: Certain aspects of the GEF-IEO TE submissions and related guidance are clear, transparency related to the GEF-IEO TE Reviews can be enhanced by ensuring that information about the TE Review and Validation process is more visible, better disseminating this information and the results of the process.

44. The KIIs revealed that the Agencies consider certain aspects of the TE Validation Process transparent, such as the submission process and the criteria and guidance provided for the completion of the TEs. Good communication exists between the GEF-IEO and most of the Agencies regarding tracking evaluations in the pipeline and ensuring submissions are up to date. The terminal evaluation guidelines are clear regarding the submission processes, and they are easily retrievable in the GEF-IEO website.

45. Information about the GEF-IEO Terminal Evaluation Reviews and Validations exists but is not easily accessible in the GEF-IEO website. The 2008 Guidelines for GEF Agencies in Conducting Terminal Evaluations (the 2008 Guidelines) available on the website are perhaps the best description of the process: one chapter (about 2 ½ pages), “Independent Review and Validation,” - presents the criteria and ratings as well as the processes used by the GEF-IEO to validate the evaluation reports.³³

46. The Guidelines have been updated since 2009, but information about the TE Review and Validation is not clearly outlined in more recent versions.³⁴ The 2019 GEF Evaluation Policy explains that “the IEO validates project terminal evaluations for those Agencies in which the evaluation function is not fully independent” and that “It does so according specific evaluation guidelines” referencing the Guidelines for GEF Agencies in Conducting Terminal Evaluation for Full-sized Projects (2017 Guidelines).³⁵ The recent Guidelines used by the GEF Agencies contain only a brief explanation, in its last section, about IEO’s use of the TE ratings and Review and Validation reports (prepared by the Agencies) but contains no information about the GEF-IEO Review and Validation process per se.³⁶ According to the KIIs, this lack of communication reduces the transparency of the validation process.

47. A few Agencies are aware of the existence of the TE validations, primarily through the GEF-IEO website and exposure to the APR reports, either via the web or participation in the GEF Council. They are aware that the ratings resulting from the TE validation processes are widely available in the APR dataset posted in the GEF-IEO public website. They use the APRs as a way

³³ In the Guideline, the GEF encourages the GEF Agencies to undertake the independent review of the findings, ratings, and quality of terminal evaluations.

³⁴ According to the GEF-IEO information about the process was removed from the 20017 Guideline to address a request by the Agencies that the Guideline focus specifically on the processes related to the conduct of the GEF evaluations. The TE guideline is was developed following extensive consultations.

³⁵ IEO, “Guidelines for GEF Agencies in Conducting Terminal Evaluation for Full-sized Projects,” (<http://www.gefio.org/sites/default/files/ieo/evaluations/files/gef-guidelines-te-fsp-2017.pdf>)

³⁶The 2017 Guideline explains that the information provided by the terminal evaluations and the independent assessments of these evaluations are used by the GEF-IEO to report annually to the Council on portfolio performance and trends. The 2017 Guidelines also mention that the evidence presented in the terminal evaluations and the independent reviews may also be an input to other evaluations that the GEF-IEO undertakes. The Guidelines also note that the GEF IEO shares the terminal evaluation reports and terminal evaluation reviews publicly through the GEF project database webpage.

to benchmark their evaluations and evaluation process against those of other Agencies, identifying areas of evaluation work where improvement is required within their agencies.³⁷

48. The large majority of Agencies who have had their TE reports validated in the period analyzed are not aware of the existence of the TE dataset on the GEF-IEO website and do not use this data. With a few exceptions, the Agencies (including those few ones who were aware of the TE reviews and validations) were not aware that the data related to their TEs had been validated and that the publicly released datasets contained the GEF-IEO validated data. According to the KIIs, most Agencies were also not aware of the existence of TE review and validation reports. Only one of the Agencies interviewed had seen a report with the IEO's full assessments and justification for ratings. The Agency considered the report valuable feedback, useful to detect strengths and weaknesses in the TE and pointing to areas for potential improvements.

49. Currently, these individual TE Validation reports (which have migrated from the PMIS to the new GEF Portal) is not accessible to the Agencies, according to KIIs. Perhaps this explains why the GEF Agencies interviewed (and those participating in the Peer Review Survey) commented on their limited awareness of the GEF-IEO TE Review and Validation process.

50. The 2008 GEF-IEO [Guidelines for GEF Agencies in Conducting Terminal Evaluations](#)³⁸ establish that upon completion of the TE review and validation, the GEF-IEO will send it to the appropriate GEF Agency, which will then have two weeks to provide comments, after which the GEF-IEO will finalize the TE Review, taking any comments into consideration. However, currently this practice does not seem to be consistently implemented; according to GEF-IEO staff, Agencies seldom provided comments, and the limited time available for validation reduced the time available for such exchanges.

51. The KIIs revealed that clear communication about the criteria and process used by the GEF-IEO to undertake the TE reviews and validations is desirable, as is information explaining the process by which the Agencies can “graduate” and have their TE ratings accepted by IEO (like those of UNDP, UNEP, and the World Bank).

To what extent does the GEF-IEO TE validation process enhance the credibility of ratings based on the evidence provided in the terminal evaluations? 

Credibility. Evaluations must be credible and based on reliable data and observations. Evaluation reports should reflect consistency and dependability in data, findings, judgments, and lessons learned, with reference to the quality of the instruments, procedures, and analysis used to collect and interpret information. (GEF-IEO Evaluation Policy 2019)

³⁷ This Peer Review information was validated by the KIIs.

³⁸ P. 35

Finding 7: The TERs assessments are well substantiated and aligned with criteria, but the limited evidence available pose challenges to proper assessments of project performance and in ensuring that evaluative judgments are consistent across reviewers.

52. The GEF-IEO desk reviews include an analysis of the TE reports, the Project Implementation Reviews (PIRs), Mid-Term Reviews, and the Project Approval Document, as a basis for the validation assessments. In the 12 TE validations analyzed, the types, numbers, quality, and depth of these TE documents provided to the reviewers were quite varied and not necessarily consistent across all TEs. For example, in addition to the TEs, about four to five Project Implementation Reviews were provided on average to guide the GEF-IEO reviewers, usually one PIR for each year of project implementation. But since the PIRs were prepared by different Agencies, their content varied widely in formats, length, and depth of information and this is likely to continue as Agencies have established their own internal processes (as described in paragraph 73). Overall, some PIRs were comprehensive in-depth documents (40 pages) and others were high level, short, and very concise (5 pages).³⁹ In one extreme case, there were data gaps in the TE related to missing ratings for five out of six criteria, and only one PIR was provided.⁴⁰ The PIRs are descriptive and could lend themselves to positive bias because they are prepared by project managers (team leaders) and may contain limited qualifying information. The Mid-Term Reviews are usually produced two to three years prior to the TE and do not necessarily contain up-to-date information.

53. The Project Approval Documents were the most consistent of those provided for each TE, but they contain limited information to guide assessment of results. Only 5 out of the 12 evaluations contained documentation related to a Mid-Term Evaluation/Reviews.⁴¹ Seven evaluations had other types of documentation, but these also contain limited information upon which to base assessments (e.g., commitment letters, amendment letters, etc.).⁴²

54. The TE reports provided by most Agencies—except the World Bank’s Implementation Completion Memorandum—were somewhat similar in terms of depth and type of information provided, but there were still variations across the assessments provided in each specified GEF evaluation criterion.⁴³

³⁹ The length is provided here to illustrate the variability across the Project Implementation Reviews. No linkage is implied here between the number of pages and the quality of information provided.

⁴⁰ Project Implementation Reviews are linked with duration of projects. The duration of GEF projects generally ranges from 2 to 10 years. A project having only one Project Implementation Review is not a problem in itself, but then it was the only one available for the TE validation .

⁴¹ Medium-size projects and those with a short duration generally do not have a Mid-Term Review.

⁴² According to GEF-IEO staff, these documents are reviewed not as primary information sources but to check for inconsistencies between the information provided in a TE and corresponding Project Implementation Reviews and the Mid-Term Review. In some instances, these may be useful to help fill information gaps and provide additional context for observations made in the TE.

⁴³ The Implementation Completion Memorandum (ICM) and the Implementation Completion and Results Report (ICR) are different documents. The ICRs are prepared for full-scale projects. ICMs are prepared for medium-size projects where the investment is low. ICRs do not vary much in quality; nor do ICMs— but ICRs are much superior in quality to ICMs (ICMs are probably the worst TEs that GEF IEO receives). The World Bank currently takes up very few, if any, medium-size GEF-projects, according to the GEF-IEO staff.

55. When comparisons are made across the ratings available in all TEs, PIRs, and the validated GEF-IEO ratings, the 2020 APR offers interesting insights. On Outcome ratings, the 2020 APR states, the “GEF IEO validations of outcome ratings show a disconnect between project implementation report ratings, terminal evaluation ratings, and IEO ratings” based on analysis of 596 completed projects for which ratings from PIRs, TEs, and validated TEs were available. A higher percentage of projects are rated in the satisfactory range for outcome by PIRs (96 percent) and TEs (92 percent) compared with TE validations conducted by the GEF-IEO (84 percent). The disconnect between the GEF-IEO validations and PIRs is 10 percent, and between GEF-IEO validations and TEs is 6 percent.⁴⁴

56. In any case, in lieu of other evidence, the GEF-IEO reviewers document the limitations of the data for drawing conclusions on certain criteria or, in certain cases (as mentioned earlier), a reviewer may be able to rely on information provided in the narrative TE for their judgments.⁴⁵ The adjustments in the ratings in the 12 TE reviews assessed were well substantiated and properly justified, often based on the TE narrative because data from other sources were limited. As such, as in any desk-based review, the reviewers have a limited basis for properly assessing project performance.⁴⁶

57. Moreover, when the evidence is inadequate, given the limited criteria used in IEO assessments raises the possibility that two or more reviewers (who conduct assessments based on subjective judgments) could potentially assess and revise ratings differently,⁴⁷. This could impact the credibility of the TE validation. As such, while these are good quality individual TE reviews, there are limitations in properly assessing project performance and in ensuring evaluative judgments are consistent and reproducible by other reviewers.

58. The quality of the TE in this context is key to the credibility of the process. Currently the quality of the large majority of the TEs validated by the GEF-IEO is assessed as Moderately Satisfactory or higher, pointing to credible evidence.

59. Most Agencies interviewed had not seen the content of the GEF-IEO validation reports (despite attempting to retrieve them directly from the Portal); therefore most KIIs could not comment on the extent to which the validation reports actually contributed to enhancing credibility and quality of future TEs. In addition, the large majority of KIIs were not aware of the desk review process used by the GEF-IEO to review the TEs and validate their ratings. This lack of communication discourages confidence across the Agencies that the process is clear, credible, and based on reliable data and observations.

⁴⁴ APR 2020.

⁴⁵ The GEF-IEO TE Validation Reports explain the adjustments made and rationale for the ratings provided where gaps existed; however, it is important to ensure that explanations are not only provided for the new rating adjusted by the consultant but also for the non-acceptance of the TE rating. Language is also an issue: some TE reports are in Spanish and French. In one case, the correct scale used by the TE was not properly translated leading the validation reviewer to rate it UA (unable to assess), though the proper rating had been provided.⁴⁵

⁴⁶ This is in line with the findings in the Peer Review report (paragraph 77) “no matter how accurate, the desk-based review of any project final report or evaluation has limited scope for properly assessing how a project actually performed”.

⁴⁷ Similar issue exists in other agencies - see UNEP example in paragraph 117.

Finding 8: The criteria and ratings scales used by the GEF-IEO are aligned with common good practice.

60. The GEF criteria and rating scales used in the TE validation are aligned with common practices: Most International Agencies use similar criteria (e.g., Outcomes, Relevance, Effectiveness, Sustainability) and the ratings criteria are also similar. However, the GEF criteria related to M&E Design, M&E Implementation, Quality of Implementation, and Quality of Execution are less common among UN Agencies and are specific to the GEF.

61. Separate documents exist to detail the GEF criteria and rating scales to be used by both the evaluators ([Guidelines for GEF Agencies in Conducting Terminal Evaluations for Full-Sized Projects](#)) when preparing the TEs and the validation reviewers (Terminal Evaluation Report Review Guidelines; see Annex 6) when preparing the reviews and validations. The documents are consistent for the most part and similar in their content, although they are not fully aligned to each other, possibly because of recent edits. The review and validation Guidelines document, for example, does not cover the set of criteria considered quite important in the TE Evaluation Guidelines related to Quality of Implementation and Quality of Execution. The GEF Evaluation Guidelines document stresses the importance of rating these criteria and provides detailed explanations.⁴⁸ The TE Review Form used by the Reviewers, is complete and contains all the criteria.

62. For the rating scales, both guidelines (TE and review) establish six-point or four-point scales to be used for the assessment of the criteria, which reflect variations.⁴⁹ For example, “Highly Satisfactory” to “Highly Unsatisfactory” ratings reflect gradual variations on a scale (e.g. from “no shortcomings” to “severe shortcomings”). Specific information about each criterion and its expected TE content is provided to guide the validation and the evaluation report preparation; however, details are not provided as to what the meaning of the scales represent (e.g. what is considered “a shortcoming” or “a severe shortcoming”).

63. Differences also exist regarding the GEF-IEO assessment of the quality of the TEs. The GEF-IEO reviewers assess the TE quality through six open-ended questions included in the Terminal Evaluation Review Form.⁵⁰ The reviewers provide ratings and short comments to: (1) extent of report coverage of assessments of relevant outcomes and impacts; (2) extent of internal consistency, completeness of evidence and substantiation of ratings; (3) proper assessment of sustainability; (4) comprehensiveness of lessons learned supported by the evidence; (5) existence of actual project cost and co-financing information; and (6) extent to which the evaluation assesses the project’s M&E systems.⁵¹ Very limited qualitative data are

⁴⁸ Quality of implementation pertains to the role and responsibilities discharged by the GEF Agencies that have direct access to GEF resources. Quality of Execution pertains to the roles and responsibilities discharged by the country or regional counterparts that received GEF funds from the GEF Agencies and executed the funded activities on ground. ⁴⁸

⁴⁹ Most criteria (Outcomes, M&E Design, M&E Implementation, etc) use a six-point scale. The Sustainability ratings use a four-point scale ranging from “Likely” to “Unlikely.”

⁵⁰ These questions are not included in the TE Report Review Guidelines.

⁵¹ Reviewers use a six-point rating scale to rate each sub-criterion and an overall rating of the TE report (Highly Satisfactory to Highly Unsatisfactory): 1. Quality of reporting on outcomes; 2. Consistency of reporting, completeness of evidence, and ratings

gathered through the current process. Additional data were collected by the GEF-IEO on the quality of evaluation reports for the purposes of the 2020 APR (Special Thematic Focus: Quality of Reporting.)⁵²

64. By contrast, the UN organizations use a comprehensive checklist by the United Nations Evaluation Group (UNEG) [Quality Checklist for Evaluation Reports](#) containing about 45 questions covering eight evaluation areas (ranging from Evaluation Report Structure, Evaluation Object, Evaluation Scope, Methodology, Findings, Conclusions, Recommendations, Gender). Their evaluation reports are assessed in terms of the quality of evaluation's design and methodological approach, the quality of its findings and evaluative evidence, and the robustness of its conclusions and recommendations.

65. While this Review also points to the existence of good quality evaluations in the pool of 12 TEs analyzed, it is important to highlight that the ratings validated by the GEF-IEO reviewers were against GEF-IEO parameters which, as noted above, are different, more high level, and less robust than the parameters established by UN Agencies (e.g. UNEG), for example. It is possible that the overall ratings related to evaluation quality could be lower as a result of comprehensive assessments against more complete checklists (such as the UNEG checklist).

How do the GEF-IEO TE reviews and validations compare with TE validations undertaken by other international development agencies? ^{ISIP}

Finding 9: Validation work similar to the GEF-IEO validations is undertaken by other international agencies, and practices used by the World Bank Independent Evaluation Group (IEG) Reviews of the Implementation Completion and Results Report (ICR) and the UNDP-IEO GEF-TE Validations could be considered for potential application in the GEF-IEO validations, particularly for the agencies that do not have internal evaluation quality assurance processes..

66. A major difference between validations by these other organizations and the GEF-IEO validations is that in both cases, the independent evaluation units are validating and reviewing TEs of their own projects, while the GEF-IEO validates TEs produced by partner agencies, and relies on the partner Agencies internal evaluation quality assurance processes. The Agencies complete TEs to comply with fiduciary agreements, and the GEF's leverage to enforce compliance and request additional requirements depends on the size of the GEF portfolios and the extent of changes needed at the institutional level in certain Agencies. However, in cases where GEF Agencies do not have internal evaluation quality assurance mechanisms, the GEF

are well substantiated; 3. Quality of reporting on sustainability; 4. Quality of lessons and recommendations; 5. Reporting on actual project costs and cofinancing; and 6. Quality of reporting on project monitoring and evaluation.

⁵² The exercise relied on protocols specially designed to guide the review of the quality of Agency reporting following three years of the implementation of the 2017 Guidelines.

IEO could consider some of the good practices on validation implemented by the more established GEF Agencies.

67. There are many different approaches and practices used by other Agencies that could be applicable to the GEF, and some of these are presented in this Report for consideration. However, a more in-depth analysis of their applicability to the specific GEF situation should be undertaken in each case. For example, desk reviews are also used by other Agencies, but mechanisms exist to enhance their credibility, such as broadening the lines of evidence used, tightening the quality assurance process, and enhancing consistency of judgments through more in-depth guidance on assessment criteria. The World Bank/IEG ICR Reviews—its main instrument for self-evaluation of full-size projects—serves as an independent validation of the results reported⁵³ and, like the GEF-IEO validation, is a desk-based, critical review of the evidence, results, and ratings of the ICR in relation to the project’s design documents.⁵⁴ However, the desk review is complemented by an additional line of evidence. The ICR Review relies on an interview with the last task team leader upon which to base its own ratings for the project. The ICR Reports are produced by the IEG in only one week and drafts are sent to the task team leaders, who have 10 days to accept the validation results or challenge them.

68. The World Bank ICR Review is intended to critically assess the evidence provided in the ICR and from the last task team leader, its quality, and the attribution of results to the activities or actions supported by the project under review. It is not simply a summary of what is in the ICR. IEG is not privy to evidence that was not included in the ICR. Since 2006, whenever insufficient evidence is provided for IEG to arrive at a clear rating, IEG downgrades the relevant ratings. According to KIIs, by penalizing limited information, the system seems to be encouraging more complete reports and the showing of the evidence (some projects are good but have no data).

69. UNDP undertakes separate validation of the results of GEF TEs using an approach very similar to that used by the GEF-IEO TE Validations. The UNDP assessments of GEF TEs are also based on desk reviews and the same types of documentation used in the GEF-IEO TE validation exercise (PIRs, Project Approval Documents, Mid-Term Reviews). Reviewers also adjust ratings and fill gaps where ratings are missing. The main difference between the process in the two Agencies is that, in addition to the TE validation, all UNDP GEF TE reviews, like all other UNDP Evaluations,⁵⁵ undergo a more complete and more robust quality assessment process (figure 1), which, as mentioned earlier, is an important element to demonstrate the quality of the TEs and add credibility. Using a set of UNEG parameters, a rating system, and weightings, in this quality assessment process, the evaluation report is assessed regarding the evaluation’s design and methodological approach, the quality of its findings and evaluative evidence, and the robustness of its conclusions and recommendations.

⁵³ IEG also reviews and validates 20 percent of their Project Performance Assessment Reports (PPARs).

⁵⁴ IEG reviews each ICR that is submitted to IEG (as opposed to only a subset of them, as the GEF-IEO).

⁵⁵ Quality assessments are carried out for all decentralized evaluations conducted by UNDP, as well as the United Nations Capital Development Fund and United Nations Volunteers program, including evaluations of United Nations Development Assistance Frameworks (UNDAFs), outcome, project, and program evaluations and thematic evaluations.

70. While the GEF-IEO TE Review Form contains only 6 open-ended questions to guide the assessments, as mentioned, the UNDP quality assessment forms are more robust and contain, for example, 16 questions for one single criterion, “Evaluation structure, methodology and data sources”—to guide the reviewers undertaking the review. This criterion is weighted 30 percent of the total, showing its importance to the overall quality of the evaluation. This additional guidance for reviewers also helps more junior reviewers to think through the criteria at more granular levels, enhancing the robustness of the assessments.

71. A platform called the QA Dashboard is used by the UNDP consultants to complete the validation; they provide responses to 40 questions (and five sections for quality assessment) and an additional section used for the validation of GEF TE ratings. By the time consultants validate the GEF ratings, the methodological approach of the evaluation has been assessed, and the bases on which the consultants justify adjustments, for example, are better documented.

72. Other Agencies provide structured feedback on the results of their validation or quality assurance processes with the goal to enhance learning and promote better evaluation work. The UNDP quality assessment of an evaluation report aims at (1) supporting bureau oversight functions by providing concurrent feedback through detailed analysis of the quality of the evaluation reports with recommendations for their improvement; and (2) contributing to corporate lessons learned by drawing from good evaluations in the annual report on evaluation.⁵⁶ The UNDP Independent Evaluation manages the quality assessment process and provides feedback to the program units and country offices, which then can introduce adjustments and strengthen areas of the evaluative evidence and the evaluation report, as well as adjust the management and implementation of evaluations to ensure usable findings and recommendations and the overall utility of decentralized evaluation reports.

73. Also, other Agencies’ guidelines for quality assessment are more detailed and visible to the public and their targeted audience, and some include information related to the GEF-IEO TE validation process performed by the UNDP-IEO. The UNDP-IEO Evaluation [Quality Assessment Guidelines](#) are posted in the UNDP-IEO website as a stand-alone document and include a series of detailed questions which will be used for the assessment of quality of the reports, as well as explanations about the GEF-IEO TE Validation process it undertakes, only and specifically for evaluations of UNDP-GEF projects and programs. Among other information about the process used (which is similar to that of the GEF-IEO), the UNDP-IEO Guidelines contain more detailed explanations and a grid which is used by UNDP to validate all GEF-related evaluation reports. While both the GEF Guidelines and the UNDP Guidelines cover the same topics, the UNDP Guidelines contains more detailed information (see Section 6.10.5). By attaching these guidelines and criteria to the ToR of every GEF TE, UNDP disseminates this information further and enhances the transparency of the process.

⁵⁶ The other purposes are (1) Improving the quality of evaluative evidence to better manage contributions to development results; (2) Supporting accountability by providing an independent assessment of the quality of decentralized evaluation reports to the UNDP Executive Board and management; and (3) Strengthening consistency in evaluation reporting and quality across projects.

74. Other agencies which do not necessarily undertake validations similar to the GEF-IEO have developed tools and practices for their own quality assurance, peer reviews of evaluation quality or evaluation assessments, and these can also be considered by the GEF-IEO for improvements, whether fully, partially, or even simply conceptually.

75. UNEP, for example, has made improvements in the consistency of evaluative judgments made across desk reviews of TEs and among evaluation reviewers by adopting better guidance and more granular definitions of the ratings used for the assessments in each criteria (see example in Annex 10). UNEP uses a matrix of evaluation criteria and performance ratings defining the conditions and supporting evidence required to assess each individual criterion.⁵⁷ This matrix enables the UNEP Evaluation Office to strengthen the consistency in the assessments and assignment of ratings across evaluations among evaluation consultants and within its own evaluation management team.⁵⁸

76. The matrix contains definitions of the meaning of each score (HU, U, MU, MS, S, HS) with more granular qualities to guide the assessments. So, for example, under Relevance, the criterion “alignment with UNEP Strategy and Strategic priorities” should be rated Highly Unsatisfactory if the “project implementation strategies and delivered contributions show lack of alignment with any parts of UNEP’s mandate” or Unsatisfactory if these “implementation strategies and contributions show weak alignment with a few secondary or peripheral parts of UNEP’s mandate” (see a 1-page example out of the 25-page matrix in the Annex 10).⁵⁹

Finding 10: International Agencies are enhancing systems to gather more qualitative data on their evaluations, investing in searchable databases that can quickly identify and retrieve data on evaluation findings, recommendations and lessons for learning purposes. They are also making this information widely available.

77. The World Bank/IEG ICR Reviews aim at contributing to both learning and accountability. Like the GEF-IEO’s work, these reviews contribute to databases for aggregation and analysis, and report on the effectiveness of projects⁶⁰ The data gathering also provides a

⁵⁷ The two main purposes of this matrix are (1) to make the foundation of UNEP performance ratings transparent and open for constructive discussion, and (2) to reduce the risk of rating the same performance features repeatedly under several criteria.

⁵⁸ In many instances projects will demonstrate features that appear within the descriptions of *different ratings levels under the same evaluation criterion*, and the decision on the final rating will be made based on (1) where the majority of the features appear, and (2) the nature and strength of the effect that an aspect of performance is found to have had on achieving the agreed results. Any negative, unintended project effects will reduce the rating against the criterion most closely associated with the affected result.

⁵⁹ Consideration could be given to developing a similar matrix to guide GEF-IEO reviewers with more granular definitions of the performance ratings, specifically designed or adapted to fit with GEF-IEO priorities.

⁶⁰ According to the Guidelines, the ICR is intended to (1) Provide a complete and systematic account of the performance and results of each project;(2) Capture and disseminate experience from the design and implementation of a project in order to: (i) improve the selection of interventions to achieve the goals of the Country Partnership Framework (or, previously, the Country Assistance Strategy); (ii) improve the design and implementation of interventions through lessons learned; and (iii) help ensure greater development impact and sustainability of projects; (3) Provide accountability and transparency at the level of individual projects with respect to the activities of the World Bank, the borrower, and involved stakeholders; (4) Provide a vehicle for realistic self-evaluation of performance by the World Bank and borrowers; and (5) Contribute to databases for aggregation, analysis, and reporting, especially by the Independent Evaluation Group (IEG), on the effectiveness of lending projects in contributing to development strategies at the sector, country, and global levels.

systematic way for IEG to critically review the evolving portfolio as projects close and to summarize the projects' objectives and key results, in addition to the ratings. However, the World Bank and other international agencies are also moving toward qualitative data collection. In addition to quantitative data, the write-ups (qualitative data) of the World Bank ICR Reviews are stored in a searchable database within IEG and, for all projects that closed since FY2011, are posted on IEG's external website. ICR Reviews more than five years old are declassified and disclosed on a quarterly basis. The data are often useful as a starting point for IEG's ICR reviewers as a quick way to identify projects of different types, or with specific objectives or activities, in preparing to undertake larger country, sector, or thematic evaluations.

78. UNDP is also mining the qualitative information in UNDP-IEO's thematic, sector, global, and regional evaluations, drawing on a large number of evaluations which have undergone quality assessments, including about 1,325 evaluations of GEF programs posted currently in the UNDP-IEO [Evaluation Resource Center](#) website. UNDP is currently preparing a series of lessons learned based on evaluative evidence, called [Reflections](#). These are a series of knowledge products (brief papers) offering lessons from past evaluations of UNDP work in crisis settings, intended to satisfy the broader request from UNDP managers for help identifying what works and what doesn't, and in what contexts. The initiative supports UNDP decision makers, especially at the country level, in their COVID-19 crisis response efforts.

79. All UNDP evaluations (including GEF-TEs) are posted in the Evaluation Resource Center, which has search capabilities not only by evaluation year, type, country, thematic area of intervention, but also by lessons learned. The quality assurance reviewer identifies Lessons Learned in the evaluation and uses a tag system for their classification (e.g., UNDP Management; Food Security Sustainability; Natural Resource Management), which enables keyword searches.⁶¹ The ratings results of UNDP's quality assurance reviews and associated reports are only available internally, and staff can filter results by evaluation quality, for example. These are useful tools for mining data in evaluation report.

Finding 11: The GEF Agencies have developed several good practices that could be shared and disseminated across all GEF Agencies.

80. This review has also identified a few good practices that some GEF Agencies are undertaking which could be useful for all Agencies, especially the new GEF Agencies which have not yet submitted TEs.

81. UNEP, UNIDO, and the World Wildlife Fund also undertake assessments of the quality of their TE reports at the draft and final stages and include the results of the final QA review as an annex in the evaluation report. At UNEP, a quality assessment checklist is used by the UNEP evaluation staff (the evaluation manager and another staff member in the office serving as a peer reviewer) to provide structured feedback to evaluation consultants, especially at the draft

⁶¹ The key steps of the UNDP Quality Assessment process include: (1) posting evaluations to the Evaluation Resource Center, (2) verification, (2) quality assessment, and (4) feedback. The Evaluation Resource Center is a public website and therefore all documents go through a final high-quality verification process.

stage of the report.⁶² The first version of the report received by the UNEP Evaluation Office from the external evaluation consultant is assessed with this tool. When the evaluation report is finalized, the report is once again assessed for quality. Any difference in quality scores between initial submission and final report is a measure of improvements resulting from the evaluation quality assurance process. UNEP only publicly discloses the final version of the evaluation report quality assessment and considers the quality assessment of the first submitted draft as an internal document. At the end of the evaluation, the UNEP Evaluation Office also assesses the compliance of the evaluation process with agreed standard procedures related to Independence, Financial Management, Timeliness, Project's Engagement and Support, Quality Assurance, and Transparency. This quality assessment is published by UNEP as an Annex of the evaluation.⁶³ Similarly, UNIDO's Office for Independent Evaluation reviews each GEF-TE draft and final evaluation report before submitting it to the GEF Evaluation Office and circulates it within UNIDO together with a management response sheet. The World Wildlife Fund evaluations are also the object of quality assurance (possibly by its own Evaluation team) and contain a three-page report with the assessment of the quality, key lessons, and recommendations.

82. Over the years, according to KIIs, the World Bank made improvements to the data collection and reporting used in GEF projects, so that staff are clear on what to collect and are able to capture of the specific results and lessons related to the capacity building and catalytic elements of the GEF projects. This experience may be useful to other GEF Agencies (ADB, AFDB, EBRD, IDB, IFAD), because their GEF projects are also smaller components of larger loan projects of the World Bank. Reporting on these component projects is challenging,⁶⁴ and some agencies have opted to undertake TEs of the GEF component only (to address fiduciary needs) which in turn may not be able to capture the other elements of the intervention, further limiting the identification of effects in the larger project. Considering that GEF work is often linked to testing new approaches and undertaking more risks, it is very important to extract lessons from the overall intervention, using TEs as a knowledge management tool and incorporating the feedback into new programming and future project design.

83. UNEP, UNDP, and UNIDO evaluations each use their own specific template designed for GEF projects which is included in the evaluation ToR, with good results regarding completeness of GEF-required information. A large number of the UNDP-GEF related evaluations posted on the Evaluation Resource Center website follow the [Guidance for Conducting Terminal](#)

⁶² The UNEP Evaluation Office is the only unit within UNEP that has the mandate to conduct "evaluations"; other UNEP assessment exercises are called "reviews." All GEF projects are subject to evaluation by the UNEP Evaluation Office, which hires external evaluators, manages the evaluation process, and assures the quality from beginning to end. All UNEP evaluations, including GEF TEs, are subject to a quality assessment by the UNDP Evaluation Office.

⁶³ The checklist contains criteria such as Quality of the Executive Summary, Evaluation Methods, Project Description, Theory of Change, Key Findings (Strategic Relevance, Quality of Project Design, Nature of the External Context, Effectiveness of Outputs and Outcomes, Likelihood of Impact, Financial Management, Efficiency, Monitoring and Reporting, Sustainability, Factors Affecting Performance), Conclusions and Recommendations, Report Structure and Presentation Quality, and Overall Report Quality Rating.

⁶⁴ There are also issues related to harmonization of development and global environmental objectives. Good practice: Assessing Global Environmental Objectives. Projects wholly or partly financed by the Global Environment Facility will likely include Global Environmental Objectives in the project appraisal document, in addition to Project Development Objectives.

Evaluations of UNDP-supported, GEF-Financed Projects and use the specific template designed for GEF projects which are included in the evaluation ToR (Annex 2). The template consists of tables and matrices to be populated by the evaluators with data. Perhaps as a result of these measures, the UNDP-GEF evaluations are consistently complete, with information on evaluation criteria and requirements, most of them are complete in coverage of GEF-required criteria, and the scales they use (also included in the ToR) are usually consistent with GEF ratings.

Finding 12: The independent evaluation quality assurance in which UN Agencies are formally engaged can be a good source of documentation for GEF-IEO to monitor the performance of some of the GEF Agencies' Evaluation Units.

84. The UN-Secretariat Inspection and Evaluation Division of the Office of Internal Oversight Services undertakes an external assessment of UN entities and presents, on a biennial basis, the UN Evaluation Dashboard. The assessment covers the quality of the agencies' evaluation reports (as part of a broader assessment of the function) and could be a good source of documentation for the GEF-IEO assessments of the quality of the Evaluation Units of several GEF partners.

85. Because Evaluation is one of the indicators against which the UN agencies report against the United Nations System-Wide Action Plan,⁶⁵ several UN Agencies undertake independent assessments of the quality of their evaluations.⁶⁶ In addition to UNEP and UNDP, other UN Agencies (UNIDO, FAO, IFAD) also report on this indicator.

VI. FINAL REMARKS

86. Overall, this Review considers that the GEF-IEO TE validations and reviews are currently at a turning point. Improvements have been made over the years, with the building of an important dataset in which to report on the GEF performance results. Over the past few years, the statistical analysis presented in each APR has been adding little interesting insight due the newly added cohort is small in relation to the overall portfolio.⁶⁷

87. More recently, following the graduation of the three larger Agencies (UNEP, UNDP, and the World Bank), improvements have also been made to TEs by UNIDO, FAO, and IFAD, which may now be well positioned to graduate. In addition to having established independent evaluation functions, these UN agencies are also undertaking more robust independent quality assurance for the purposes of UN System-Wide Action Plan reporting. For IFAD, perhaps some discussions could take place to ensure that the GEF TEs undergo formal reviews by the IFAD Independent Evaluation Office. Should the GEF-IEO consider this approach feasible, the

⁶⁵ For more details see <https://www.unwomen.org/en/how-we-work/un-system-coordination/promoting-un-accountability>

⁶⁶ The first phase of UN System-Wide Action Plan (UN-SWAP) implementation focused on gender mainstreaming and planning, the second phase is designed to focus on results, and includes monitoring activities and outcomes for gender-related Sustainable Development Goal results.

⁶⁷ As identified in the Peer Review Report, the limitations identified in the IEO validation process and in the comparability of TE ratings diminish the utility of APRs.

validation efforts could focus on a sample of their Agency portfolio for quality assurance purposes, similar to the arrangement with the three larger Agencies. This would free up resources at the GEF-IEO with no impact on data quality for the purposes of the APR.

88. The analysis in this Review points to about 3 percent of missing data in reports from the remaining 10 percent of the portfolio (approximately), which mostly includes the MDBs and the NGOs⁶⁸: if the missing data could be simply be noted as “blank” or “missing,” the overall impact on the APR cohort analysis would probably be small, especially considering that new data in each APR (relative to the overall dataset) deviate little from previous years. Validation efforts could then also perhaps only target a sample of the portfolio for quality assurance purposes, freeing resources even further.

89. Considering that gaps in data are relatively small and the evaluation quality is good, there are perhaps opportunities to address data gaps through more targeted guidance to this group of Agencies. Opportunities may also exist for targeted capacity building to the newer Agencies (which have not yet submitted TEs).⁶⁹ It is important to ensure that guidance is provided for the development of evaluation ToRs and the GEF-IEO expectations regarding quality of the TEs, which will likely result in fewer gaps in the data and better quality of the evaluation products.⁷⁰

90. In addition to sharing the datasets with the Agencies via the website, the validation reports could be more consistently shared and discussed with the Agencies submitting TEs as a way to provide feedback for future TE improvements. The 2008 Guidelines recommended the practice of sharing the review and validation reports with the GEF Agencies, could be more consistently and widely applied through the introduction of automated messaging systems in which the GEF Agencies receive a notification through the Portal that their TE review and validation process is initiated and finalized and that results are available.⁷¹

91. The GEF-IEO could also support evaluation functions in the Agencies through the building of communities of practice or groups for targeted capacity building where issues exist. Though the Agencies are very different (and a “one-size-fits-all” approach would likely not work), the KIIs consider it important to have mechanisms for sharing experiences across agencies, especially at the level of the evaluation units. Some more well-established evaluation functions could share experiences to guide the work of those newer Agencies which are not yet advanced. Some earlier practices such as the 2014 Annual Meeting of APR (which brought together face-to-face UNEP, UNDP, and the World Bank) were identified as good practices which the GEF-IEO could champion again. It could also be appropriate to set up specific groups

⁶⁸ This is a rough estimate, based on the fact that 4 out of 12 TEs analyzed in this Review (30 percent) had a large portion of ratings missing. Considering that these evaluations were conducted by the Asian Development Bank, Inter-American Development Bank, and World Bank, they make up about 4 percent of the total TEs rated in the 2017–20 period.

⁶⁹ The size of their portfolio has not been confirmed by this Review.

⁷⁰ From 2016–18, the GEF-IEO provided a training module on TE preparation to the GEF Agencies, including the new Agencies.

⁷¹ *Upon completion of the review, the Office will send it to the appropriate GEF Agency, which will then have two weeks to provide its comments, after which the GEF Evaluation Office will finalize the review, taking these into consideration.* (GEF-IEO 2008 [Guidelines for GEF Agencies in Conducting Terminal Evaluations](#) P. 35)

of Agencies to share information, such as for example, the MDBs, which could benefit from the World Bank experience in the capture of results of GEF catalytic interventions as components of larger loan projects.

92. The introduction of more quality checks is likely to enhance the desk reviews and ultimately the validations (should these be continued or streamlined), by providing complementary and additional documentation and evidence on the quality or lack of information on evaluations methodology, data collection, and analysis, on the basis of which evaluation assessments are made. For example, the desk reviews could include interviews with the evaluation managers, or the GEF-IEO TE validation team could participate in a select number of terminal evaluation missions as a means of assessing the evaluation practices of the GEF Agencies; or, for a select group of projects, the team may be able carry out a technical assessment of project impacts after project completion.

93. The existence of more qualitative data related to findings, recommendations, and lessons learned that are included in these evaluations could further GEF-IEO's efforts in mining qualitative data, especially if these were also identified through the TE validations and reviews and their quality assured through additional checks.

94. The recent focus in the APRs to include qualitative findings, in addition to the quantitative ratings of the TEs, has been an important improvement, in line with the trends (also identified in other Agencies) of mining evaluation data. As the validation and quality assurance processes evolve and data is used for GEF-IEO Evaluations (e.g. country evaluations or thematic or strategic country clusters), opportunities may exist for GEF-IEO TE validation teams to participate more in these other evaluation missions, cross-checking TE and validation information as well as getting feedback from country missions on the validated TEs and enhancing synergies across the work of various GEF-IEO teams.

95. A few suggestions for future GEF-IEO interventions are presented for consideration in Annex 1. These require more in-depth analysis in light of the specific GEF-IEO situation and the resources for validation in IEO; but also, in light the future strategy of the GEF-IEO as a whole. There may be more opportunities to streamline resources or develop joint initiatives, or share practices more widely to facilitate innovation and cross-fertilization of ideas.

ANNEX 1: SUGGESTED AREAS FOR GEF-IEO FUTURE INTERVENTIONS

Issue identified by the GEF-IEO TER Review	Suggested intervention	Related to Finding #
1. Good guidance provided for GEF TEs/established processes for TE submission	<ul style="list-style-type: none"> - Continue to maintain/update GEF TE Evaluation Guidelines and maintain the APR datasets; 	Finding # 1; Finding #2;
2. GEF-IEO validations currently providing about 24% of data for APR purposes (2020)	<ul style="list-style-type: none"> - Consider discontinuing TE validation in its present form; or alternatively, invest more resources to boost lines of evidence (e.g. trips for reviewers; more time for reviewers to undertake consultations with project/evaluation staff personnel) as is done by other agencies. - Keep data gaps, noting them as “missing” for the purposes of APR and other internal data use. - Maintain some level of QA reviews targeting newer agencies and TE samples from 6 Agencies). 	Finding #3; Finding #5
3. Data gaps coming mostly from MDBs, NGOs, currently representing only a very small percentage of rubrics		Finding # 5
4. The TEs are the main line of evidence used in desk reviews for TERS assessments to fill in gaps		Finding #7; Finding #9 (para. 108–111)
5. Desk reviews based on limited evidence and potential inconsistencies in validation assessments (credibility of validation process)	<ul style="list-style-type: none"> -Cross-check TE information and use opportunities of other GEF-IEO evaluations for additional line of evidence (country consultations) to validate TEs in countries where evaluations are taking place (e.g. country missions) (see item #8) -Consider adopting/providing more granular guidance on the meaning of the scales to be used for each of the criteria assessments (e.g. what is considered “Satisfactory” or “Highly Satisfactory”). 	Finding #7; #9
6. TER quality assessment criteria are limited to 6 open-ended questions and high-level guidance on scales,	<ul style="list-style-type: none"> -Provide more granular guidance for TE validation assessment criteria (e.g. UNEG criteria) on performance ratings to strengthen consistency in assessments and assignment of ratings among reviewers, especially for newer Agencies. -Adopt a robust quality review process using/adapting UNEG Quality Checklist for Evaluation Reports; -Assess the quality of TEs (all those submitted by newer agencies and a sample of those submitted by UNDP, UNEP, the World Bank, UNIDO, FAO, and IFAD) in line with more robust practices to build confidence in their quality; 	Finding #9 (para 112–118)

	<ul style="list-style-type: none"> -Ensure qualitative assessments provide complementary and additional documentation and evidence on the quality/lack of quality of evaluations methodology, data collection and analysis, lessons learned, etc); 	
	<ul style="list-style-type: none"> - Consider an annual independent evaluation quality assurance (similar to those conducted by UN Agencies) to benchmark evaluation quality. 	Finding #13
<p>7. Perception of limited transparency of GEF-IEO TE validation process</p>	<ul style="list-style-type: none"> - Disseminate information better (make it more visible) about the process used by the GEF-IEO to assess quality of TEs (and/or validate ratings) ensuring awareness among all GEF Agencies; - Support evaluation functions in the Agencies by building communities of practices, for targeted capacity building and/or formalizing informal relationships and/or sharing experiences across agencies, especially at the level of the evaluation units; - Use experienced Agencies (UNDP, the World Bank, UNEP) to share good practices and exchange information; - Group agencies by type as appropriate (MDBs, NGOs, UN Agencies) - Share QA review documents with all Agencies as a capacity building means; 	Finding # 6; Finding 11 (para 124)
<p>8. Other Agencies use qualitative data in TEs to document/extract lessons for other evaluations and/or future programming</p>	<ul style="list-style-type: none"> -As other Agencies do, start building a database to store qualitative (and quantitative) data on lessons learned from evaluations for mining purposes; -Pre-establish a set of “tags” (e.g. sector areas, type of lesson, type of results, etc.) to enable searching capabilities; -Ensure the QA reports (and TEs) are uploaded in the database of qualitative experiences, where the “tags” can be searched, so as to retrieve valuable qualitative (and quantitative) data for GEF-IEO thematic, sectoral, and country-level evaluations. 	Finding #10
<p>9. Data issues (missing/Inconsistent data) mostly found in small number of TEs from MDBs, NGOs</p>	<ul style="list-style-type: none"> - Upon receiving TEs, verify that ratings are provided and communicate missing/data existing gaps (with a view of accepting or rejecting the TEs). Some evaluation units (NGOs) may be able to retrieve the missing data. - Replace TE Validations with a robust quality review process and assess the quality of all of TEs from this group to verify quality and for data mining purposes. 	Finding #5; Finding 11 (para 125)

	<ul style="list-style-type: none"> - Provide more hands-on training and ToR templates to ensure proper coverage of TEs in all GEF evaluation criteria; - Ensure GEF Agencies receive an automated notification through the Portal that their TE validation process is initiated and finalized, and that results are available; 	
<p>10. Graduate Agencies with established evaluation/QA processes</p>	<ul style="list-style-type: none"> - Determine/negotiate with Agencies to ensure GEF TEs undergo internal QA or validation process by internal evaluation units - Graduate UNIDO, FAO, and IFAD—no longer validate their ratings; - Sample a few TEs and undertake more robust QA with a view to documenting lessons and ensuring availability of “qualitative” data on TE evaluations for the database, to be used for future programming. - Share QA review documents with these Agencies as feedback; - Use UN Agencies’ Independent Quality Assurance processes to monitor performance of GEF Agencies’ Evaluation units. 	<p>Finding #4; Finding #6; Finding #12</p>

Note: APR = Annual Performance Report; FAO = Food and Agriculture Organization; IFAD = International Fund for Agriculture; MDB = multilateral development bank; NGO = nongovernmental organization; QA = quality assurance; TE = terminal evaluation; TER = terminal evaluation review; UNEG = United Nations Evaluation Group; UNDP = United Nations Development Programme; UNEP = United Nations Evaluation Programme; UNIDO = United Nations Industrial Development Organization.

ANNEX 2: ASSESSMENT MATRIX

Questions	SuB-questions	Performance Indicators /Variables to consider	Potential Sources of Data Collection/Triangulation
1. To what extent is the GEF-IEO terminal evaluation validation process transparent? ^[1] _[SEP]	<ul style="list-style-type: none"> - Is the GEF-IEO TE validation process clear and information on criteria easily accessible to the participating agencies? - Are there opportunities for the IEO and GEF Agencies to engage in discussions about the results obtained by the validation process? - What are the areas where further needs/opportunities for GEF-IEO capacity building related to the process? 	<ul style="list-style-type: none"> • Information/guidance documents exist to explain the process and are accessible to GEF Agencies and stakeholders • Feedback mechanisms exist for GEF Agencies to challenge TE validation results • Evidence of communications/meetings between GEF Agencies and IEO officials about TE validation results • Degree to which the GEF Agencies and IEO Staff perceive the TE validation process as transparent • Views and opinions of KIs on capacity building activities and future needs 	<ul style="list-style-type: none"> • Project Reports/ Document reviews • Interviews with Key Informants
2. To what extent does the GEF-IEO TE validation process enhance the credibility of ratings based on the evidence provided in the terminal evaluations? ^[1] _[SEP]	<ul style="list-style-type: none"> - Are the current GEF-IEO TE validation process (criteria, overall/individual quality ratings) appropriate to measure the quality of TE evaluations? - Do the parameters of the GEF-IEO TE validation process (e.g. criteria/ratings) enable comparability and identification of trends in the quality of GEF TE evaluations? - Are the GEF-IEO TE validation criteria applied consistently across the universe of GEF TE evaluations? - Are there adjustments needed in the process, parameters, criteria and ratings and tools to enhance compliance of a diverse community of Partner Agencies? - Does the GEF-IEO TE validation process rely on experienced professionals to conduct 	<ul style="list-style-type: none"> • Information/guidance documents containing information about criteria/rating and explanations about the process • Perceived value-added of TE validation process; by GEF Agencies and IEO staff and Stakeholders; • Trends of QA ratings before and after validation process • Perception of GEF Agencies and IEO staff that the process is adequate, and that credibility of ratings are enhanced as a result of the TE validation process • Evidence of use of TE validation results by GEF Agencies 	<ul style="list-style-type: none"> • Project Reports/ Document reviews/TE Validation Data/Reports • Interviews with Key Informants

	<p>quality assessments?</p> <ul style="list-style-type: none"> - What are the strengths of the GEF-IEO TE validation process? - What are the main weaknesses or gaps in the current GEF-IEO TE validation process? 		
<p>3. How does the GEF IEO TE validation process compare with TE validation process of other international development agencies? ^[1]_[SEP]</p>	<ul style="list-style-type: none"> - To what extent are the GEF-IEO TE validation criteria and ratings aligned with/comparable to that of other international development agencies? - How does the level of investment/resources by other Agencies in reference to their TE validation processes compare with that of the GEF IEO TE validation process? - How do the results of TE validation process compare across the GEF IEO and International Development Agencies? 	<ul style="list-style-type: none"> - Guidelines and documentation about International Agencies TE Validation Processes - Views and opinions of selected International Development Agencies on cost/benefits of their TE validation systems - Views and opinions of selected International Development Agencies on enhancements in quality and/or credibility before and after their TE validation - International Development Agencies trends of QA ratings before and after introduction of validation process and/or other analysis of their systems 	<ul style="list-style-type: none"> • Project Reports/ Document reviews • Interviews with Key Informants • International Development Agencies Website data/User analytics (if available)
<p>4. What are the good practices and lessons that may be incorporated in the GEF ^[1]_[SEP] IEO's validation process?</p>	<ul style="list-style-type: none"> - What are lessons and practices can be incorporated to the validation process to address its key weaknesses identified in this Review? - What are the key/priority areas of intervention that can be addressed with funds/ resources allocated to the GEF-IEO? - Are there technologies and/or systems that could be leveraged to improve the GEF-IEO validation process? 	<ul style="list-style-type: none"> • Degree of alignment between existing resources and GEF-IEO TE validation process expectations • Evidence of existence of GEF IEO funds/resources for future enhancements; 	<ul style="list-style-type: none"> • Project Reports/ Document reviews • Interviews with Key Informants

ANNEX 3: LIST OF TE VALIDATIONS REVIEWED

#	GEF ID	Evaluation title	Organization	Region	Overall Quality rating	GEFIEO
1	5439	Fighting Against Wildlife Poaching and Illegal Trade in Africa: the Case of African Elephants	World Bank	Global	4	Y
2	3362	SIP: Catchments and Landscape Management	IFAD	AFR	5	Y
3	3535	Creating Markets for Renewable Power in Ukraine	EBRD	ECA	4	Y
4	3589	CTI Coastal and Marine Resources Management in the Coral Triangle: Southeast Asia under Coral Triangle Initiative	ADB	Asia	5	Y
5	3593	Market Transformation Programme on Energy Efficiency in Greenhouse Gas-Intensive Industries in Russia	JOINT	ECA	4	Y
6	3777	CBSP Sustainable Management of the Wildlife and Bushmeat Sector in Central Africa	FAO	AFR	4	Y
7	3908	Industrial Energy Efficiency for Malaysian Manufacturing Sector	UNIDO	Asia	5	Y
8	4113	Mainstreaming Biodiversity in Palm Cropping in Colombia with an Ecosystem Approach	IDB	LAC	4	Y
9	5771	Improving Mangrove Conservation across the Eastern Tropical Pacific Seascape (ETPS) through Coordinated Regional and National Strategy Development and Implementation	WWF	LAC	5	Y
10	9163	Enabling the use of Global Data Sources to assess and Monitor Land Degradation at Multiple Scales	CI	Global	5	Y

ANNEX 4: LIST OF KEY INFORMANTS

#	Key Informant	Organization	Category
1	Maria Rosario Catalina Narciso mrcnarciso.consultant@adb.org ADB/GEF Portfolio Management Officer (Consultant) Sustainable Development and Climate Change Department	ADB	Agency
2	AYANLEH DAHER ADEN, A.DAHERADEN@AFDB.ORG , Senior Environment & Climate Finance Officer Coordinator, Global Environment Facility (GEF) & Adaptation Fund (AF) Unit Environment and Climate Finance Division (PECG1)	AfDB	Agency
3	Guirane NDIAYE G.NDIAYE@AFDB.ORG GEF Portfolio Supervision and Results (M&E) Specialist	AfDB	Agency
4	Orissa Samaroo osamaroo@conservation.org Senior Director, GEF Policy and Portfolio Management, Conservation International	CI	Agency
5	Alexis Franke, FrankeA@ebrd.com External Relations and Partnerships, Vice Presidency for Policy and Partnerships	EBRD	Agency
6	Oley Sybira SYBIRAO@ebrd.com	EBRD	Agency
7	Henley, Guy HenleyG@ebrd.com	EBRD	Agency
8	Genevieve Braun Genevieve.Braun@fao.org , Programme officer, FAO-GEF Coordination Unit	FAO	Agency
9	Amélie Solal Celigny, Amelie.SolalCeligny@fao.org , Evaluation officer, FAO Evaluation Office	FAO	Agency
10	Ortega Rada, Alexandra ALEXANDRAO@iadb.org IDB-GEF Technical Specialist, Climate Change and Sustainable Development Department	IDB	Agency
11	Killmer, Annette Bettina ANNETTEK@iadb.org Operations Advisor, Climate Change and Sustainable Development Sector (CSD)	IDB	Agency
12	Margarita Astralaga < m.astralaga@ifad.org > Director of the Environment, Climate, Gender and Social Inclusion Division (ECG)	IFAD	Agency
13	Fabrizio FELLONI, Interim Officer in Charge of the Independent Office of Evaluation of IFAD	IFAD	Agency
14	Margarita Arguelles margarita.arguelles@undp.org Focal Point for Vertical Funds Guidance, Bureau for Policy and Programme Support	UNDP	IDA
15	Michael Spilsbury michael.spilsbury@un.org Director, Evaluation Office, UN Environment Programme	UNEP	IDA
16	Michaela BERNDL M.BERNDL@unido.org , Senior Evaluation Assistant, UNIDO Independent Evaluation Division Team	UNIDO	Agency
17	Christopher David Nelson, cnelson2@worldbank.org ; Manager, Infrastructure and Sustainable Development Unit, Independent Evaluation Group	WB	IDA
18	Kissick, Amelia Amelia.Kissick@wwfus.org Senior Program Officer, Adaptive Management, Evaluation Office, World Wildlife Fund	WWF	Agency
19	Ms. Tullia Aiazzi (Adviser)	Consultant	Peer Review
20	Molly Watts Sohn	GEF-IEO	Staff
21	Neeraj Kumar Negi	GEF-IEO	Manager
22	Laura Nissley	GEF-IEO	Validation Reviewer
23	Ritu Kanotra	GEF-IEO	Validation Reviewer

ANNEX 5: LIST OF KEY DOCUMENTS AND WEBSITES REVIEWED

ADB Asian Development Bank Independent Evaluation

<https://www.adb.org/site/evaluation/main>

EBRD Independent Evaluation Department <https://www.ebrd.com/what-we-do/evaluation-overview.html>

FAO – Office of Evaluation <http://www.fao.org/about/who-we-are/departments/office-of-evaluation>

GEF-IEO.(No date). Terminal Evaluation Report Review Guidelines

GEF-IEO. 2017. The GEF in the Changing Environmental Finance Landscape: FINAL REPORT OF OPS6 https://www.gefio.org/sites/default/files/ieo/evaluations/files/ops6-report-eng_1.pdf

GEF-IEO. 2020. *Third Professional Peer Review of the IE Function of the GEF (Peer Review)*, Peer review: https://www.thegef.org/sites/default/files/council-meeting-documents/EN_GEF.E_C58_inf_04_Third_Professional_Peer_Review_of_the_IE_Function_of_the_GEF.pdf

GEF-IEO. APR 2016: <https://www.gefio.org/sites/default/files/ieo/evaluations/files/project-performance-2017.pdf>

GEF-IEO. APR 2018 <https://www.gefio.org/council-documents/annual-performance-report-2019>

GEF-IEO. APR 2020. <https://www.gefio.org/council-documents/annual-performance-report-2020>

GEF-IEO. APR 2004 <https://www.gefio.org/content/gef-annual-performance-review-2004>

GEF. 2008. [Guidelines for GEF Agencies in Conducting Terminal Evaluations](#)

GEF. 2017. [Guidelines for GEF Agencies in Conducting Terminal Evaluation for Full-sized Projects](#)

GEF. 2019. Evaluation Policy <https://www.gefio.org/evaluations/gef-evaluation-policy-2019>

IDB Office of Evaluation and Oversight <https://www.iadb.org/en/ove/home>

IFAD Independent Evaluation Office <https://www.ifad.org/en/web/ioe/about>

UN Women <https://www.unwomen.org/en/about-us/accountability>

UNDP Evaluation Resource Center <https://erc.undp.org/>

UNEG <http://www.unevaluation.org/document/detail/607>

UNEP Evaluation Office - <https://www.unenvironment.org/about-un-environment-programme/evaluation-office/overview>

UNIDO Independent Evaluation Division <https://www.unido.org/resources/evaluation-and-internal-oversight>

UNIDO. 2019. Charter of the Evaluation and Internal Oversight.
https://www.unido.org/sites/default/files/files/2019-04/DGB-2019-07_DG_Bulletin_Charter_of_the_Evaluation_and_Internal_Oversight.pdf

World Bank IEG. 2018. [Guidelines for Reviewing World Bank Implementation Completion and Results Reports: A Manual for Evaluators](#)

World Bank Independent Evaluation Group <https://ieg.worldbankgroup.org/>

ANNEX 6: TERMINAL EVALUATION REPORT REVIEW GUIDELINES

TERMINAL EVALUATION REPORT REVIEW GUIDELINES

- The assessments in the terminal evaluation reviews will be based largely on the information presented in the terminal evaluation report. If insufficient information is presented in a terminal evaluation report to assess a specific issue such as, for example, quality of the project's monitoring and evaluation system or a specific aspect of sustainability, then the preparer of the terminal evaluation reviews will briefly indicate so in that section and elaborate more if appropriate in the section of the review that addresses quality of report. If the review's preparer possesses other first-hand information such as, for example, from a field visit to the project, and this information is relevant to the terminal evaluation reviews, then it should be included in the reviews only under the heading "Additional independent information available to the reviewer." The preparer of the terminal evaluation review will take into account all the independent relevant information when verifying ratings.

B.1 Criteria for Outcome Ratings

- Based on the information provided in the terminal evaluation report, the terminal evaluation review will make an assessment of the extent to which the project's major relevant objectives were achieved or are expected to be achieved⁷², relevance of the project results, and the project's cost-effectiveness. The ratings on the outcomes of the project will be based on performance on the following criteria:⁷³

- **Relevance.** Were project outcomes consistent with the focal area/operational program strategies and country priorities? Explain.
- **Effectiveness.** Are project outcomes commensurate with the expected outcomes (as described in the project document) and the problems the project was intended to address (that is, the original or modified project objectives)?
- **Efficiency.** Include an assessment of outcomes and impacts in relation to inputs, costs, and implementation times based on the following questions: Was the project cost-effective? How does the project's cost/time versus outcomes equation compare to that of similar projects? Was the project implementation delayed due to any bureaucratic, administrative, or political problems and did that affect cost-effectiveness?

⁷² *Objectives* are the intended physical, financial, institutional, social, environmental, or other development results to which a project or program is expected to contribute (OECD DAC 2002).

⁷³ *Outcomes* are the likely or achieved short-term and medium-term effects of an intervention's outputs. Outputs are the products, capital goods, and services that result from a development intervention; these may also include changes resulting from the intervention that are relevant to the achievement of outcomes (OECD DAC 2002). For the GEF, environmental outcomes are the main focus.

- An overall rating will be provided according to the achievement and shortcomings in the three criteria ranging from highly satisfactory, satisfactory, moderately satisfactory, moderately unsatisfactory, unsatisfactory, highly unsatisfactory, and unable to assess.
- The reviewer of the terminal evaluation will provide a rating under each of the three criteria (relevance, effectiveness, and efficiency). Relevance of outcomes will be rated on a binary scale: a 'satisfactory' or an 'unsatisfactory' rating will be provided. If an 'unsatisfactory' rating has been provided on this criterion, the overall outcome achievement rating may not be higher than "unsatisfactory". Effectiveness and Efficiency will be rated as following:
 - (a) **Highly satisfactory.** The project had no shortcomings.
 - (b) **Satisfactory.** The project had minor shortcomings.
 - (c) **Moderately satisfactory.** The project had moderate shortcomings.
 - (d) **Moderately unsatisfactory.** The project had noticeable shortcomings.
 - (e) **Unsatisfactory.** The project had major shortcomings.
 - (f) **Highly unsatisfactory.** The project had severe shortcomings.
 - (g) **Unable to assess.** The reviewer was unable to assess outcomes on this dimension.
- The calculation of the overall outcomes score of projects will consider all three criteria, of which relevance criterion will be applied first - the overall outcome achievement rating may not be higher than "unsatisfactory". The second constraint that is applied is that the overall outcome achievement rating may not be higher than the "effectiveness" rating. The third constraint that is applied is that the overall rating may not be higher than the average score of effectiveness and efficiency criteria calculated using the following formula:

Outcomes = (b + c) ÷ 2

- In case the average score is lower than the score obtained after application of the first two constraints, then the average score will be the overall score. The score will then be converted into an overall rating with mid values being rounded up upwards.

B.2 Impacts

- Has the project achieved impacts, or is it likely that outcomes will lead to the expected impacts? Impacts will be understood to include positive and negative, primary and secondary long-term effects produced by a development intervention. They could be produced directly or indirectly and could be intended or unintended. The terminal evaluation review's preparer will take note of any mention of impacts, especially global environmental benefits, in the terminal evaluation report including the likelihood that the project outcomes will contribute to their achievement. Negative impacts mentioned in the terminal evaluation report should be noted and recorded in section 2 of the terminal evaluation reviews template in the subsection on "Issues that require follow-up." Although project impacts will be described, they will not be rated.

B.3 Criteria for Sustainability Ratings

- Sustainability will be understood as the likelihood of continuation of project benefits after completion of project implementation (GEF 2000). To assess sustainability, the terminal evaluation reviewer will identify and assess the key risks that could undermine continuation of benefits at the time of the evaluation. Some of these risks might include the absence of or inadequate financial resources, an enabling legal framework, commitment from key stakeholders, and enabling economy. The following four types of risk factors will be assessed by the terminal evaluation reviewer to rate the likelihood of sustainability of project outcomes: financial, sociopolitical, institutional frameworks and governance, and environmental.
- The following questions provide guidance to assess if the factors are met:
 - (a) **Financial resources.** What is the likelihood that financial resources will be available to continue the activities that result in the continuation of benefits (income-generating activities, and trends that may indicate that it is likely that in future there will be adequate financial resources for sustaining project outcomes)?
 - (b) **Sociopolitical.** Are there any social or political risks that can undermine the longevity of project outcomes? What is the risk that the level of stakeholder ownership is insufficient to allow for project outcomes/benefits to be sustained? Do the various key stakeholders see in their interest that the project benefits continue to flow? Is there sufficient public/stakeholder awareness in support of the long-term objectives of the project?
 - (c) **Institutional framework and governance.** Do the legal frameworks, policies, and governance structures and processes pose any threat to the continuation of project benefits? While assessing this parameter, consider if the required systems for accountability and transparency, and the required technical know-how, are in place.
 - (d) **Environmental.** Are there any environmental risks that can undermine the future flow of project environmental benefits? The terminal evaluation should assess whether certain activities in the project area will pose a threat to the sustainability of project outcomes. For example, construction of dam in a protected area could inundate a sizable area and thereby neutralize the biodiversity-related gains made by the project.
- The reviewer will provide a rating as follows:
 - (a) **Likely.** There are no risks affecting that criterion of sustainability.
 - (b) **Moderately likely.** There are moderate risks that affect that criterion of sustainability.
 - (c) **Moderately unlikely.** There are significant risks that affect that criterion of sustainability.
 - (d) **Unlikely.** There are severe risks affecting that criterion of sustainability.
 - (e) **Unable to assess.** Unable to assess risk on this dimension.
 - (f) **Not applicable.** This dimension is not applicable to the project.

B.4 Criteria for Assessment of Quality of Project M&E Systems

- GEF projects are required to develop M&E plans by the time of work program inclusion, to appropriately budget M&E plans, and to fully carry out the M&E plan during implementation. Project managers are also expected to use the information generated by the M&E system during project implementation to improve and adapt the project to changing situations. Given the long-term nature of many GEF projects, projects are also encouraged to include long-term monitoring plans that measure results (such as environmental results) after project completion. Terminal evaluation reviews will include an assessment of the achievement and shortcomings of M&E systems.

- (a) **M&E design.** Project should have a sound M&E plan to monitor results and track progress in achieving project objectives. An M&E plan should include a baseline (including data, methodology, and so on), SMART (specific, measurable, achievable, realistic, and timely) indicators and data analysis systems, and evaluation studies at specific times to assess results. The time frame for various M&E activities and standards for outputs should have been specified. Questions to guide this assessment include: In retrospect, was the M&E plan at entry practicable and sufficient (sufficient and practical indicators identified; timely baseline; targets created; effective use of data collection; analysis systems including studies and reports; practical organization and logistics in terms of what, who, and when for M&E activities)?
 - (b) **M&E plan implementation.** The M&E system was in place and allowed the timely tracking of results and progress toward project objectives throughout the project. Annual project reports were complete, accurate, and with well-justified ratings. The information provided by the M&E system was used to improve and adapt project performance. An M&E system should be in place with proper training for parties responsible for M&E activities to ensure that data will continue to be collected and used after project closure. Question to guide this assessment include: Did the project M&E system operate throughout the project? How was M&E information used during the project? Did it allow for tracking of progress toward project objectives? Did the project provide proper training for parties responsible for M&E activities to ensure data will continue to be collected and used after project closure?
 - (c) **Other questions.** This includes questions on funding and whether the M&E system was a good practice.
 - (d) Was sufficient funding provided for M&E — in the budget included in the project document?
 - (e) Was sufficient and timely funding provided – for M&E during project implementation?
 - (f) Can the project M&E system be considered – a good practice?
- A number rating 1–6 will be provided for each criterion according to the achievement and shortcomings with highly satisfactory = 6, satisfactory = 5, moderately satisfactory = 4, moderately

unsatisfactory = 3, unsatisfactory = 2, highly unsatisfactory = 1, and unable to assess = no rating. The reviewer of the terminal evaluation will provide a rating under each of the three criteria (M&E design, M&E plan implementation, and M&E properly budgeted and funded) as follows:

- (a) Highly satisfactory. There were no shortcomings in that criterion of the project M&E system.
- (b) Satisfactory. There were minor shortcomings in that criterion of the project M&E system.
- (c) Moderately satisfactory. There were moderate shortcomings in that criterion of the project M&E system.
- (d) Moderately unsatisfactory. There were significant shortcomings in that criterion of the project M&E system.
- (e) Unsatisfactory. There were major shortcomings in that criterion of the project M&E system.
- (f) Highly unsatisfactory. There was no project M&E system.

The rating for M&E during implementation will be the overall rating of the M&E system:

Rating on the Quality of the Project Monitoring and Evaluation System = b

B.5 Criteria for Assessment of Quality of Terminal Evaluation Reports

- The ratings on quality of terminal evaluation reports will be assessed using the following criteria:

- (a) The report presents an assessment of all relevant outcomes and achievement of project objectives in the context of the focal area program indicators if applicable.
- (b) The report was consistent, the evidence presented was complete and convincing, and ratings were well substantiated.
- (c) The report presented a sound assessment of sustainability of outcomes.
- (d) The lessons and recommendations are supported by the evidence presented and are relevant to the portfolio and future projects.
- (e) The report included the actual project costs (totals, per activity and per source) and actual co-financing used.
- (f) The report included an assessment of the quality of the M&E plan at entry, the M&E system used during implementation, and whether the information generated by the M&E system was used for project management.

- A number rating 1–6 will be provided for each criterion according to the achievement and shortcomings with highly satisfactory = 6, satisfactory = 5, moderately satisfactory = 4, moderately unsatisfactory = 3, unsatisfactory = 2, highly unsatisfactory = 1, and unable to assess = no rating.

Each criterion to assess the quality of the terminal evaluation will be rated as follows:

- (a) Highly satisfactory. There were no shortcomings in the terminal evaluation on this criterion.
 - (b) Satisfactory. There were minor shortcomings in the terminal evaluation on this criterion.
 - (c) Moderately satisfactory. There were moderate shortcomings in the terminal evaluation on this criterion.
 - (d) Moderately unsatisfactory. There were significant shortcomings in the terminal evaluation on this criterion.
 - (e) Unsatisfactory. There were major shortcomings in the terminal evaluation on this criterion.
 - (f) Highly unsatisfactory. There were severe shortcomings in the terminal evaluation on this criterion.
- The first two criteria (of all relevant outcomes and achievement of project objectives and report consistency and substantiation of claims with proper evidence) are more important and have therefore been assigned a greater weight. The quality of the terminal evaluation reports will be calculated by the following formula:

Quality of the Terminal Evaluation Report = $0.3 \times (a + b) + 0.1 \times (c + d + e + f)$

The total number will be rounded and converted to the scale of highly satisfactory to highly unsatisfactory.

B.6 Assessment of Processes Affecting Attainment of Project Outcomes and Sustainability

- This section of the terminal evaluation review will summarize the factors or processes related to implementation delays and co-financing that may have affected attainment of project results. This section will summarize the description in the terminal evaluation on key causal linkages of these factors:

- (a) Co-financing and project outcomes and sustainability. If there was a difference in the level of expected co-financing and actual co-financing, what were the reasons for it? To what extent did materialization of co-financing affect project outcomes and/or sustainability? What were the causal linkages of these effects?

- (b) Delays and project outcomes and sustainability. If there were delays, what were the reasons for them? To what extent did the delay affect project outcomes and/or sustainability? What were the causal linkages of these effects?
- (c) Country ownership and sustainability. Assess the extent to which country ownership has affected project outcomes and sustainability. Describe the ways in which it affected outcomes and sustainability, highlighting the causal links.

ANNEX 7: TERMINAL EVALUATION REVIEW FORM – SECTION “QUALITY OF THE TERMINAL EVALUATION REPORT”⁷⁴

Quality of the Terminal Evaluation Report

A six-point rating scale is used for each sub-criteria and overall rating of the terminal evaluation report (Highly Satisfactory to Highly Unsatisfactory)

Criteria	GEF IEO comments	Rating
To what extent does the report contain an assessment of relevant outcomes and impacts of the project and the achievement of the objectives?	The report thoroughly assesses project outcomes and impacts using the project’s results framework.	S
To what extent is the report internally consistent, the evidence presented complete and convincing, and ratings well substantiated?	The report is consistent, and the evidence presented is complete and convincing.	S
To what extent does the report properly assess project sustainability and/or project exit strategy?	Although the TE does not provide ratings for different aspects of sustainability, it does break down the risks in these areas.	S
To what extent are the lessons learned supported by the evidence presented and are they comprehensive?	The lessons learned are supported with evidence and are comprehensive.	S
Does the report include the actual project costs (total and per activity) and actual co-financing used?	The report does include actual co-financing and project costs; however, it doesn’t explain the reasons behind the difference between expected and actualized co-financing or the implications.	MS
Assess the quality of the report’s evaluation of project M&E systems:	The report satisfactorily assesses the project’s M&E system.	S
Overall TE Rating		S

⁷⁴ Terminal Evaluation Review form, GEF Independent Evaluation Office, APR 2020

ANNEX 8: INTERVIEW GUIDE – FOR AGENCIES

Interview Guide – for Agencies

Review of the GEF-IEO Terminal Evaluation (TE) Validation Process

August 2020

The GEF-IEO is currently undertaking a Review of its TE Validation Process. The Process is used to validate the Terminal Evaluations submitted by Agencies recipient of GEF funds. This Interview is part of the consultations, which will be used to inform the findings of the Review process.

Below are some of the key questions to guide the 1 hour interviews.

Thank you for participating. Your feedback is very valuable to us.

Questions

- (a) Please explain your involvement with the GEF-IEO TE Validation process. When and how have you been involved?
- (b) To what extent is the GEF-IEO terminal evaluation validation process transparent? ^[L]_[SEP]
 - 1. Is the GEF-IEO TE validation process clear and information on criteria easily accessible to the participating agencies?
 - 2. Are there opportunities for the IEO and GEF Agencies to engage in discussions about the results obtained by the validation process?
 - 3. What are the areas where further needs/opportunities for GEF-IEO capacity building related to the process?
- (c) To what extent does the GEF-IEO TE validation process contributes to enhancing the credibility of you Agency's evaluations? ^[L]_[SEP]
 - 1. Are the current GEF-IEO TE validation criteria and ratings adequate to validate the TE information?
 - 2. Does the TE validation process contributes to assuring the quality of the assessments made by the TEs?
 - 3. Are there adjustments needed in the process, parameters, criteria and ratings and tools to enhance compliance of a diverse community of Partner Agencies?
- (d) What are the strengths of the GEF-IEO TE validation process? ^[L]_[SEP]
- (e) What are the main weaknesses or gaps in the current GEF-IEO TE validation process? ^[L]_[SEP]
- (f) Are you familiar with other TE validation processes used by other international development agencies? If so, how does the GEF IEO TE validation process compare with them?

ANNEX 9: RESULTS OF THE GEF-IEO VALIDATIONS AND REVIEWS FOR THE 12 PROJECTS ANALYZED

ADB – 3589 Criteria	Final PIR	IA Terminal Evaluation	IA Evaluation Office Review	GEF IEO Review
Project Outcomes	S	UA	-	S
Sustainability of Outcomes		L	-	L
M&E Design		S	-	S
M&E Implementation		S	-	S
Quality of Implementation		S	-	S
Quality of Execution		S	-	MS
Quality of the Terminal Evaluation Report		-	-	S

IFAD – 3362- Criteria	Final PIR	IA Terminal Evaluation	IA Evaluation Office Review	GEF IEO Review
Project Outcomes	-	S	-	MS
Sustainability of Outcomes		MU	-	MU
M&E Design		MU	-	MS
M&E Implementation		MU	-	MU
Quality of Implementation		S	-	S
Quality of Execution		N/A	-	MS
Quality of the Terminal Evaluation Report		-	-	S

WWF 5771 Criteria	Final PIR	IA Terminal Evaluation	IA Evaluation Office Review	GEF IEO Review
Project Outcomes	UA	S-HS	--	S
Sustainability of Outcomes		ML	--	ML
M&E Design		UA	--	S
M&E Implementation		S	--	S
Quality of Implementation		S	--	S
Quality of Execution		S	--	S
Quality of the Terminal Evaluation Report		--	--	S

WB 5439 Criteria	Final PIR	IA Terminal Evaluation	IA Evaluation Office Review	GEF IEO Review
Project Outcomes	S	S	-	S
Sustainability of Outcomes	-	NR	-	UA
M&E Design	-	NR	-	MS
M&E Implementation	-	NR	-	UA
Quality of Implementation	-	NR	-	S
Quality of Execution	-	NR	-	S
Quality of the Terminal Evaluation Report	-	-	-	MS

UNIDO 3908 Criteria	Final PIR	IA Terminal Evaluation	IA Evaluation Office Review	GEF IEO Review
Project Outcomes	S	S	-	S
Sustainability of Outcomes		ML	-	ML
M&E Design		S	-	S
M&E Implementation		S	-	S
Quality of Implementation		S	-	S
Quality of Execution		MS	-	MS
Quality of the Terminal Evaluation Report		S	-	S

Criteria – JOINT EBRD/UNIDO 3593	Final PIR	IA Terminal Evaluation	IA Evaluation Office Review	GEF IEO Review
Project Outcomes	-	S	-	S
Sustainability of Outcomes		L	-	ML
M&E Design		HS	-	MS
M&E Implementation		S	-	MS
Quality of Implementation		S	-	MS
Quality of Execution		S	-	S
Quality of the Terminal Evaluation Report		-	-	MS

Criteria IFAD 3362	Final PIR	IA Terminal Evaluation	IA Evaluation Office Review	GEF IEO Review
Project Outcomes	-	S	-	MS
Sustainability of Outcomes		MU	-	MU
M&E Design		MU	-	MS
M&E Implementation		MU	-	MU
Quality of Implementation		S	-	S
Quality of Execution		N/A	-	MS
Quality of the Terminal Evaluation Report		-	-	S

Criteria IADB 5838	Final PIR	IA Terminal Evaluation	IA Evaluation Office Review	GEF IEO Review
Project Outcomes	S	MS	-	MS
Sustainability of Outcomes		ML	-	UA
M&E Design		NR	-	S
M&E Implementation		NR	-	MS
Quality of Implementation		NR	-	UA
Quality of Execution		NR	-	MU
Quality of the Terminal Evaluation Report		-	-	U

Criteria IADB 4113	Final PIR	IA Terminal Evaluation	IA Evaluation Office Review	GEF IEO Review
Project Outcomes	S	MS	-	MS
Sustainability of Outcomes		ML	-	ML
M&E Design		NR	-	S
M&E Implementation		NR	-	MS
Quality of Implementation		NR	-	UA
Quality of Execution		NR	-	MS
Quality of the Terminal Evaluation Report		-	-	MS

Criteria – FAO 3984	Final PIR	IA Terminal Evaluation	IA Evaluation Office Review	GEF IEO Review
Project Outcomes	-	MS	-	MS
Sustainability of Outcomes		ML	-	ML
M&E Design		MS	-	MS
M&E Implementation		MS	-	MS
Quality of Implementation		S	-	MS
Quality of Execution		MS	-	S
Quality of the Terminal Evaluation Report		-	-	MU

Criteria FAO 3777	Final PIR	IA Terminal Evaluation	IA Evaluation Office Review	GEF IEO Review
Project Outcomes	MS	MU	--	U
Sustainability of Outcomes		ML	--	ML
M&E Design		MU	--	MS
M&E Implementation		U	--	U
Quality of Implementation		MU	--	MU
Quality of Execution		MU	--	U
Quality of the Terminal Evaluation Report		--	--	MS

Criteria EBRD 3535	Final PIR	IA Terminal Evaluation	IA Evaluation Office Review	GEF IEO Review
Project Outcomes	HS	UA ⁷⁵	--	S
Sustainability of Outcomes	--	S ⁷⁶	--	ML
M&E Design	--	UA	--	U
M&E Implementation	--	S	--	MU
Quality of Implementation	--	HS	--	MS
Quality of Execution	--	S	--	MS
Quality of the Terminal Evaluation Report	--	--	--	MS

⁷⁵ The TE does not provide overall relevance, effectiveness, and efficiency ratings.

⁷⁶ The IA TE used a different scale for sustainability.

Criteria CI 9163	Final PIR	IA Terminal Evaluation	IA Evaluation Office Review	GEF IEO Review
Project Outcomes	S	HS	-	S
Sustainability of Outcomes		ML	-	ML
M&E Design		HS	-	S
M&E Implementation		HS	-	S
Quality of Implementation		HS	-	S
Quality of Execution		HS	-	S
Quality of the Terminal Evaluation Report		-	-	S

ANNEX 10: UNEP MATRIX (EXAMPLE OF ONE CRITERIA)

A. Strategic Relevance

The evaluation will assess *‘the extent to which the activity is suited to the priorities and policies of the target group, recipient and donor’*. The evaluation will include an assessment of the project’s relevance in relation to UNEP’s mandate and its alignment with UNEP’s policies and strategies at the time of project approval. Under strategic relevance an assessment of the complementarity of the project with other interventions addressing the needs of the same target groups will be made. This criterion comprises four elements (see below) and the overall rating for Strategic Relevance is calculated through the Weighted Ratings Table:

- (a) *Alignment to the UNEP Medium Term Strategy⁷⁷ (MTS), Programme of Work (POW) and Strategic Priorities*
- (b) *Alignment to Donor/GEF/Partner Strategic Priorities*
- (c) *Relevance to Global, Regional, Sub-regional and National Environmental Priorities*
- (d) *Complementarity with Existing Interventions/ Coherence⁷⁸*

A1. Alignment to the UNEP Medium Term Strategy⁷⁹ (MTS), Programme of Work (POW) and Strategic Priorities

The evaluation should assess the project’s alignment with the MTS and POW under which the project was approved and include, in its narrative, reflections on the scale and scope of any contributions made to the planned results reflected in the relevant MTS and POW. UNEP strategic priorities include the Bali Strategic Plan for Technology Support and Capacity Building 80 (BSP) and South-South Cooperation (S-SC). The BSP relates to the capacity of governments to: comply with international agreements and obligations at the national level; promote, facilitate and finance environmentally sound technologies and to strengthen frameworks for developing coherent international environmental policies. S-SC is regarded as the exchange of resources, technology and knowledge between developing countries.

Highly Unsatisfactory	Unsatisfactory	Moderately Unsatisfactory	Moderately Satisfactory	Satisfactory	Highly Satisfactory
The project’s implementation strategies and delivered contributions (results) show:	The project’s implementation strategies and delivered contributions (results) show: - weak alignment (i.e. consistency) with a few	The project’s implementation strategies and delivered contributions (results) show: - partial alignment (i.e. consistency) with many	The project’s implementation strategies and delivered contributions (results) show:	The project’s implementation strategies and delivered contributions (results) show:	The project’s implementation strategies and delivered contributions (results) show:

⁷⁷ UNEP’s Medium Term Strategy (MTS) is a document that guides UNEP’s programme planning over a four-year period. It identifies UNEP’s thematic priorities, known as Sub-programmes (SP), and sets out the desired outcomes, known as Expected Accomplishments (EAs), of the Sub-programmes.

⁷⁸ This sub-category is consistent with the new criterion of ‘Coherence’ introduced by the OECD-DAC in 2019.

⁷⁹ UNEP’s Medium Term Strategy (MTS) is a document that guides UNEP’s programme planning over a four-year period. It identifies UNEP’s thematic priorities, known as Sub-programmes (SP), and sets out the desired outcomes, known as Expected Accomplishments (EAs), of the Sub-programmes.

⁸⁰ <http://www.unep.fr/ozonaction/about/bsp.htm>

A. Strategic Relevance

The evaluation will assess ‘the extent to which the activity is suited to the priorities and policies of the target group, recipient and donor’. The evaluation will include an assessment of the project’s relevance in relation to UNEP’s mandate and its alignment with UNEP’s policies and strategies at the time of project approval. Under strategic relevance an assessment of the complementarity of the project with other interventions addressing the needs of the same target groups will be made. This criterion comprises four elements (see below) and the overall rating for Strategic Relevance is calculated through the Weighted Ratings Table:

- (a) Alignment to the UNEP Medium Term Strategy⁷⁷ (MTS), Programme of Work (POW) and Strategic Priorities
- (b) Alignment to Donor/GEF/Partner Strategic Priorities
- (c) Relevance to Global, Regional, Sub-regional and National Environmental Priorities
- (d) Complementarity with Existing Interventions/ Coherence⁷⁸

A1. Alignment to the UNEP Medium Term Strategy⁷⁹ (MTS), Programme of Work (POW) and Strategic Priorities

The evaluation should assess the project’s alignment with the MTS and POW under which the project was approved and include, in its narrative, reflections on the scale and scope of any contributions made to the planned results reflected in the relevant MTS and POW. UNEP strategic priorities include the Bali Strategic Plan for Technology Support and Capacity Building 80 (BSP) and South-South Cooperation (S-SC). The BSP relates to the capacity of governments to: comply with international agreements and obligations at the national level; promote, facilitate and finance environmentally sound technologies and to strengthen frameworks for developing coherent international environmental policies. S-SC is regarded as the exchange of resources, technology and knowledge between developing countries.

Highly Unsatisfactory	Unsatisfactory	Moderately Unsatisfactory	Moderately Satisfactory	Satisfactory	Highly Satisfactory
- lack of alignment (i.e. not consistent) with any parts of UNEP’s mandate and thematic priorities, as represented in the Medium-Term Strategy and Programme of Work under which the project was approved.	secondary or peripheral parts of UNEP’s mandate and thematic priorities, as represented in the Medium-Term Strategy and Programme of Work under which the project was approved.	secondary or peripheral parts of UNEP’s mandate and thematic priorities, as represented in the Medium-Term Strategy and Programme of Work under which the project was approved.	- partial alignment (i.e. consistency) with a single key focus of UNEP’s mandate and thematic priorities, as represented in the Medium-Term Strategy and Programme of Work under which the project was approved.	- full alignment (i.e. consistency) with more than one key focal area of UNEP’s mandate and thematic priorities, as represented in the Medium-Term Strategy and Programme of Work under which the project was approved. BUT No anticipated identifiable contribution to reported	- full alignment (i.e. consistency) with more than one key focal area of UNEP’s mandate and thematic priorities, as represented in the Medium-Term Strategy and Programme of Work under which the project was approved AND Anticipated identified contribution(s) to reported

A. Strategic Relevance

The evaluation will assess ‘the extent to which the activity is suited to the priorities and policies of the target group, recipient and donor’. The evaluation will include an assessment of the project’s relevance in relation to UNEP’s mandate and its alignment with UNEP’s policies and strategies at the time of project approval. Under strategic relevance an assessment of the complementarity of the project with other interventions addressing the needs of the same target groups will be made. This criterion comprises four elements (see below) and the overall rating for Strategic Relevance is calculated through the Weighted Ratings Table:

- (a) *Alignment to the UNEP Medium Term Strategy⁷⁷ (MTS), Programme of Work (POW) and Strategic Priorities*
- (b) *Alignment to Donor/GEF/Partner Strategic Priorities*
- (c) *Relevance to Global, Regional, Sub-regional and National Environmental Priorities*
- (d) *Complementarity with Existing Interventions/ Coherence⁷⁸*

A1. Alignment to the UNEP Medium Term Strategy⁷⁹ (MTS), Programme of Work (POW) and Strategic Priorities

The evaluation should assess the project’s alignment with the MTS and POW under which the project was approved and include, in its narrative, reflections on the scale and scope of any contributions made to the planned results reflected in the relevant MTS and POW. UNEP strategic priorities include the Bali Strategic Plan for Technology Support and Capacity Building 80 (BSP) and South-South Cooperation (S-SC). The BSP relates to the capacity of governments to: comply with international agreements and obligations at the national level; promote, facilitate and finance environmentally sound technologies and to strengthen frameworks for developing coherent international environmental policies. S-SC is regarded as the exchange of resources, technology and knowledge between developing countries.

Highly Unsatisfactory	Unsatisfactory	Moderately Unsatisfactory	Moderately Satisfactory	Satisfactory	Highly Satisfactory
				Expected Accomplishment indicator(s)	Expected Accomplishment indicator(s)

ANNEX 11: TERMS OF REFERENCE

SHORT-TERM CONSULTANT FOR THE REVIEW OF THE TERMINAL EVALUATION REVIEW PROCESS

Terms of Reference

Short-Term Consultant for the review of the Terminal Evaluation Review Process

The Global Environment Facility (GEF) provides support to address global environmental concerns related to biodiversity, climate change, international waters, land degradation, and chemicals. Since its inception in 1991, the GEF has provided developing countries and countries with economies in transition US \$ 20 billion in grants. These grants are implemented on ground through a network of 18 accredited agencies. The GEF receives its funds through a four-year replenishment.

The GEF Independent Evaluation Office (GEF IEO) has a central role in ensuring the independent evaluation function within the GEF. The GEF IEO is based in Washington DC. It is administered by the World Bank but is independent of its management as well as the management of the GEF. Its Director reports directly to the GEF Council, the GEF governing body. All contracts with the IEO are World Bank contracts. More information about the GEF IEO can be found at Office's website: www.gefeo.org.

The GEF Agencies implement GEF supported projects in recipient countries. At the end of implementation completion of a project, GEF Agencies prepare a terminal evaluation to give an account of implementation experience and project performance. The GEF IEO validates these terminal evaluations to ensure that the project performance ratings are comparable, consistent and evidence based. The validation results in a terminal evaluation review report, which includes validated performance ratings that cover areas such as outcome, sustainability, implementation and M&E. These ratings are incorporated in GEF IEO's terminal evaluation review dataset and analyzed to assess performance trends, correlations and causal relationships. Analysis of these performance ratings along with qualitative information from the terminal evaluations is presented regularly in the GEF IEO's Annual Performance Report. The data is also used in several evaluations prepared by the GEF IEO. The process for terminal evaluation validation was instituted in 2004. Since then there have been minor changes in the process along with that in the instrument used for validation

The GEF IEO is conducting an evaluation of the self-evaluation systems across the GEF Partnership. The evaluation assesses the Agency self-evaluation systems along with the GEF IEO's validation process. The review of the terminal evaluation validation process will be undertaken within the framework of the evaluation of the self-evaluation systems. The purpose of the review is to identify gaps in the GEF IEO's validation process, measures that may be used to strengthen the process, and the cost implications of these measures. The review will seek to answer following questions:

- (a) To what extent is the terminal evaluation validation process transparent?
- (b) To what extent does the validation process enhance the credibility of ratings based on the evidence provided in the terminal evaluations?
- (c) How does the GEF IEO's terminal evaluation validation process compare with that of other international development agencies?
- (d) What are the good practices and lessons that may be incorporated in the GEF IEO's validation process?

The review will be conducted from July to September 2020.

Responsibilities and Accountabilities

Ms Claudia Marcondes was selected based on a competitive selection process. The consultant will perform following tasks:

1. Develop a framework for reviewing the IEO TE validation process.
2. Assess the GEF IEO's validation process. Review relevant documents (Terminal evaluations of GEF agencies) and conduct interviews of the GEF IEO staff and key stakeholders, to reflect on all aspects of the process including quality, and efficiency.
3. Compare GEF IEO's process with that of other international development agencies to provide examples of good practice (including those that implement activities through other Agencies). Review documents and conduct interviews of the relevant agencies.

4. Identify any areas for improvement which would include a discussion of implications for efficiency and quality.
5. Prepare a report that presents the findings of the review.

Compensation, payments and other arrangements

96. The assignment will be for an initial period of 30 work days. The consultant daily rate will be USD xx per day net. International travel is not anticipated. However, in case such travel is required, it would be undertaken following the World Bank rules and procedures for such travel. The consultant will report to Ms Geeta Batra, Chief Evaluation Officer, IEO.