

62nd GEF Council Meeting

June 21 – 23, 2022

Washington, D.C.

GUIDELINES FOR CONDUCTING PROGRAM EVALUATION

(Prepared by the Independent Evaluation Office of the GEF)

Table of Contents

Summary	iii
Applicability	iii
Key Words	iii
I. Introduction.....	1
II. Guidelines for Conduct of Program Evaluations	2
A. Roles and Responsibilities	2
Lead GEF Agency.....	2
Evaluators	3
GEF Operational Focal Points	4
Other GEF Agencies	4
B. Content of the Evaluation Report.....	4
Introduction and Background.....	4
Methodology	6
Program Performance Additionality	7
Program Implementation.....	10
Program Monitoring and Evaluation.....	11
Program Cofinancing	12
Alignment with GEF Policies	12
Lessons and Recommendations	13
Performance Ratings	13
C. Submission of Program Evaluation.....	14
D. Use of Program Evaluation.....	14
Annex: Rating Scales	15
A. Outcome Ratings.....	15
B. Sustainability Ratings	16
C. Program Implementation Ratings	17
D. Program M&E Ratings.....	20
Bibliography	23
Table 1: Ratings	14

SUMMARY

The *GEF Evaluation Policy* (2019) requires that the respective lead Global Environment Facility (GEF) Agency will evaluate a program at the end of its implementation to provide a comprehensive and systematic assessment of its performance and the factors that affect performance. Program evaluations are expected to support accountability, learning, transparency, knowledge sharing, and decision making in the GEF. The purpose of these guidelines is to support GEF Agencies in conducting program evaluations, and to facilitate the consistency and completeness of their reporting.

APPLICABILITY

Program evaluation reports are required, in English, for all programs funded through the GEF Trust Fund and, where applicable, other GEF-administered trust funds.¹

These guidelines are relevant and applicable to the actors across the GEF Partnership who have a role in the conduct of program evaluation. These include the lead GEF Agency of a program, evaluators, other GEF Agencies involved in implementation, and the GEF operational focal points (OFPs) of the countries in which a program is implemented.

These guidelines are aimed at programs that were approved from GEF-6 onwards but may also be used to evaluate programs that were approved earlier. Even though a program evaluation will cover all child projects implemented as part of the program, Agencies will continue to follow the applicable terminal evaluation guidelines to prepare terminal evaluations for these child projects.

KEY WORDS

Program; Program Evaluation; Coordinating Project; Child Project; Additionality; Theory of Change; Relevance; Integration; System; System Boundary; Transformation; Scale; Effectiveness; Efficiency; Results; Output; Outcome; Sustainability; Impact; Knowledge Management; Cofinancing; Implementation; Monitoring and Evaluation; Core Indicators; Lessons.

¹ In August 2006, the Council decided that all GEF operational policies, procedures, and governance structures are applied to these funds unless the Least Developed Countries Fund/Special Climate Change Fund (LDCF/SCCF) Council decides otherwise (GEF/C.29/5, *Governance of the Climate Change Funds*, Joint Summaries of the Chairs, 29th GEF Council Meeting, Decision on Agenda Item 6). Accordingly, the LDCF/SCCF currently applies the GEF Monitoring and Evaluation Policy.

I. INTRODUCTION

1. The GEF Project and Program Cycle Policy defines the term “program” as “a longer-term and strategic arrangement of individual yet interlinked projects that aim at achieving large-scale impacts on the global environment” (GEF 2018, 4). GEF uses the program modality as one of its main means of providing support to recipient countries.² GEF programs often consist of a *coordinating child project* that provides the overall direction and support for program activities and several decentralized *child projects* that are tied together by a shared vision and framework.

2. Although the GEF has used the program modality since its inception, up to GEF-3 total funding through the modality was low. In GEF-4, the GEF Council endorsed the objectives and basic principles for programmatic approaches (GEF 2008). Thereafter, the share of GEF funding provided through programs increased substantially.

3. The nature of the programs has changed over time with a shift toward the use of *integrated approaches*. Although the GEF started providing support for the use of integrated approaches through its Operational Program on Integrated Ecosystem Management in 1999, the use of this approach was limited through GEF-5. In GEF-6 the GEF approved a set of integrated approach pilot programs. These pilots aimed to achieve transformational change at scale to address drivers of global environmental degradation (GEF IEO 2021). These pilots also placed a strong emphasis on knowledge management facilitated through a coordinating project. Other programs approved during GEF-6 also emphasized integrated approaches, system transformation, and addressing challenges at scale. Reliance on integrated approaches was deepened in GEF-7 through support for *impact programs* (GEF IEO 2021).

4. The GEF Evaluation Policy 2019 defines evaluation as “the systematic and impartial assessment of planned, ongoing, or completed activities, projects, programs in specific focal areas or sectors, policies, strategies and their implementation, or other topics relevant to the GEF partnership and organization” (GEF IEO 2022, 3). The term “program evaluation” combines two terms “program” and “evaluation.” Within the context of the GEF program cycle, it means evaluation of a GEF-supported program.

5. A program evaluation is expected to provide a comprehensive and systematic account of a program’s performance, including aspects related to its design, implementation, achievement of objectives and targets, and factors that affected performance. The evaluation is expected to discuss the strengths and weaknesses of a program and lessons that may be drawn from its implementation. Program evaluation is expected to support accountability, learning, transparency, knowledge sharing, and decision making in the GEF.

6. The Guidelines for Conducting Program Evaluation have been prepared within the framework of the GEF Evaluation Policy 2019 and draw on the Ethical Guidelines (2007) of the GEF IEO. The guidelines also draw upon international norms and standards such as *Applying Evaluation Criteria Thoughtfully* (OECD DAC 2021), *Norms and Standards for Evaluation* (UNEG

² Other modalities include full-size projects, medium-size projects, and enabling activities.

2017), and *Ethical Guidelines for Evaluation* (UNEG 2020). The guidelines are aimed at facilitating evaluation of GEF programs, especially those approved from GEF-6 onwards. They clarify the roles and responsibilities of the relevant actors and the content of the evaluation report, its submission, and use.

II. GUIDELINES FOR CONDUCT OF PROGRAM EVALUATIONS

A. Roles and Responsibilities

Lead GEF Agency

7. The lead GEF Agency³ will perform the following:
 - (a) Evaluate a GEF program close to or at the end of its implementation.⁴ The earliest it may initiate an evaluation is at a point when implementation of at least 60 percent of the child projects are complete and at least 80 percent of the GEF funding for the program has been utilized.⁵ The latest a program may be evaluated is up to six months after the end of program implementation.
 - (b) Develop specific terms of reference for a program’s evaluation and provide guidance, documentation, and support to the evaluation team. The documentation submitted at program framework document approval, and endorsement/approval of coordinating project and child projects, will be useful to the evaluators. Similarly, project implementation reports and midterm reviews of program activities will be useful.
 - (c) Archive information gathered during program implementation and evaluation and, at the request of the GEF Independent Evaluation Office (IEO), make it available for follow-up studies.
 - (d) Ensure that the evaluation team is composed of evaluators with expertise in the thematic area addressed by the program, integrated approaches and other important elements of the program’s strategy, fiduciary oversight, social and environmental safeguards, and gender-related concerns.

³ As of this writing, there are 18 GEF Agencies accredited by the GEF that may have direct access to GEF resources. Any of these Agencies may be involved in a program as its lead Agency.

⁴ End of implementation is understood as completion of the last program activity—including activities of child projects and the coordinating project but excluding program evaluation—supported through GEF funding.

⁵ The coordinating projects of some programs, especially programs approved in GEF-6, are designed for a shorter duration than may be necessary to meet the threshold conditions for starting program evaluation. In such a situation, the lead Agency may undertake program evaluation at the time of closing of the coordinating project even if the threshold conditions in terms of program progress have not been met. The evaluation report will clearly explain the reasons for conducting the evaluation earlier than required.

- (e) Ensure that the program evaluation team is independent, unbiased, and free of conflicts of interest as required by Minimum Requirement 3 of the GEF Evaluation Policy 2019. The evaluators should not have been involved in the design, approval, implementation, or execution of the program or its child projects.
- (f) Consult with other GEF Agencies that are involved in implementation of coordinating projects and/or implementation of child projects to develop the terms of reference for the evaluation.
- (g) Share the draft reports of the evaluation with the GEF OFPs in the countries covered by the program (i.e., the countries in which the program activities were implemented) for comment and transmit the final report to them.
- (h) Actively seek and address feedback from relevant stakeholders when developing the program evaluation's terms of reference and evaluation report.
- (i) Submit the program evaluation report to the GEF Portal within 6 months of the evaluation's completion and within 12 months of the end of program implementation.
- (j) Share the final evaluation report with the relevant stakeholders and ensure that the report is publicly available.

8. The credibility of program evaluation relies on independence from program implementation. GEF Agencies will use one of two broad approaches to ensure the independence of program evaluation: (1) the evaluation unit of an Agency commissions and/or manages program evaluation; (2) the operations unit of a GEF Agency manages and/or commissions the program evaluation, and its evaluation unit validates the information and performance ratings provided in the evaluation report.

Evaluators

9. The following guidelines shall be observed by the evaluators in conducting terminal evaluations:

- (a) Evaluators will have knowledge of, or will familiarize themselves with, GEF programs and strategies, and with relevant GEF policies such as those on program approval, the program cycle, monitoring and evaluation (M&E), cofinancing, fiduciary standards, gender, stakeholder engagement, and environmental and social safeguards.
- (b) Evaluators will take the perspectives of all relevant stakeholders into account. They will gather information on program performance and results from multiple sources to facilitate triangulation. They will seek the necessary contextual information to assess the significance and relevance of observed performance and results.

- (c) Evaluators will present a balanced account that is clear, candid, and consistent with evidence.
- (d) Evaluators will apply the evaluation criteria discussed in these guidelines (see paragraph 17). They will also evaluate performance in other areas that are important to the GEF. Where necessary, they may cover other topics to enhance utility of evaluation.
- (e) Evaluators will use the rating scales provided in the annex to these guidelines.
- (f) Evaluators will abide by the GEF Evaluation Office Ethical Guidelines (GEF IEO 2007).

GEF Operational Focal Points

10. Minimum Requirement 4 of the GEF Evaluation Policy 2019 requires involvement of the relevant OFPs in the program evaluation process. Where applicable, the OFPs will receive the draft report of an evaluation from the lead Agency for comment. The OFPs are encouraged to provide inputs while respecting the independent nature of the evaluation.

Other GEF Agencies

11. Other GEF Agencies that are involved in implementation of the program will facilitate the conduct of the evaluation by providing necessary support to the evaluation team. They will provide documentation that may be a useful source of information on implementation and results of the child projects. The Agencies will also provide data on utilization of GEF resources and materialization of cofinancing for child projects. When requested, they will facilitate the evaluation team in accessing child project sites for verification and help the team establish contact with key informants and project beneficiaries. They will provide feedback on the draft terms of reference and the draft evaluation report.

B. Content of the Evaluation Report

Introduction and Background

Program Information and Description

12. The evaluation report will provide general information on the program including its name, list of coordinating and child projects and their respective GEF project IDs; countries covered by the program; the lead GEF Agency and other Agencies involved in implementation; the program implementation team, the program evaluation team; the program milestone dates (e.g., start date, expected end date, actual implementation completion date, or anticipated completion date when the evaluation was conducted); GEF financing, and promised and materialized cofinancing. The report will list program objectives and expected outcomes, the program's strategy, a description of program activities, coordinating project and child projects, and the institutional arrangements for implementation.

Theory of Change

13. A program's theory of change describes the causal relationships through which a program would deliver its expected impact, along with the assumptions about the conditions necessary to make the causal relationships work. and provides a basis for evaluating the program. Where available, the program evaluation will use the theory of change presented in the program framework documents as a basis for evaluation. If necessary, the evaluators will update the theory in consultation with the key stakeholders to address gaps and reflect changes made to the program during implementation. Where the program framework documents do not explicitly provide a theory of change, the evaluators will develop one based on the information provided in the program documents and in consultation with the key stakeholders.

Box 1: Definitions of Key Terms

Impact: The positive and negative, primary and secondary, long-term effects produced by a project or program, directly or indirectly, intended or unintended.

Output: A product or service that results from the completion of activities implemented within a project or program.

Outcome: An intended or achieved short- or medium-term effect of a project or program's outputs.

Results: Include intervention outputs, outcomes, and progress toward longer-term impact including global environmental benefits. They should be discernible and measurable.

Sustainability: The continuation/likely continuation of positive effects from the intervention after it has come to an end, and its potential for scale-up and/or replication; interventions need to be environmentally as well as institutionally, financially, politically, culturally, and socially sustainable.

Source: GEF IEO 2019.

14. The program evaluation report will describe the program's theory of change including a description of the outputs, outcomes, intermediate states, and intended long-term impacts of the program; the causal pathways for the long-term impacts; mechanisms through which these impacts are to be achieved; and implicit and explicit assumptions. The theory should illustrate how child projects, along with the coordinating child project, contribute to achieving the program's intended outcomes and long-term impacts, including any planned interactions among the child projects that are expected to contribute to these ends. The theory of change should outline synergies the program intends to tap to achieve the intended systemic change, and the trade-offs involved in key choices made in program design. Where applicable, the theory of change should indicate how the given environmental challenges have been addressed through integration.

Scope of Evaluation

15. The evaluation report will define its scope in terms of the program's objectives, theory of change, supported activities (e.g., coordinating child project and other child projects), the context in which the program was designed and implemented, and program M&E data. The

scope will also be determined based on the questions the evaluation seeks to answer, the boundaries of the system covered, the features of the program and its components, geographical coverage, period under review, key users, and likely uses of the evaluation.

Methodology

16. The evaluation report will describe the methodology used to carry out the evaluation. It will discuss the evaluation questions, analytical framework, and evaluation criteria with reference to the program theory of change; information sources; and limitations. The program's theory of change will provide a basis for the type of evidence needed to assess whether the theory is valid, and to assess whether the actual program results and causal relationships are consistent with the theory.

17. The Organization for Economic Co-operation and Development Development Assistance Committee (OECD DAC) defines six main criteria for evaluation—relevance, coherence, effectiveness, efficiency, impact, and sustainability (OECD DAC 2021). These criteria will be applied to evaluate GEF programs. The evaluation report will also cover other areas of significance to the GEF such as additionality, program M&E, performance of GEF Agencies, materialization of cofinancing, application of environmental and social safeguards, gender equality, stakeholder engagement including involvement of the private sector, integration, and knowledge management. These last two considerations—integration and knowledge management—have been key features of the GEF programs approved from GEF-6 onwards (see paragraphs 25 and 35, respectively). Similarly, several programs target transformative change at scale; this may be covered in the evaluation report using applicable approaches (Uitto and Batra 2022). Additional themes and topics may additionally be covered to enhance the utility of the evaluation.

18. The evaluation report will draw from different information sources to facilitate triangulation. Evaluations that draw upon *desk reviews* will provide information on the documents reviewed, the rules for document inclusion and exclusion, and the instrument(s) used for review. For *interviews*, the evaluation report will provide information on interviewee names, positions, and affiliations; the interview date(s), and the interviewee selection process. For *field verifications*, information such as the basis for site selection, which sites were visited and when, and who visited the sites will be provided. Where *surveys* (including online surveys) are conducted, the report will discuss the sample frame, sampling approach, response rates, and period during which the survey was conducted.

19. The program's M&E system is expected to gather information on environmental stress reduction (greenhouse gas emissions reduction, reduction of waste discharge, etc.) and environmental status change (change in population of endangered species, forest stock, pollutants in the waterbody, water retention in degraded lands, etc.). The M&E system will also be an important source of information on program implementation, M&E plan implementation, and adaptive management. The evaluation should assess the quality of information gathered through the program M&E system at both the program and child project levels and—where the quality of information is adequate—use it.

Program Performance Additionality

20. Within the context of GEF support, *additionality* is understood as the extent to which GEF support has contributed to environmental and socioeconomic impacts; legal, institutional, and financial results; socioeconomic benefits; and innovation (GEF IEO 2020). To determine additionality, the report will compare the benefits of GEF support to a scenario without GEF support. It will identify specific areas where GEF support has contributed additional results and the nature of these additional results.

21. The evaluation will assess the extent to which benefits delivered using a programmatic approach are quantitatively and qualitatively different from a nonprogrammatic baseline approach—that is, are the benefits of a program different from a comparable cluster of stand-alone projects with similar objectives, inputs, and activities? A comparable GEF stand-alone investment may provide a basis to assess the benefits of the program modality. It will also assess the additionality of the use of an integrated approach over other program design choices that do not rely on integration. Depending on the context and objectives, an integrated approach or, alternatively, a simpler singular approach may be appropriate. The evaluation will weigh and discuss the extent to which use of an integrated approach was appropriate given the context, objectives, and costs involved.

Relevance and Coherence

22. The report will assess a program's *relevance* by considering the extent to which the program's objectives and design is responsive to the environmental and development priorities of the participating countries and congruent with GEF strategic priorities and objectives.

23. The report will discuss the *external coherence* of the program by considering its compatibility with other programs and activities being implemented in the participating countries, and in targeted sectors and institutions. It will discuss the extent to which the program involves socioeconomic trade-offs. It will also assess whether the program is well targeted (in terms of countries, regions, and/or communities covered) to generate global environmental benefits.

24. The report will discuss the *internal coherence* of the program. It will assess the extent to which program design tackled the drivers of environmental challenges it sought to address. It will assess the extent to which the theory of change, governance structure, activities, child projects, and M&E system are aligned with program objectives. It will also report on whether and how the program design incorporated lessons from past GEF activities and other relevant non-GEF interventions.

Integration

25. GEF programs approved from GEF-6 onwards use integrated approaches to address drivers of environmental degradation. Therefore, the evaluation will assess how the program

design applied integrated approaches, the extent to which integration was achieved, and the results.

26. An integrated program applies *systems thinking* by specifying the boundaries of the targeted system, addressing multiple drivers of environmental degradation simultaneously, addressing relationships among system elements across scales, addressing important risks and vulnerabilities, and considering system resilience (Bierbaum et al. 2018). Individual GEF programs may apply systems thinking in different ways. For example, the Sustainable Cities impact program uses the circular economy approach, and the Commodities integrated approach pilot is designed to provide support across a commodity's supply chain. An integrated program includes *feedback loops* to facilitate quick adaptation (Bierbaum et al. 2018). The specified results indicators cover the major pathways through which the different drivers of environmental degradation are mitigated. The indicators are tracked regularly and used to improve program effectiveness. A program should have applied these principles to generate global environmental benefits. The evaluation report will discuss the level of integration achieved, highlighting areas where integration was evident and effective, as well as those where it was not evident and/or may be improved.

Effectiveness

27. The evaluation report will discuss the extent to which program objectives and targets—as stated in the program framework documents—have been achieved. It will report if there were any changes in the program's design and/or expected results after start of implementation.

28. Consistent with Minimum Requirement 3 of the GEF Evaluation Policy 2019, the report will present an assessment of achievement of program results. The report will describe the relevant systems boundaries, the drivers of environmental degradation, and the system-level transformation outcomes targeted by the program. It will present an assessment of the extent to which progress has been made in achieving the targeted outcomes and long-term impacts of the program framework document. If the document provides targets for expected program outputs, their level of achievement will also be covered.

29. The report will aggregate data on expected and actual contributions of the program and its child projects to GEF corporate results targets (core indicators). The targets provided in the program framework document will provide a basis for comparison with actual results. For child projects that may still be under implementation, progress made up to the point of evaluation will be reported.

30. The evaluation report should provide an account of processes such as mainstreaming, replication, scaling up, and market change that indicate progress toward environmental and socio-economic changes at scales beyond that directly addressed by the program activities. The report should point out the specific elements of the program that are being adopted beyond GEF support. It should also discuss the extent to which experiences from one child project have influenced activities in others and, where applicable, the role played by the coordinating project in the process.

31. The report should discuss factors that affected delivery of outputs and outcomes, and progress to impacts, such as program design, implementation, linkages with other interventions, materialization of cofinancing, stakeholder involvement, and macroeconomic factors. If there are substantial differences in the performance of child projects, the report should discuss the factors that may explain the variance.

32. The evaluators should document the unintended impacts—both positive and negative—of the program and assess their overall scope and implications. Where these impacts are undesirable from an environmental or socioeconomic perspective, the evaluation should suggest corrective actions, along with preventive/mitigative measures that may be useful in future interventions.

Efficiency

33. The criterion of efficiency is aimed at assessing the extent to which the inputs are converted into results in an economic and timely way (GEF IEO 2019; OECD DAC 2021). The term “economic” is understood as the cost-effective conversion of inputs such as funds, expertise, personnel, and equipment into results compared to feasible alternatives. The term “timely” is understood as the extent to which the program activities were started and completed within a reasonable time frame. The report will discuss the extent to which the program was efficient in achieving its results. It will discuss areas of efficiency and inefficiency, and factors that affected efficiency.

Sustainability

34. Minimum Requirement 3 of the GEF Evaluation Policy 2019 specifies that a program evaluation will assess *sustainability*. The assessment will weigh risks to the continuation of net benefits from the program given their probability and severity. Specifically, it will discuss key risks and explain how they may affect the continuation of net benefits over a reasonable time frame. The report will also discuss the extent to which the flow of program benefits is resilient to shocks. This analysis should cover financial, sociopolitical, institutional, and environmental risks and shocks, with other likely risks discussed as applicable. The report should address the variation in levels of risk faced by the different program implementation sites and the reasons for these differences. It should discuss the measures that were undertaken, and/or may be needed, to mitigate risks.

Knowledge Management

35. Programs approved from GEF-6 onwards give considerable attention to knowledge management. The coordinating child projects of these programs are expected to facilitate knowledge exchange among other child projects, GEF Agencies, executing agencies, and other stakeholders. These programs are often expected to develop knowledge platforms that aid data capture, transformation, and sharing.

36. The evaluation will discuss the extent to which, and how, the knowledge platforms created by the program are contributing to the generation of global environmental benefits. It will discuss the effectiveness of the information technology system put in place to facilitate capture, storage, retrieval, and use of knowledge; as well as the types of knowledge products planned and delivered by the program and their utility. The report will highlight strong and weak areas of performance in knowledge management in supporting the achievement of program outcomes. It will also discuss good practices from the program that may be relevant for other GEF activities, along with areas where knowledge management in future programs and projects may be improved.

Program Implementation

37. The evaluation report will discuss how well the program was implemented by the GEF Agencies. It will describe the program's implementation experience, challenges that were faced at the program and project levels at different stages of the program cycle, how these challenges were addressed, and the circumstances that aided or hindered implementation. It will describe how information from program M&E was used to steer the program and its use of adaptive management. The report will also document instances where opportunities for adaptive management were not fully explored.

38. The report will discuss the performance of the lead GEF Agency and other partner Agencies, providing an account of how these Agencies have performed in their respective roles. In cases where the lead Agency is involved in child project implementation, its role in this capacity will be discussed along with those of other Agencies involved in child project implementation; its role as lead Agency will be assessed separately. The report will also discuss issues that may need follow-up from the lead Agency and other GEF Agencies.

Lead GEF Agency

39. The report will discuss how well the lead Agency led and coordinated the program. The operational units of the lead Agency are responsible for overall delivery of program outputs and outcomes. The key functions of a lead Agency may include—but are not restricted to—development of the program framework document, coordination among child projects, program-level monitoring and reporting, identification and management of program-level risks, application of GEF fiduciary standards (see paragraph 49) in implementation of the coordinating project (and child projects implemented by it), knowledge management across the program, midterm review, and program evaluation. The evaluation report will discuss how well the lead Agency performed its responsibilities, the challenges it faced, and the measures it took to respond to these challenges.

Other GEF Agencies

40. The report will discuss how well the Agencies involved in the implementation of child projects performed their responsibilities as relevant to overall program progress and performance. It will assess how well these Agencies collaborated with the lead GEF Agency and

among themselves. It will cover topics such as choice of executing agencies, timeliness of child project start-up and implementation, supervision during implementation, how well relevant risks were identified and managed for child projects, application of GEF fiduciary standards, implementation of M&E plans for child projects and collection of data on core indicators and/or tracking tools, and support for program midterm review and final evaluation. A tabular presentation that summarizes information by child project and responsible Agency may be useful.

Program Monitoring and Evaluation

41. Minimum Requirement 1 of the GEF Evaluation Policy 2019 calls for a fully developed and budgeted program M&E plan during program preparation; Minimum Requirement 2 calls for implementation of this plan during program implementation. The evaluation will assess strengths and weaknesses of the program M&E plan and its implementation. The focus will be on arrangements to track and report on program implementation and its results. Most coordinating projects include activities to coordinate program M&E. M&E-related discussion included in the terminal evaluation of the coordinating project will be a useful source of information on program M&E for the program evaluation.

M&E Design

42. The report will present an assessment of the program's M&E plan. The assessment will consider the extent to which the M&E plan was practical and well thought out. For example, did it clearly identify the M&E activities at the program level and how these will draw on the M&E activities to be implemented at the child project level? The report will also assess if the M&E plan for the program adequately addressed the program's theory of change, GEF M&E requirements, incorporated applicable core indicators and tracking tools, and provided baseline information. It will discuss whether the indicators specified to track environmental, gender, socioeconomic, and other results are appropriate (SMART: specific, measurable, attributable, realistic, and time-bound) and consistent across child projects.

M&E Implementation

43. The evaluation report will present an assessment of M&E implementation. It will discuss the extent to which the M&E system operated as per the plan; and if changes were made, it will note these changes, the reasons for them, and whether the changes were timely. The report will describe the extent to which information on the specified indicators was gathered systematically following a robust methodology, and in a timely manner. It will discuss how data from the child projects and coordinating project were aggregated and whether and how these data were used for decision making. It will discuss whether the resources allocated for M&E were sufficient and used prudently. Finally, it will discuss how the information from the M&E system was used in decision making and adaptive management.

Program Cofinancing

44. Cofinancing is defined as financing that is additional to GEF funding and that supports the implementation of a GEF-financed activity and the achievement of its objective(s) (GEF 2018). The use of cofinancing in GEF activities ensures that GEF finances only the incremental (or additional) costs of generating global environmental benefits. Cofinancing may also enhance the scale of the supported activities, support sustainability, and strengthen partnerships. Timely materialization of cofinancing is crucial to ensure that sufficient resources are available for planned activities and that these activities are completed on time.

45. The evaluation report will provide information on cofinancing sources, cofinancing commitments, and activities supported through cofinancing. It will report on the type of cofinancing mobilized—cash or in kind, grant, loan, or equity—managed by the GEF Agency implementing the program or in parallel by other partner organizations. It will provide information on the extent to which expected cofinancing materialized and was timely. Where materialization differed from commitments, or where there was a delay in materialization, the report will discuss the reasons for this and its effects on program implementation and results.

Alignment with GEF Policies

Environmental and Social Safeguards

46. The evaluation report will discuss the extent to which appropriate environmental and social safeguards were applied, especially in the design and implementation of child projects. It is expected that GEF activities will not cause any harm to the environment or to any stakeholder; where applicable, Agencies will take measures to prevent and/or mitigate any adverse effects.

Gender

47. The report will discuss the extent to which gender considerations were addressed in designing and implementing the program, including in its child projects. The report will look at whether a gender analysis was conducted by child projects, the extent to which the program was implemented in a manner that ensured gender-equitable participation and benefits, and whether gender-disaggregated data on beneficiaries was gathered. If the GEF program under evaluation or any of its activities disadvantaged or may disadvantage women, this should be documented and reported. The evaluators should also determine the extent to which relevant gender-related concerns were tracked through the program's M&E system, including for the individual child projects.

Stakeholder Engagement

48. The report will provide an account of stakeholder engagement in various program activities including its design and implementation. "Stakeholder" refers to an individual or group that has an interest in a GEF-financed activity or is likely to be affected by it, such as local communities, indigenous peoples, civil society organizations, and private sector entities (GEF 2017b). Stakeholder engagement is understood as a process involving stakeholder identification

and analysis; planning of stakeholder engagement; consultations with, and participation of, stakeholders in monitoring, evaluation, and learning through program implementation; addressing stakeholder grievances, and ongoing reporting to stakeholders (GEF 2017b). Considering the program objectives, the report will discuss the extent to which, and how, the program identified and engaged with the relevant stakeholders including the private sector, civil society organizations, government agencies, and program beneficiaries. The report will discuss the role stakeholder engagement has played in determining program ownership, building partnerships, and drawing on stakeholder expertise and capacities in program design and implementation. The report should discuss the extent to which, and how, stakeholder engagement has affected implementation, results, and risks to program sustainability.

Fiduciary Standards

49. The report will discuss the extent to which fiduciary standards were applied during program implementation. The GEF fiduciary standards are aimed at ensuring “accountability, integrity, transparency, efficiency and effectiveness in the use of GEF financing” by the GEF Agencies (GEF 2019b). The standards cover areas including, but not limited to, program appraisal standards, procurement processes, financial management and control, financial audits, financial disclosure, conflict of interest, code of ethics, whistle-blower protections, and safeguards against anti-money laundering and financing of terrorism. The evaluators will take stock of the extent to which program implementation was consistent with GEF fiduciary standards, relying on Agency responses to evaluator questions and information gathered through review of terminal evaluations, midterm reviews, and child project implementation reports. Where necessary, the program evaluation report will point out areas for improvement.

Lessons and Recommendations

50. Evaluators should provide a few well-formulated lessons that are based on program experience, including experiences from individual child projects. Wherever possible, evaluation reports should include examples of good practices in design and implementation that have led to effective stakeholder engagement, successful broader adoption of GEF initiatives by stakeholders, and large-scale environmental impacts. The evaluators should describe aspects of program performance that worked well, along with reasons for it. They should discuss the conditions under which these good practices may or may not be replicated.

51. Recommendations should be well formulated and targeted. They should clearly state the need for action, the recommended action along with its likely consequences vis-à-vis the status quo and other courses of action, the specific actor(s) to carry out the recommended action, and the time frame.

Performance Ratings

52. Program performance will be rated on criteria for outcomes, sustainability, implementation, and M&E (table 1). The ratings will be on an even-number Likert scale, as described in the annex. These ratings may be provided in a table in the main body of the report

or as an annex to the report. They may also be embedded in the corresponding report sections covering each topic.

Table 1: Ratings

Criterion	Dimension	Rating scale
Outcomes	Relevance, effectiveness, efficiency	Six points: highly satisfactory to highly unsatisfactory
Sustainability	Financial, sociopolitical, institutional, environmental	Four points: likely to unlikely
Implementation: lead GEF Agency	Performance of the Agency leading the program in its role as program lead	Six points: highly satisfactory to highly unsatisfactory
Implementation: other Agencies	Performance of Agencies implementing child projects of the program	Six points: highly satisfactory to highly unsatisfactory
M&E design	Quality of design of program M&E plan	Six points: highly satisfactory to highly unsatisfactory
M&E implementation	Quality of program M&E during implementation	Six points: highly satisfactory to highly unsatisfactory

C. Submission of Program Evaluation

53. The GEF Agencies will submit the final program evaluation reports on the GEF Portal no later than six months after completion of the program evaluation and no later than one year after completion of program activities. Where program evaluation has been conducted by its operations unit, the lead GEF Agency will also submit a validation report from its evaluation unit up to one year after completion of the program evaluation. If preparation of the program evaluation was delayed, the evaluation report will explain the reasons for the delay.

54. The GEF IEO will track and report on Agency submission of completed program evaluations. The IEO will also follow up with the respective Agencies on any errors noted in their submissions.

D. Use of Program Evaluation

55. The information provided by the program evaluations is used by the GEF IEO to report to the Council on overall GEF performance. It is also an input to the comprehensive evaluations prepared by the GEF IEO. The IEO synthesizes information presented in the program evaluations to assess performance of GEF investments through programs. It validates a sample of the program evaluation reports. It shares the evaluation reports publicly through its website.

56. The GEF Secretariat and GEF Agencies may also use program evaluation to inform their future work and for knowledge sharing. Similarly, the evaluations may be used by the GEF Scientific and Technical Advisory Panel to develop advisory products for the GEF partnership.

ANNEX: RATING SCALES

The evaluators will rate program performance on the following criteria: outcomes, sustainability, implementation, and M&E. The rated dimensions are described below along with the level of performance for a specific rating. The ratings will be provided for performance at the program level—that is, considering the performance of child projects as well as synergies and dis-synergies at the programmatic level, because a program is more than the sum of its parts.

Actual performance may not completely correspond to any of the rating descriptions that follow. Therefore, rating should be assigned based on the description that provides the best fit.

A. Outcome Ratings

The following criteria will be used to develop an overall rating of program outcome performance:

- (a) **Relevance and coherence.** Are the program outcomes aligned with the GEF focal area/operational program strategies, country priorities, needs of the beneficiaries, and mandates of the Agencies? Is the program compatible with other relevant programs and interventions being undertaken in the recipient countries covered by the program? Are the program outcomes consistent with the mandate of the GEF Agency? Was the program design appropriate for delivering the expected outcomes and consistent with the applicable GEF policies? Is/are the theory of change, program governance structure, activities, child projects, and M&E system aligned with the program objectives?
- (b) **Effectiveness.** To what extent were the outcomes commensurate with the ex ante expectations and targets provided in the program framework document? Has the program made the expected level of contributions to transformation of the targeted system? To what extent has it made progress in mitigating the targeted drivers of environmental degradation? To what extent have completed child projects achieved outcomes expected at implementation completion?
- (c) **Efficiency.** Was the program cost-effective? How does the program cost/time versus output/outcomes equation compare to other options to deliver similar results? Were the activities completed in a timely manner?

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

- **Highly satisfactory.** The outcomes substantially exceed expectations, and they are highly relevant and cost-effective.

- **Satisfactory.** The level of outcomes achieved was as expected. The outcomes are relevant and cost-effective.
- **Moderately satisfactory.** The level of outcomes achieved was close to expectation: the majority of targets met, but some were not achieved. The outcomes are generally relevant and cost-effective.
- **Moderately unsatisfactory.** Overall, the level of outcome achieved is lower than expected, although in some areas it was met or is close to expectations. The outcomes are generally relevant but somewhat cost-ineffective or, alternatively, generally cost-effective but not adequately relevant.
- **Unsatisfactory.** The expected outcome was not achieved or was substantially lower than expected, and/or the achieved outcome is not relevant. Alternatively, the outcome was cost-ineffective compared to alternatives.
- **Highly unsatisfactory.** None or a negligible level of outcome was achieved, or the program had substantial negative consequences that outweigh its benefits.
- **Unable to assess.** The available information does not allow an assessment of the level of outcome achievement.

B. Sustainability Ratings

The rating for program sustainability will be based on overall likelihood of incidence and magnitude of the effect of risks on continuation of net benefits from the program, along with the program's resilience to such risks. The rating will take into account the likelihood of the continuation of the benefit stream over a reasonable time frame. The evaluators will consider financial, sociopolitical, institutional, environmental, and other risks. At the time of the evaluation, these risks may not have materialized or may just be beginning to materialize. The assessment should be based on evidence that indicates the level of risk.

- **Financial resources.** What is the likelihood that financial resources will be available to continue the activities that result in benefits being sustained (e.g., income-generating activities, regular government budget allocations for the activities, and trends indicating likely future financial resources adequate for sustaining the program outcome)?
- **Sociopolitical.** Are there any social or political risks that can undermine the longevity of program outcomes? What is the risk that the level of stakeholder ownership is insufficient to allow for program outcomes/benefits to be sustained? Do key stakeholders see that it is in their interest that the program benefits continue? Is there sufficient public/stakeholder awareness in support of the long-term objectives of the program?
- **Institutional framework and governance.** Do the legal frameworks, policies, and governance structures and processes pose any threat to the continuation of program

benefits? While assessing these risks, consider if the required systems for accountability and transparency, and the required technical know-how, are in place.

- **Environmental.** Are there any environmental risks that can undermine the future flow of program benefits? The terminal evaluation should assess whether certain activities in the program area will pose a threat to the sustainability of program outcomes. For example, program outcome may be especially vulnerable to climate change risks. Similarly, biodiversity-related gains made by a program targeting marine protected areas may be affected by an increase in pollutant accumulation.

Program sustainability will use a four-point scale. The top two ratings (likely and moderately likely) comprise the likely range for sustainability, and the bottom two ratings the unlikely range.

- **Likely.** There are either no risks to continuation of benefits or there are some risks, but the magnitude of their effect is too small and/or the probability that they will materialize is too small. It is likely that the net benefits of the program will continue.
- **Moderately likely.** There are some risks to sustainability, and they may have some effect on the continuation of benefits if they materialize. However, the probability of materialization of these risks is low. Net benefits are more likely to continue than abate.
- **Moderately unlikely.** There are significant risks to sustainability. The effect on continuation of benefits will be substantial if these risks materialize, and the probability of materialization of these risks is significant. The net benefits of the program are likely to abate.
- **Unlikely.** There are severe risks to sustainability. These have a high probability of materialization and are likely to adversely affect continuation of benefits if they materialize. Overall, it is unlikely that net benefits will continue to accrue, and the long-term intended impacts of the program are unlikely to be achieved.
- **Unable to assess.** The available information does not allow an assessment of the expected incidence and magnitude of risks to sustainability.

C. Program Implementation Ratings

Program implementation ratings will be based on an assessment of the extent to which the GEF lead Agency and the GEF Agencies involved in implementation of child projects fulfilled their respective roles satisfactorily. This assessment excludes the performance of those agencies that execute activities on the ground under the supervision of GEF Agencies.

Two ratings will be provided for program implementation: for the lead GEF Agency of the program in its role as program lead, and, collectively, for the GEF Agencies involved in child project implementation. That is, a single rating is provided for implementation of *all* child projects; the rationale being that each child project will be rated in its respective terminal

evaluation. Table A.1 provides a description of the scale used to rate the performance of the lead Agency and other Agencies.

Table A.1: Program Implementation Rating Scale

Rating	Lead program Agency	Agencies implementing child projects
Highly satisfactory	Performance of lead implementing Agency was exemplary and exceeds expectations. Program framework document was robust, and program was well led. The Agency maintained strong oversight and ensured that relevant GEF policies were implemented. The Agency identified and addressed emerging concerns in a timely manner. Program M&E was robust, and its activities were implemented in a timely manner. The lead Agency ensured that program implementation was on track.	Performance of Agencies that implemented child projects was exemplary and exceeds expectations. Preparation of child projects was robust and well aligned with the program objectives. Child projects were well implemented. The Agencies collaborated well with the lead Agency and among themselves. They identified and addressed emerging concerns in the child projects in a timely manner. The M&E plans of child projects were robust, and their activities were implemented in a timely manner. The Agencies ensured that implementation of all or almost all child projects was completed on time or is on track to be completed on time.
Satisfactory	Performance of the lead Agency did not have any major weaknesses, and expectations were met. Program preparation and implementation were robust, and relevant GEF policies were applied. The Agency supervised the program well—it identified and addressed emerging concerns in a timely manner, although some weaknesses remained. The M&E plan was robust, and its activities were implemented in a timely manner. The lead Agency ensured that program implementation was on track.	Performance of the GEF Agencies implementing the child projects did not have any major weaknesses, and expectations were met. Preparation and implementation of child projects was robust, and relevant GEF policies were applied. The GEF Agencies supervised the child projects well—they identified and addressed emerging concerns in a timely manner, although some weaknesses may have remained. The M&E plans of the child projects were robust, and M&E activities were implemented in a timely manner. The GEF Agencies ensured that implementation of about two-thirds or more of the child projects was completed on time or is on track to be completed on time.
Moderately satisfactory	Performance of the lead Agency had some weaknesses, but overall expectations were met. Program	Performance of the GEF Agencies implementing child projects had some weaknesses, but overall expectations

Rating	Lead program Agency	Agencies implementing child projects
	<p>preparation and implementation were adequate and relevant. GEF policies were applied, although there were some areas of weak implementation. The lead Agency supervised the program adequately—it identified and addressed emerging concerns, although some remained unaddressed or were inadequately addressed. The M&E plan was solid, although some activities were dropped or were implemented with delays. Program implementation had minor delays, and a few activities were dropped.</p>	<p>were met. Preparation and implementation of child projects was adequate and relevant GEF policies were applied, although there are some areas of weak implementation. The GEF Agencies supervised implementation of the child projects adequately—they identified and addressed emerging concerns, although some concerns may have remained inadequately addressed. M&E plans of the child projects were solid. At least half of the projects were completed on time or are on track to be completed on time.</p>
Moderately unsatisfactory	<p>While there were several areas of solid performance, the lead Agency overall did not meet the expectations. Program preparation and implementation had weaknesses, although these were not too severe. Program supervision was somewhat weak. Although most emerging concerns were identified, many remained unaddressed or inadequately addressed. M&E implementation was somewhat weak. Program implementation was delayed, and a few activities were dropped or reduced in scale.</p>	<p>While there were several areas of solid performance, the GEF Agencies implementing the child projects overall did not meet the expectations. Preparation and implementation of child projects had weaknesses, although these were not too severe. Supervision of child project implementation was somewhat weak. Most emerging concerns were identified, but several remained unaddressed or inadequately addressed. M&E plan and/or implementation in child projects was somewhat weak. Implementation of at least half of the child projects experienced delays. Consequently, a few activities were dropped or reduced in scale.</p>
Unsatisfactory	<p>There were major shortcomings in implementation, and the lead Agency did not meet the expected level of performance. Program preparation and implementation had severe weaknesses. Adherence to fiduciary standards was somewhat weak. Program supervision was weak—emerging concerns were not identified in time</p>	<p>There were major shortcomings in implementation, and the GEF Agencies that implemented the child projects did not meet the expected level of performance. Child project preparation and implementation had severe weaknesses. Adherence to fiduciary standards was somewhat weak. Supervision of child projects was weak—emerging concerns were not identified in</p>

Rating	Lead program Agency	Agencies implementing child projects
	and remained unaddressed or inadequately addressed. M&E implementation was weak, and M&E activities were not implemented in a timely manner. Program implementation was delayed, and several activities were dropped or reduced in scale.	time and remained unaddressed or inadequately addressed. M&E implementation was weak, and M&E activities were not implemented in a timely manner. Implementation of at least two-thirds of the child projects was delayed, and several activities were dropped or reduced in scale.
Highly unsatisfactory	There were severe shortcomings in the quality of implementation. The lead Agency did not perform its fiduciary responsibilities satisfactorily. Supervision was poor. Emerging concerns were not identified in time, including those that should have been obvious. Although instances of mismanagement were discovered, corrective actions were not undertaken. Program activities were very poorly implemented, and several had to be dropped.	There were severe shortcomings in the quality of implementation of child projects. The Agencies did not discharge their fiduciary responsibilities satisfactorily. Supervision of implementation of a substantial number of child projects was poor. Emerging concerns were not identified in time, including those that should have been obvious. Although instances of mismanagement were discovered, corrective actions were not undertaken. All or almost all child projects were very poorly implemented, experienced delays, and had activities dropped.
Unable to assess	The available information does not allow assessment of the performance of the lead Agency.	The available information does not allow assessment of the performance of the Agencies that implemented child projects.

D. Program M&E Ratings

The M&E arrangements will be rated at the program level, with a focus on program-level M&E. The evaluation will cover the coordinating project’s M&E activities focused on program M&E. The evaluation will also cover the extent to which M&E arrangements in child projects were aligned with, and fed into, program M&E. The quality of program M&E will be assessed in terms of the following:

- **Design.** The assessment will consider the practicality and sufficiency of the M&E plan developed for the program. It will examine the extent to which the program and child project documents provide a baseline. It will consider the extent to which the program M&E plan specifies: (1) applicable core indicators and clear targets, including appropriate (SMART—specific, measurable, attributable, realistic, and time-bound) indicators to track environmental, gender, and socioeconomic results; (2) a proper

methodological approach for data collection; (3) logistics of the M&E activities, including a schedule and responsibilities for data collection; and (4) a budget with adequate funds for M&E activities.

- **Implementation.** The assessment will consider the extent to which the program’s M&E system operates as per the M&E plan. It will examine whether there was a need to revise the plan, and if so, it will consider if revisions were carried out in a timely manner. It will consider whether sufficient resources were provided for M&E. It will determine if the data on indicators—including core indicators, tracking tools, and other specified indicators—were gathered in a systematic manner. It will consider the extent to which appropriate methodological approaches were used, and the extent to which program management made use of the information from M&E during program implementation.

The quality of M&E on these two dimensions will be assessed separately on a six-point scale, as shown in table A.2.

Table A.2: Program Monitoring and Evaluation Rating Scale

Rating	M&E plan	M&E implementation
Highly satisfactory	The program M&E plan was robust and did not have any weaknesses. Baseline data were complete. The specified indicators were appropriate, and arrangements for M&E plan implementation were adequate. The M&E plan exceeds expectations and is exemplary.	M&E plan implementation was excellent. Weaknesses in M&E were addressed in a timely manner. M&E activities were conducted in a timely manner, and data from M&E were used to improve program implementation. Quality of M&E plan implementation exceeded expectations and was exemplary.
Satisfactory	The program M&E plan was robust and did not have any or had only minor weaknesses. Baseline data were provided, or its collection is planned at project start. The specified indicators were appropriate, and arrangements for M&E plan implementation were adequate. The plan meets expectations.	M&E plan implementation was generally as per the plan. Weaknesses in M&E were addressed in a timely manner. M&E activities were conducted in a timely manner, and data from M&E were used in improving program implementation. Overall, M&E implementation meets expectations.
Moderately satisfactory	On balance, the program M&E plan was solid. It did not have any major weaknesses, although it did have some minor weaknesses. The specified indicators were generally appropriate, and arrangements for M&E plan implementation were	M&E plan implementation was generally as per the plan. Weaknesses in M&E were generally addressed, although some weaknesses remained. Some M&E activities were delayed. M&E data were used for reporting but had little use in improving program implementation. Overall, M&E implementation meets

Rating	M&E plan	M&E implementation
	adequate. The plan meets expectations.	expectations with some areas of low performance.
Moderately unsatisfactory	On balance, the M&E plan was weak with strengths in some areas. The specified indicators were generally appropriate, but additional indicators were required to adequately capture program results. There were gaps in arrangements for M&E plan implementation. The plan needs several improvements to meet expectations.	On balance, M&E during implementation was weak with strengths in some areas. Most M&E activities were completed, but some were dropped or delayed. M&E data were generally not reported in a timely manner, and use of M&E data to improve program implementation was limited. Overall, M&E implementation does not meet expectations, although there are some areas where the performance is adequate.
Unsatisfactory	The M&E plan had severe shortcomings. No baseline data were, provided or any indication that baseline data would be collected at project start. Indicators do not adequately address project outcomes and other results; for several results, relevant indicators have not been specified. There are gaps in arrangements for M&E plan implementation—no budget or an inadequate budget was provided for M&E.	M&E plan implementation was flawed and/or did not address severe weaknesses of the M&E plan. Several M&E activities were either dropped or were incomplete. The data collection methodology was not sound. M&E data were not reported in a timely manner, and there is little evidence to suggest that the data were used to improve program implementation. M&E implementation does not meet expectations.
Highly unsatisfactory	No M&E plan was prepared.	No, or negligible, M&E activity was implemented other than conduct of the program evaluation.

BIBLIOGRAPHY

Bierbaum, R., et al. 2018. *Integration: to solve complex environmental problems*. Washington, DC: Global Environment Facility Scientific and Technical Advisory Panel.

GEF (Global Environment Facility). 2008. "Joint Summary of the Chairs—GEF Council Meeting. April 22–25, 2008." GEF/C.33/JointSummary.

—. 2011. "GEF Minimum Fiduciary Standards: Separation of Implementation and Execution Functions in GEF Partner Agencies." GEF/C.41/06/Rev.01.

—. 2012. "Clarifying the Responsibilities of the GEF's Key Actors with Respect to the Use of GEF Resources." GEF/C.42/04.

—. 2017a. "Policy on Gender Equality." GEF/C.53/04.

—. 2017b. "Policy on Stakeholder Engagement." GEF/C.53/05/Rev.01.

—. 2018. "[Project and Program Cycle Policy](#)." OP/PL/01.

—. 2019a. *Instrument for the Establishment of the Restructured Global Environment Facility*. Washington, DC: GEF.

—. 2019b. "Minimum Fiduciary Standards for GEF Partner Agencies." GA/PL/02.

—. 2019c. "Policy on Environmental and Social Safeguards." SD/PL/03.

—. 2020. "Guidelines on the Project and Program Cycle Policy." GEF/C.59/Inf.03.

GEF IEO (Global Environment Facility Independent Evaluation Office; previously the Evaluation Office). 2007. [GEF Evaluation Office Ethical Guidelines](#). Evaluation Document No. 2.

—. 2020. *An Evaluative Approach to Assessing GEF's Additionality*. Evaluation Report No. 139.

—. 2021. [Formative Evaluation of the GEF Integrated Approach to address the Drivers of Environmental Degradation](#).

—. 2022. *The GEF Evaluation Policy 2019*. Washington, DC: GEF IEO.

GEF STAP (Global Environment Facility Scientific and Technical Advisory Panel). 2021. "The Circular Economy and Climate Mitigation—A STAP Advisory Document."

OECD DAC (Organisation for Economic Co-operation and Development Development Assistance Committee). 2002. *Glossary of key terms in evaluation and results-based management*. Working Party on Aid Evaluation.

—. 2021. [*Applying Evaluation Criteria Thoughtfully*](#). Paris: OECD Publishing.

Uitto, Juha I., and Geeta Batra. 2022. *Transformational Change for People and the Planet: Evaluating Environment and Development*. Springer, Cham.

UNEG (United Nations Evaluation Group). 2016. *Norms and Standards for Evaluation*. New York: UNEG.

—. 2020. *Ethical Guidelines for Evaluation*. New York: UNEG.