

Evaluation of the Agency Self Evaluation Systems

(Prepared by the Independent Evaluation Office of the GEF)

- Revised Approach Paper -
July 2020

Point of contact:

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

Contents

Background	2
Literature review.....	4
Assessments Across GEF Partnership	7
Purpose	9
Theory of Change	10
Key Questions and Hypothesis.....	14
Evaluation Design.....	16
Sources of Information	17
Risks and Limitations.....	19
Peer feedback and Stakeholder Involvement.....	20
Expected Outputs, Outreach and Tracking	20
Resources and Schedule	21
Evaluation Team	21
Schedule of Work Activities	21
References	23

Background

The Agency self-evaluation systems are expected to facilitate learning and accountability across the GEF partnership. At the project level, these are reflective exercises that help implementors to learn from the experience. At a higher level in an organization, self-evaluations facilitate in monitoring of the portfolio and may provide early warnings on trends that warrant attention. Factors such as policy framework, quality assurance arrangements, incentives for candor in reporting, harmonization of practices, information sharing arrangements, and adequacy of resources for self-evaluation provided by an Agency may affect the extent to which its self-evaluation system meet the needs of the GEF partnership.

The GEF Independent Evaluation Office (IEO) is undertaking the ‘Evaluation of the Agency Self-Evaluation Systems’ to assess the extent to which Agency self-evaluation systems provide quality and timely information and comply with the GEF requirements. This draft approach paper presents a discussion on the literature, scope of the enquiry, key questions, and proposed methods for the evaluation.

OECD DAC defines independent evaluation as an “*evaluation carried out by entities and persons free of the control of those responsible for the design and implementation of the development intervention*” (OECD 2002). In contrast, it defines self-evaluation as “*an evaluation by those who are entrusted with the design and delivery of a development intervention*” (OECD 2002). The definitions used across the United Nations system and across the multi-lateral development banks consider self-evaluation to be a management project/program management responsibility and primarily for the management’s own use (UN, 2018; ECG, 2018). Both independent evaluation and self-evaluation are aimed at enhancing learning and accountability. Although, independent evaluation is generally regarded as more credible for accountability.

The term self-evaluation system includes components such as scope, responsibilities, methods, data, and products, that cover different stages of the project and program cycle. There are considerable variations across the GEF Agencies in terms of their mandates, scale of operation, and structure of the relationship between their evaluation function and management. As a result, the self-evaluations needs, and practices, of the Agencies vary.

The expectations from the self-evaluation systems of the Agencies are outlined in several GEF policy documents and policies of the GEF Agencies. For example, The GEF Evaluation Policy (GEF IEO 2019), The Guidelines for GEF Agencies in Conducting Terminal Evaluation for Full-sized Projects (GEF IEO 2017), and Minimum Fiduciary Standards for GEF Partner Agencies (GEF 2018) specify several requirements.

The GEF Evaluation Policy (2019) requires the Agencies to prepare mid-term reviews (where applicable), terminal evaluations, and monitor their respective GEF portfolios. The terminal evaluation guidelines for full-sized projects provides guidance on reporting on topics such as project details, stakeholder consultations, theory of change, methodology, outcome, sustainability, implementation, project M&E, safeguards, gender, and lessons. While the guidelines are applicable to the full-size projects that involve more than US \$ 2.0 million in GEF funding, the Agencies are encouraged to use these to develop terms of reference for evaluation of the GEF funded medium-size projects and programs. The GEF Monitoring Policy (2019) addresses the guiding principles for monitoring along with other requirements including reporting through project implementation reports and tracking tools.

The evaluation policies of several GEF Agencies address self-evaluations (EBRD 2013, IFAD 2015, UNDP 2016, UNIDO 2018, IDB 2019). These policies generally cover the relationship between self and independent evaluation functions, responsibilities related to self-evaluation, and reporting requirements.

Networks and groups such as The OECD Development Assistance Committee (DAC), ECG of the international development banks, United Nations Evaluation Group (UNEG), and Multilateral Organisation Performance Assessment Network (MOPAN), promote – among other things – coherence and harmonization in M&E across multilateral organizations.

Nonetheless, there are considerable variations in self-evaluation practices of the GEF Agencies.

The GEF IEO is undertaking this evaluation in response to requests from the GEF Council members and the GEF Secretariat to assess the extent to which Agency self-evaluation systems provide quality and timely information and comply with the GEF requirements. The GEF Council and the Secretariat are interested in ensuring that the Agencies monitor their GEF portfolios well and that Agency self-evaluation practices facilitate learning and accountability for GEF supported activities. Another motivation is that GEF Agencies that have a strong independent evaluation function along with a culture of self-evaluation, may provide lessons to other Agencies where these systems are less well developed.

For this evaluation, self-evaluation will include any evaluation conducted by those that are involved in implementation of evaluated activities with or without involvement of the Agency evaluation unit. This includes project level evaluations such as terminal evaluations, mid-term reviews, and progress reports prepared by those involved in project and program implementation. The evaluation offices of some of the GEF Agencies such as UNEP and UNIDO are involved in commissioning and managing the conduct of terminal evaluations and/or other reviews. In such instances, for the purposes of this evaluation, these terminal evaluations will also be considered as self-evaluations.

There may be a difference in how a GEF Agency addresses self-evaluation of the GEF supported activities vis-à-vis non-GEF activities. Some Agencies may have different arrangements for the evaluation of GEF supported activities, whereas others may have identical arrangements and procedures for the evaluation of GEF and non-GEF activities. This evaluation will focus on how the Agency self-evaluation systems address the GEF supported activities. However, for comparison, the evaluation would consider the arrangements for self-evaluation of the non-GEF activities.

Literature review

Two streams of scholarly literature – knowledge management and monitoring and evaluation (M&E) in international development organizations – are relevant to evaluation of the Agency self-evaluation systems. Although much of the work on knowledge

management is based on experiences in business organizations, some of these experiences are relevant to the international development context. The literature on M&E in international development organizations – although important for this evaluation – is relatively less developed.

Self-evaluation systems

Several practitioners have discussed establishment of self-evaluation systems in international development organizations (Zall Kusek and Rist 2004; Bester 2012). Picciotto (1999) distinguishes self-evaluation from independent evaluation by noting that the former aims primarily at assisting decision makers whereas the latter focuses primarily on accountability.

Self-evaluations serve many purposes for which independent evaluation may not be as well suited. Self-evaluations are useful for communicating implementation progress and impact of an intervention to the decision makers, donors, and general public (Zall Kusek and Rist, 2004). They are useful in situations where decisions are urgent and require close synchronization (Picciotto, 1999). Self-evaluation also provides practitioners opportunities for conversion of tacit knowledge into explicit knowledge (Spender 1996; Nonaka 1994). Taut's (2007) 'action researcher' takes advantage of the rich information gained by being an insider and generates knowledge that facilitates adaptive management. This approach also allows for rapid feedback to others on lessons that may be applicable in other contexts with similar challenges (Taut 2007).

Mayne (2020) argues for greater attention to development of an evaluative culture in community focused organizations. He makes a case that to have a strong evaluative culture an organization need to have capabilities for self-evaluation, opportunities to conduct these evaluations, and be motivated to use evidence from self-evaluation to improve their services to clients.

One of the major challenges with self-evaluation is the issue of credibility. Scriven (1975) argues that a self-evaluation has less credibility because of the perceived conflict of interest. This may be understood within the framework of the agent-principal problem as agents may lack incentives for candor (Ross 1973; Arrow 1984; and Grossman and Hart,

1992). However, Scriven (1975) notes that measures such as issuance of guidelines on conducting self-evaluations, use of checklists, and standardization of evaluation criteria and practices may enhance credibility. While in theory self-evaluation should promote learning, lack of incentives to do so well may compromise its utility. There is a risk that self-evaluation may become a bureaucratic requirement for those responsible for conducting them and may result in mechanical tracking of indicators without attention to their broader implication (WB IEG 2016).

Scriven (1975) argues that independent evaluation of at least some of the activities along with self-evaluations will enhance the credibility of the latter. Picciotto (2012) argues that independent evaluation should focus on the higher-level questions that are not adequately assessed by self-evaluation, and the rest should be left to the latter. He also notes that self-evaluations are more likely to be owned and implemented by decision makers since they are self-generated. Picciotto (2002) argues that regardless of the type of evaluation, they add value only if they result in lessons and institutional learning.

Knowledge Management

There is agreement among scholars that knowledge is a critical resource for organizations (Drucker 1993; Quinn 1992; Reich 1992). Effective organizations create knowledge and integrate it in their work (Lam, 2000; Spender 1996a; Grant 1996; Tsoukas 1996). They facilitate knowledge transfer among their staff (Szulanski, 1996), and may gain a competitive advantage from it (Arrow 1974; Kogut and Zander, 1992). As a result, it is important to know whether and how organizations create, acquire, and manage knowledge.

Several factors influence effectiveness of knowledge transfer. Individual effort and motivation, and strength of ties among individuals, facilitate knowledge transfer along with an individual's ability to frame and translate knowledge (Reagans and McEvily, 2003). It may be more efficient to use 'strong ties' to transfer tacit knowledge and 'weak ties' to transfer codified knowledge (Reagans and McEvily, 2003). Szulanski et al. (2004) found that perceived trustworthiness of the source aids effectiveness of intra-organizational knowledge transfer.

Following Mintzberg's (1979) typology, most international development organizations may be classified as professional bureaucracies, marked by the presence of a complex but stable work environment, and where coordination is achieved by design and by application of standards. Lam (2000) argues that the learning focus of a professional bureaucracy tends to be narrow and constrained within the boundary of formal specialist knowledge. Lam analyzes knowledge within an organization along two dimensions – epistemological and the ontological. She uses modes of expression of knowledge – explicit and tacit knowledge – and locus of knowledge – individual and collective – in a matrix form to describe four different forms of organizational knowledge: 'embrained' (individual-explicit), 'embodied' (individual-implicit), 'encoded' (collective-explicit), and 'embedded' (collective-tacit) knowledge. She concludes professional bureaucracies have a higher dependence on 'embrained' knowledge than other types of organizations.

International development organizations provide development aid primarily through a project-based modality. According to Ajmal and Koskinen (2008), project team members need to learn things that are already known in other contexts and need to acquire and assimilate knowledge that resides in organizational memory. Ability of team members to learn determines their individual effectiveness and eventually organizational effectiveness (Huber, 1991). Documentation and sharing of experiences from completed projects may help a project-based organization avoid repetition of past errors (Ajmal and Koskinen, 2008).

Assessments Across GEF Partnership

Several GEF Agencies already assess some aspects of performance of their self-evaluation systems. Usually these analyses focus on the quality of products of self-evaluation. For example, evaluation units of several GEF Agencies such as UNDP, UNEP, and IFAD assess, and report, aggregated quality ratings of self-evaluations of their completed projects through their annual reports. These analyses of self-evaluation reports may address issues such as candor and consistency of reporting, and lessons and recommendations, but usually these take the form of quality ratings and generally these topics are not dealt with in detail. Systemic and process related issues relevant to self-evaluation may be addressed occasionally through detailed reviews.

'Behind the Mirror: A Report on the Self-Evaluation Systems of the World Bank Group' (IEG 2016) is the first detailed review of a GEF Agency self-evaluation system. The evaluation reviewed the self-evaluation practices of the World Bank Group comprehensively and compared these to the practices of the other multilateral development banks. The evaluation found that self-evaluation systems of the World Bank Group complement the independent evaluation systems, but the focus of the former is more on results reporting and meeting the accountability needs and less on learning to enhance performance. The evaluation found that the information generated by the self-evaluation systems is not regularly mined for knowledge and learning except by the Independent Evaluation Group of the World Bank. The evaluation called for greater use of the information generated by the self-evaluation systems to meet the learning needs of the management and teams, and for improvement of portfolio performance.

Several other GEF Agencies have undertaken or are presently conducting evaluations to assess performance of their self-evaluation systems. The Independent Evaluation Department (IED) of the ADB recently completed a review of ADB's project level self-evaluation system (IED ADB 2020). The evaluation assessed the ADB's project level self-evaluation system to be robust and credible, and to be useful for both learning and accountability. However, the evaluation also found reliability of gathered information to be a concern and concluded that much of the tacit knowledge that is internal to the ADB staff remains untapped. Similarly, the Independent Development Evaluation (IDEV) unit of the AfDB has conducted an evaluation of the self-evaluation systems and processes of AfDB recently (IDEV 2019). Several other GEF Agencies such as EBRD and UNDP¹ are presently undertaking an assessment of their self-evaluation systems (EBRD 2019).

An assessment undertaken by the Joint Inspection Unit (JIU) of United Nations compares evaluation function across UN organizations based on structure and reporting lines, size, budget, and utility (JIU, 2014). However, it does not clearly distinguish between self and independent evaluation function or compare self-evaluation system performance. In assessments undertaken by the United Nations Evaluation Group (UNEG), the focus on

¹ Communications with UNDP Independent Evaluation Office

independent evaluation function and coverage of self-evaluation systems is nominal. The ECG has prepared a practice note focused on normative expectations from a self-evaluation system but has not assessed and compared the performance of the self-evaluation systems of its members. Assessments carried out by the Multilateral Organisation Performance Assessment Network (MOPAN) touch upon some of the issues that are relevant to self-evaluation, such as adaptive management, results focus, and evaluation policy. However, the assessments do not consider self-evaluation as a specific area of performance.

The GEF IEO has covered some aspects of the self-evaluation system performance through its Annual Performance Report (APR). APR regularly presents analysis of quality of terminal evaluations prepared by the GEF Agencies and gaps in submission of these evaluations. These terminal evaluations are generally prepared or commissioned by the project management at implementation completion, and these evaluations give account of the project or program implementation experience and performance. Occasionally, APRs have also covered other issues relevant to self-evaluation. For example, APR 2005 (GEF IEO 2006) included an assessment of the project-at-risk systems of the GEF Partner Agencies, which also covered arrangements for monitoring and reporting of project and portfolio performance. APR 2006 (GEF IEO 2007) included an assessment of project supervision practices of World Bank, UNDP and UNEP, and – among other topics – covered quality of reporting through the annual project implementation reports. APR 2015 covered gaps in submission of tracking tools by the GEF Agencies. More recently, APR 2020 presented a detailed analysis of the quality of terminal evaluation reports submitted by the GEF Agencies including an assessment of the extent to which these reports comply with the GEF IEO terminal evaluation guidelines. However, GEF IEO is yet to conduct a comprehensive assessment of the self-evaluation systems of GEF Agencies.

Purpose

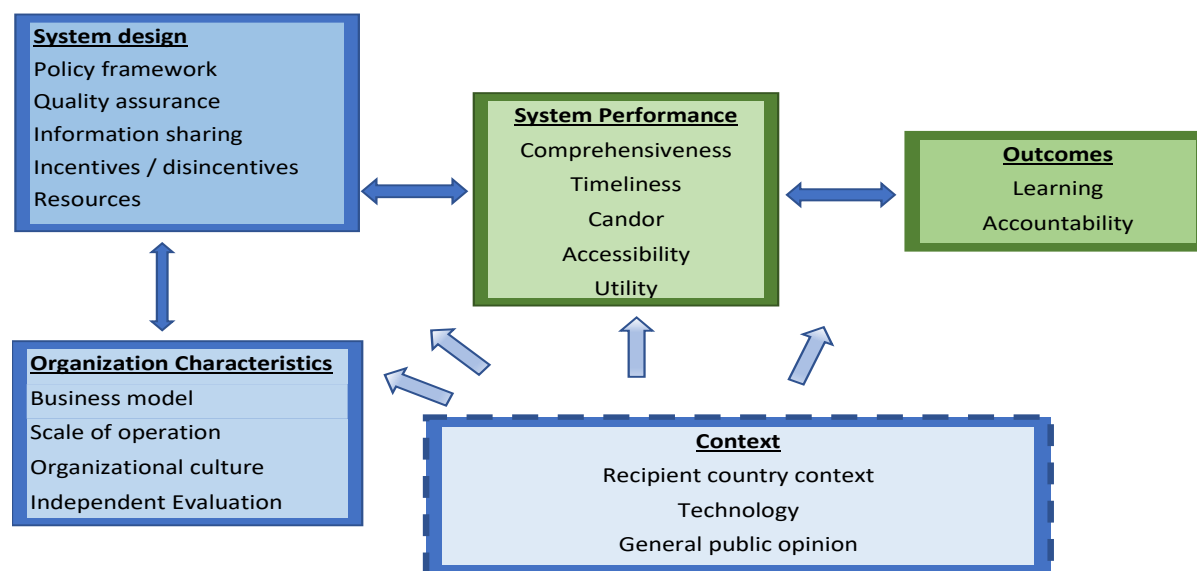
The GEF Independent Evaluation Office (IEO) is undertaking the ‘Evaluation of the Agency Self-Evaluation Systems’ to assess the extent to which the Agency self-evaluation systems provide credible and quality information on GEF supported activities. It will assess how these self-evaluation systems are implemented, how information on GEF activities is generated, and how this information is used in decision making. The evaluation will assess

the factors that affects quality of reporting on GEF supported activities and identify areas for improvement.

Theory of Change

The evaluation will be based on a theory of change presented in Figure 1. The socio-political and technological context in which an agency operates may affect the characteristics of the given agency, the type of self-evaluation system it puts in place, and how this system performs. The embedded assumption being that the agency characteristics and system design will be in sync with each other, and with the contextual conditions in which these operate. How well a self-evaluation system is designed may be expected to affect system performance. The underlying assumption being that the system will be well implemented and will function as designed. A well performing self-evaluation system may be expected to provide information that is comprehensive, timely, candid, accessible and useful. However, even if a self-evaluation system provides such information (intermediate outcome), it may not lead to learning and accountability if the information is not used in decision making. Thus, the assumptions need to hold to ensure learning and accountability (final outcomes).

Figure 1. A simple theory of change for a self-evaluation system



The focus of this evaluation will be on the concerns related to system design, system performance, and outcomes. Other factors related to operational context and organizational characteristics will be examined to understand how these are affecting system performance.

Within the context of GEF Partnership the theory of change model presented in figure 1 may be applied both at the corporate level – i.e. at the GEF management level – and at the Agency level. The former is beyond the scope of this evaluation as it is being addressed in a separate review of the GEF RBM system and of the GEF IEO terminal evaluation validation process. This review will focus on the evaluation of the self-evaluation system at the GEF Agency level.

System Effectiveness (The Ys)

The system effectiveness related dependent variables are classified as intermediate outcomes and final outcomes (Figure 1). Final outcomes of an effective self-evaluation system are enhanced learning and accountability in the organization. Learning is reflected in actions taken based on knowledge generated by the self-evaluation systems. These actions may be at the project level, thematic level, regional and/or at the corporate level. At the project level this reflects in terms of actions taken for adaptive management of project, design improvements in follow up activities, or incorporation of lessons in design of similar projects. At the corporate level it is likely to show in terms of improvements in the policies, guidelines, and business processes.

Greater accountability implies that the organization not only set targets and milestones for indicators of institutional performance but tracks actual performance using credible methods and owns responsibility for target achievement. An effective self-evaluation system facilitates accountability by gathering information on various indicators, assuring quality of collected data and data gathering processes; and, making this information accessible to decision makers and other users within the organization, and, where applicable, to general public. When targets and milestones are not met, an Agency clearly communicates non-achievement and, where applicable, facilitates corrective actions.

A self-evaluation system may generate information on how the project was implemented and the results of the project. This includes information on the extent to which and how various policies and safeguards were applied, the challenges that were faced during implementation and their effect, and how these were addressed. It also generates information on the use of inputs and its results in terms of outputs and outcomes. It also generates information on the risks that may pose threat to sustainability.

While reports and datasets generated by a self-evaluation system may be regarded as the key system outputs, the extent to which knowledge contained in these products is comprehensive, timely, candid, accessible, and useful, may be regarded as intermediate outcome. If knowledge produced by the self-evaluation system scores high on these dimensions, then it may be expected to facilitate learning and accountability. However, whether it does so need to be examined.

If the information generated by the self-evaluation system is relevant and covers important areas of institutional performance, it may be regarded as comprehensive. An effective self-evaluation system will track what is important and track it well without overburdening the organizations involved in data gathering – i.e. the GEF Agency and their counterparts in the recipient countries that execute the GEF activities on ground. Comprehensive coverage of issues that are of concern may be expected to facilitate learning and accountability.

An effective self-evaluation system would provide decision makers with timely information on emerging risks and challenges. Timely availability of information to decision makers and other users will facilitate its uptake for corrective actions and would facilitate learning and accountability. When information from the self-evaluation system is easily accessible, users will be able to access and use the information with ease. Accessibility will include data format, explanation, and retrieval.

Candor in reporting is an important dimension of a self-evaluation system's effectiveness: the higher the level of candor, the greater the trust in the information reported through terminal evaluations, mid-term reviews and project implementation reports, and by extension greater the ability of these products to facilitate learning. While candor in reporting is useful to ensure accountability, exclusive reliance on any single information

source that is involved in design and implementation of the evaluated activity may create disincentives for candor. The self-evaluation systems that incentivize candor, along with mitigation of risks associated with its use for accountability, are more likely to be effective.

Utility of the information generated by the self-evaluation system is a key intermediate outcome. If the generated information is useful for decision making and to deepen the understanding of relevant issues, it will help an organization incorporate this knowledge in its work and improve. Evidence of utility may be found in use of the generated information in strategic decisions at the corporate level, adaptive management at the project level, and for reporting through corporate performance scorecards and/or performance reviews. It will also be useful in designing new activities and policies.

Factors that affect the system performance (the Xs')

Outcomes and performance of a self-evaluation system may be affected by several factors such those related to system design and implementation, organizational characteristics, capacities of those partners that execute activities on ground, and broader context.

System design related variables include the self-evaluation policy framework; presence of a functioning centralized self-evaluation function and arrangements for quality assurance; information management arrangements; incentives to promote candor; and sufficiency of resources allocated for self-evaluation. The extent to which system design is in sync with other structures and systems will also have a bearing on the system effectiveness.

Presence of a centralized self-evaluation unit and adequate quality assurance arrangements are expected to raise the quality of the self-evaluation through follow up, feedback, and by addressing information gaps. In due course, quality assurance arrangements are expected to build evaluation capacities and promote candor.

A robust information management system is likely to enhance effectiveness of the self-evaluation system. It would provide for systematic recording of data, quality assurance, and access to data in a form that is easy to use. It would also be a repository of data that may be analyzed to draw lessons from design and implementation of activities and policies.

How incentives for self-evaluation are designed, may affect the level of candor in reporting. In organizations where timely reporting of risks and concerns helps a manager in garnering greater support for implementing corrective actions for a given activity or policy, the likelihood of candor in reporting increases. However, if the reporting of risks and concerns is taken to be an indicator of poor performance, then the responsible manager is less likely to report it or report it in an opaque manner.

Provision of adequate resources for a self-evaluation system is important. Lack of (staff) time and budget affects implementation of the system and is likely to affect the quality of evidence gathered. Under resourced systems are unlikely to ensure that the generated information is of high quality, timely, and accessible.

Among the organizational characteristics, variables such as business model, scale of operation, organizational culture, and relationship with independent evaluation affect the design and performance of the self-evaluation system. Scale of operation may affect the extent to which self-evaluation systems need to be elaborate. The international organizations that work at scale and have a multi-country footprint need to have more systematic self-evaluation systems because barriers to knowledge sharing among staff are higher due to geographical distance and weaker ties. At the same time, economies of scale allow them to have more elaborate arrangements for self-evaluation. Organizational culture is an important influence on the self-evaluation traditions. Staff diversity, leadership, type of business model, and arrangements that provide individual staff agency, may influence organizational culture.

A mutually reinforcing relationship with the independent evaluation system may enhance effectiveness of the self-evaluation system. An independent evaluation unit may build capacities for self-evaluation by providing guidance and training, and by providing feedback on the quality of self-evaluation. Self-evaluation, on the other hand, may be a source of quality data for independent evaluations. We may expect the self-evaluation system to benefit from a well-functioning independent evaluation system.

Key Questions and Hypothesis

The evaluation aims to answer the following questions:

How do policy frameworks in the GEF Agencies support their self-evaluation

systems? The evaluation will assess how self-evaluation is addressed by various policies of Agencies. The assumption is that an enabling policy framework will lead to sound arrangements for self-evaluation, which will then lead to good quality self-evaluations. The evaluation will assess the extent to which policies explain the purpose and role of self-evaluations, provide guidance on how the self-evaluations ought to be conducted, and clarify relationship with independent evaluation.

To what extent do the agency self-evaluation systems provide credible, quality and timely information to support accountability and learning?

The evaluation will consider the arrangements that are in place in the Agencies to conduct self-evaluations (or equivalent²). It will consider arrangements to address quality assurance, harmonization, information management and knowledge sharing, as they relate to self-evaluation of the GEF projects. Where the Agencies have separate arrangements for evaluating GEF projects than for other projects in their portfolio, the evaluation will document the differences in the arrangements and the rationale for the differences. The evaluation will assess the extent to which these arrangements allow for robust self-evaluation of GEF supported activities.

To what extent do the Agency self-evaluation systems meet the GEF requirements according to the relevant GEF policies and guidelines?

The evaluation will assess the extent to which GEF Agencies are meeting the minimum requirements relevant to self-evaluation. These include the minimum requirements related to project M&E stipulated by the GEF Evaluation Policy (2019) and its predecessor the GEF M&E Policy (2010), and the Guidelines for GEF Agencies in Conducting Terminal Evaluation for Full-sized Projects (2017). It will assess the quality of the products such as project implementation reports, mid-term reviews, and terminal evaluations, that are being produced by the self-evaluation systems (or equivalent arrangements). From a normative perspective, the evaluation will record perceptions of the Agency staff, national counterparts, and consultants, on the

² The term equivalent is used for Agencies where the evaluation unit may be involved in conducting evaluations such as terminal evaluation or mid-term reviews that are generally conducted by those implementation of the activities.

extent to which the Agency self-evaluation systems are effective in supporting the learning needs of the GEF partnership.

What are the factors that influence the effectiveness of the self-evaluation systems which could impact the quality and timeliness of information provided to the GEF?

The evaluation will assess whether and how different variables such as policy framework, information management arrangements, incentives to promote candor, quality assurance arrangements, level of resources provided, and capacities of the local partners, affect self-evaluation system effectiveness. It will assess if, and how, presence of a robust independent evaluation function affects a self-evaluation system's effectiveness.

Evaluation Design

The evaluation will use relevance, effectiveness, efficiency and impact criteria to assess Agency self-evaluation system performance as it related to GEF supported activities (OECD, UNEG). These criteria are embedded in the evaluation questions discussed earlier.

Questions related to policies, evaluation arrangements, and compliance with GEF requirements address issues related to relevance. Questions related to compliance with the GEF requirements and the extent to which the self-evaluation systems support the learning needs of the Partnership address the effectiveness criterion. Similarly, examination of resource use and its sufficiency address the efficiency criterion. Questions that assess effect on learning and accountability, address the impact criterion.

The evaluation will be based on the principles of design thinking (Ramaprasad and Syn, 2013; Adams, and Nash, 2016). The evaluation process will focus on a participatory approach through collaboration and learning, in addition to accountability. It will diagnose weaknesses in the self-evaluation systems and help identify desirable solutions through an active engagement process

The evaluation will use a multiple-case design and cover all the GEF Agencies (Yin 2018). Self-evaluation system of a GEF Agency – as it related to the GEF supported activities – will be the unit of analysis. The self-evaluation system of a given GEF Agency will be assessed based on its specific circumstances and how it addresses evaluation of the GEF supported activities. Within each Agency information on the self-evaluation system will be gathered at

three levels: corporate, middle-management, and project management level. At the corporate level, the unit responsible for RBM and independent evaluation, along with senior management will be an important source of information.

For each of the Agencies, a few projects will be selected to gather information on the extent to which the self-evaluation system is operating at the project level as designed. The number of projects covered will differ based on whether a given Agency adopts the same process for all projects, or adopts different processes depending on the type of GEF project. In case an Agency follows a uniform process for GEF projects, two projects – one recently closed and one under implementation – would suffice. In other cases, more projects may be reviewed. The projects will be covered to illustrate how the self-evaluation system design works in ideal conditions – it is not aimed at ensuring representativeness. Ensuring representativeness requires a different magnitude of effort, which is beyond the scope of this evaluation.

Sources of Information

Literature Review: the evaluation will draw from the literature relevant to self-evaluation systems especially on topics such as knowledge management and M&E in international development organizations. Some of this work has already been incorporated in this proposal. The work will be further deepened through systematic identification of the relevant literature, synthesis, and incorporation of its findings in the report based on the research. Another important area will be the emerging literature on design thinking-based evaluation, which emphasizes attention to needs and experience of the users. This literature will be a reference for consultative process for conduct of the evaluation.

Desk Reviews: the source material from GEF Agencies will be reviewed. This will include Agency policies related to evaluation, monitoring, results-based management (RBM), and activity cycle; performance score cards; templates for appraisal of project proposals, regular reporting on projects, tracking progress on corporate results indicators; and, annual portfolio monitoring reports and thematic reviews conducted by the operations. Review of evaluation, monitoring, and RBM policies, will help in understanding the policy framework for self-evaluation within each of the selected organizations. Templates and

related guidance used by organizations for regular reporting on projects will be reviewed to determine what is being collected, why, how, and at what frequency, and for what use. Review of a sample of annual project implementation reports, mid-term reviews, and implementation completion reports, along with relevant guidance will facilitate a comparison of the information being gathered through these tools and quality of information provided. Reports prepared by UNEG, ECG, MOPAN, and JIU, that cover at least some aspects of self-evaluation in GEF Agencies will also be reviewed.

Datasets: The evaluation will draw on different datasets maintained by the GEF IEO. This includes data on project performance and quality of reporting. It will also draw from the PIR datasets maintained by the GEF Secretariat.

Interviews: Interview of different sets of respondents will be an important source of information. GEF Secretariat staff involved in coordination of the self-evaluations at the GEF corporate level will be interviewed.

Several categories of respondents from the GEF Agencies will be interviewed. Staff involved in design and implementation of the self-evaluation system in Agencies will be an important source for information on how the system is supposed to work, how it is working at the corporate level, and what arrangements are there for GEF supported activities. They will provide details on the information management system design, submission of self-evaluation reports, quality assurance arrangements, and conduct of targeted analysis and synthesis of information from the self-evaluation system. They would also be a useful source of information on the policy framework for self-evaluation and relationship with the independent evaluation function. The staff of the evaluation units will be another source for information on functioning of the self-evaluation system and its relationship with the independent evaluation function. The senior and mid-level managers of the organization will be tapped for information on expectations from self-evaluation and actual use of the information generated by it. The staff and consultants involved in implementation and self-evaluation of the projects will be an important source for documenting the working of the self-evaluation system at the project level. About 12-15 interviews per selected organization may be sufficient. However, the eventual number will

depend on whether each additional interview continues to bring in new information and helps deepen the understanding of the self-evaluation system.

National counterparts will be interviewed to gather their perceptions on the performance of the Agency self-evaluation systems. Depending on the interviewee one or more GEF Agencies may be covered through a single interview.

Different modules will be developed to gather information from the different sets of interviewees. Some of the information gathered through desk reviews will be validated through interviews.

Online survey: An online survey will be conducted to gather perception on credibility and use of information provided by the self-evaluation system of the organization. Targeted respondents include staff of the selected organizations and their partners in recipient countries. Other stakeholders such as the GEF Secretariat staff, GEF Operational Focal Points, and CSO network members, will also be covered. The effort required from the respondents will not exceed 15 minutes including time required to read the questions and background information. The list of potential respondents will be acquired from the selected organizations. All of the selected organizations maintain these lists for dissemination of their knowledge products and sharing of official publications.

Workshops and focus groups: In addition to interviews, workshops and focus groups will be an important source of information. These avenues will also be used for validation and analysis of information gathered through other sources. The number of workshops and focus groups will be determined based on the need but at the minimum a start-up workshop will be conducted to share the evaluation approach with the Agencies and final workshop to share the emerging findings. In between focus groups and workshops would be conducted to gather and analyze information on Agency self-evaluation systems.

Risks and Limitations

The evaluation covers all the 18 GEF Agencies. Given the number of Agencies, it will be difficult to accomplish the evaluation without active involvement and cooperation of the

Agencies and their staff. Despite their support, it may still be difficult to execute all the planned activities of the evaluation given the level of complexity in the required coordination.

Although online surveys are economical, the response rates to these surveys are low. Experience at the GEF IEO shows that response rate may be around 10 percent – although these may be doubled through follow up. We still make this choice because online survey is a complementary source of information. The information may be used to identify issues that are of concern and need to be explored further through interviews and focus groups.

Peer feedback and Stakeholder Involvement

The evaluation will benefit from the feedback from two peer reviewers: Garrett Kilroy, Senior Evaluation Specialist at Independent Evaluation Office of ADB, and Sonja Sabita Teelucksingh, Senior Environmental Specialist at the GEF Secretariat. The peer reviewers have provided feedback on the draft version of this approach paper. They will also provide feedback on the intermediary products and the draft report of the evaluation.

The first and the last workshops are planned as information sharing events – the first to share information on the evaluation approach and the final workshop to share the findings and identify remedial measures. Other workshops and focus groups in the interim will be a source of information and will facilitate analysis of the gathered information. The participants will include Agency staff from operations and evaluation units, and GEF Secretariat. Other stakeholders that will be consulted include the STAP and the CSO Network.

The draft report of the evaluation will be shared with the key stakeholders to get their feedback on the emerging conclusions, and to identify errors of analysis and of omission and commission.

Expected Outputs, Outreach and Tracking

The evaluation is primarily intended for the GEF Council and the GEF corporate audience, including the GEF Secretariat, the GEF Partner Agencies, STAP, and the CSO Network. The evaluation report will be delivered during the FY2021. The evaluation report will be

published on the GEF IEO website and distributed via email among the GEF Council members, GEF country focal points, GEF Secretariat, Partner Agencies, and the CSO network. A four-page summary of the findings will also be prepared for circulation among a wider audience.

Resources and Schedule

Evaluation Team

The evaluation will be led by Neeraj Kumar Negi, Senior Evaluation Officer at the GEF IEO. Molly Sohn, Evaluation Analyst, will be the other member of the core team of the evaluation. The review of self-evaluation products such terminal evaluation, project implementation report, and mid-term review, will be conducted by the GEF IEO staff and consultants. They would also conduct the desk review of the Agency policies and guidelines relevant to self-evaluation.

The GEF IEO will hire a firm to conduct the collaborative process based on systems design thinking to gather and analyze information on the Agency self-evaluation systems and to identify remedial actions. The firm will use workshops, focus groups, interviews, and surveys, to gather information. It will also draw on the related work conducted by the GEF IEO. The firm will submit a report to the evaluation leader from the GEF IEO. The findings of this report will be integrated in the GEF IEO report that will be presented to the GEF Council.

Schedule of Work Activities

The GEF IEO report will be delivered in February 2021, in time for the first meeting of the replenishment group in March 2021 and for June 2021 GEF Council meeting. Table 1 shows the schedule of work activities for completion and presentation of the findings of the evaluation. The schedule of work has been prepared keeping in mind the GEF Council meeting schedule.

Table 1. Schedule of work activities

Project milestone	Work period or completion date
Approach paper	July 20 th , 2020

First workshop	August 30 th , 2020
Desk reviews, interviews, online survey, interim workshops and focus groups	September to December 2020
Analysis/Synthesis of gathered data	December 2020
Final workshop	January 2021
Final report by the firm	20 th of January 2021
Draft evaluation report	February 7 th 2021
Report for the replenishment meeting	February 20 th 2021
Preparation of the four-page flier	March 2021
Council document of the evaluation uploaded	May 1 st , 2021
Presentation of the evaluation to the Council	June 2021
Publication of the finalized report	June 2021

References

- Adams, Chithra, and John B. Nash. "Exploring design thinking practices in evaluation." *Journal of MultiDisciplinary Evaluation* 12, no. 26 (2016): 12-17.
- Ajmal, Mian M., and Kaj U. Koskinen. "Knowledge transfer in project-based organizations: an organizational culture perspective." *Project Management Journal* 39, no. 1 (2008): 7-15.
- Arrow, Kenneth. *The economics of agency*. No. TR-451. Stanford University California Institute for Mathematical Studies in the Social Sciences, 1984.
- Arrow, Kenneth. *The limits of organization*. WW Norton & Company, 1974.
- Audia, Pino G., Sebastien Brion, and Henrich R. Greve. "Self-assessment, self-enhancement, and the choice of comparison organizations for evaluating organizational performance." In *Cognition and strategy*, pp. 89-118. Emerald Group Publishing Limited, 2015.
- Bester, Angela. "Results-based management in the United Nations Development System: Progress and challenges." A report prepared for the United Nations Department of Economic and Social Affairs, for the Quadrennial Comprehensive Policy Review (2012).
- Chelimsky, Eleanor, ed. *Program evaluation: Patterns and directions*. American Society for Public Administration, 1985.
- Coase, Ronald H. "The problem of social cost." In *Classic papers in natural resource economics*, pp. 87-137. Palgrave Macmillan, London, 1960.
- Drucker, Peter. "Post-capitalist society." 1993.
- ECG. "ECG Practice Note: Self-evaluation in ECG member institutions." 2018. Available at: <https://www.ecgnet.org/documents/46856/download>
- European Bank for Reconstruction and Development (EBRD). 2013. *Evaluation Policy*. Available at: <https://www.ebrd.com/cs/Satellite?c=Content&cid=1395241631988&pagename=EBRD%2FContent%2FDownloadDocument>

---2019. EVD Work Programme 2019 to 2020 and Budget 2019. Available at:
<https://www.ebrd.com/documents/evaluation/evd-work-programme-201920-and-budget-2019.pdf>

Global Environment Facility (GEF). 2018. GEF Corporate Scorecard. Available at:
<https://www.thegef.org/sites/default/files/documents/ScorecardMay2018.pdf>

---2018. Minimum Fiduciary Standards for GEF Partner Agencies. Available at:
https://www.thegef.org/sites/default/files/documents/Fiduciary_Standards.pdf

---2019. Policy on Monitoring. Available at:
https://www.thegef.org/sites/default/files/council-meeting-documents/EN_GEF.C.56.03.Rev_.01_Policy_on_Monitoring.pdf

Global Environment Facility Independent Evaluation Office (GEF IEO). 2006. GEF Annual Performance Report (APR) 2005. Available at:
<http://www.gefio.org/sites/default/files/ieo/evaluations/apr-2005.pdf>

---2007. GEF Annual Performance Report 2007. Available at:
<http://www.gefio.org/sites/default/files/ieo/evaluations/apr-2007.pdf>

---2017. GEF Annual Performance Report 2015. Available at:
<http://www.gefio.org/sites/default/files/ieo/evaluations/files/apr%202015.pdf>

---2017. The Guidelines for GEF Agencies in Conducting Terminal Evaluation for Full-sized Projects. Available at:
<https://www.gefio.org/sites/default/files/ieo/evaluations/files/gef-guidelines-te-fsp-2017.pdf>

---2019. The GEF Evaluation Policy. Available at:
<http://www.gefio.org/sites/default/files/ieo/evaluations/files/gef-me-policy-2019.pdf>

---2020. GEF Annual Performance Report 2020. Available at:
<https://www.gefio.org/sites/default/files/ieo/evaluations/files/apr-2020.pdf>

Grossman, Sanford J., and Oliver D. Hart. "An analysis of the principal-agent problem." In *Foundations of Insurance Economics*, pp. 302-340. Springer, Dordrecht, 1992.

Huber, George P. "Organizational learning: The contributing processes and the literatures." *Organization science* 2, no. 1 (1991): 88-115.

IDEV. Evaluation of the African Development Bank's self-evaluation systems and processes. Independent Development Evaluation unit of African Development Bank, 2019. Available at:

<https://idev.afdb.org/sites/default/files/documents/files/SESP%20Evaluation%20Summary%20Report%20%28EN%29.pdf>

IED ADB. 2020 - Annual Evaluation Review: ADB's Project Level Self-Evaluation System. Independent Evaluation Department of Asian Development Bank. 2020. Available at:

<https://www.adb.org/sites/default/files/evaluation-document/539791/files/aer2020.pdf>

IEG. Behind the Mirror: A Report on the Self-Evaluation Systems of the World Bank Group. World Bank Independent Evaluation Group. 2016. Available at:

https://ieg.worldbankgroup.org/sites/default/files/Data/Evaluation/files/behindthemirror_0716.pdf

Inter-American Development Bank Office of Evaluation and Oversight (IDB OVE). 2019. Evaluation Policy Framework – IDB Group. Available at:

<http://idbdocs.iadb.org/wsdocs/getdocument.aspx?docnum=EZSHARE-872199154-11142>

International Fund for Agricultural Development (IFAD). 2015. Revised IFAD Evaluation Policy. Available at: <https://webapps.ifad.org/members/eb/102/docs/EB-2011-102-R-7-Rev-3.pdf>

Joint Inspection Unit. 2014. Analysis of the Evaluation Function in the United Nations system. Joint Inspection Unit of United Nations, JIU/REP/2014/6. Available at:

https://www.unjiu.org/sites/www.unjiu.org/files/jiu_document_files/products/en/report-s-notes/JIU%20Products/JIU_REP_2014_6_English.pdf

King, Gary, Robert O. Keohane, and Sidney Verba. *Designing social inquiry: Scientific inference in qualitative research*. Princeton university press, 1994.

Kogut, Bruce, and Udo Zander. "Knowledge of the firm, combinative capabilities, and the replication of technology." *Organization science* 3, no. 3 (1992): 383-397.

Kruger, Justin, and David Dunning. "Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments." *Journal of personality and social psychology* 77, no. 6 (1999): 1121.

Lam, Alice. "Tacit knowledge, organizational learning and societal institutions: An integrated framework." *Organization studies* 21, no. 3 (2000): 487-513.

Liverani, Andrea, and Hans E. Lundgren. "Evaluation systems in development aid agencies: an analysis of DAC Peer Reviews 1996—2004." *Evaluation* 13, no. 2 (2007): 241-256.

Mayne, John. "Building evaluative culture in community services: Caring for evidence." *Evaluation and program planning* 80 (2020): 101450.

Miller, Dale T., and Michael Ross. "Self-serving biases in the attribution of causality: Fact or fiction?." *Psychological bulletin* 82, no. 2 (1975): 213.

Mintzberg, Henry. *The structuring of Organizations*. Prentice Hall. 1979.

MOPAN. 2017a. "African Development Bank (AfDB): Institutional Assessment Report." MOPAN 2015-16 Assessments, 2017.

---2017b. "Inter-American Development Bank (IDB): Institutional Assessment Report." MOPAN 2015-16 Assessments, 2017.

---2017c. "The World Bank: Institutional Assessment Report." MOPAN 2015-16 Assessments, 2017.

---2017d. "United Nations Development Programme (UNDP): Institutional Assessment Report." MOPAN 2015-16 Assessments, 2017.

---2019a. "Asian Development Bank (ADB): Institutional Assessment Report." MOPAN 2017-18 Assessments, 2019.

---2019b. "Food and Agriculture Organization (FAO): Institutional Assessment Report." MOPAN 2017-18 Assessments, 2019.

---2019c. "International Fund for Agricultural Development (IFAD): Institutional Assessment Report." MOPAN 2017-18 Assessments, 2019.

---2019d. "World Health Organization (WHO): Institutional Assessment Report." MOPAN 2017-18 Assessments, 2019.

Nonaka, Ikujiro. "A dynamic theory of organizational knowledge creation." *Organization science* 5, no. 1 (1994): 14-37.

The Organization for Economic Co-operation and Development (OECD). 2002. Glossary of Key Terms in Evaluation and Results Based Management. Available at:

<http://www.oecd.org/development/peer-reviews/2754804.pdf>

--- "Press Release: Development Aid Drops in 2018, Especially to Neediest Countries." 10 April 2019. (2019). Available at: <https://www.oecd.org/newsroom/development-aid-drops-in-2018-especially-to-neediest-countries.htm>

--- Better Criteria for Better Evaluation: Revised Evaluation Criteria Definitions and Principles for Use. OECD/DAC Network on Development Evaluation. Available at:

<http://www.oecd.org/dac/evaluation/revised-evaluation-criteria-dec-2019.pdf>

Picciotto, Robert. "The logic of evaluation independence and its relevance to international financial institutions." *Independent Evaluation* (2012): 37.

Picciotto, Robert. "The logic of mainstreaming: a development evaluation perspective." *Evaluation* 8, no. 3 (2002): 322-339.

Picciotto, Robert. "Towards an economics of evaluation." *Evaluation* 5, no. 1 (1999): 7-22.

Quinn, James Brian. *Intelligent Enterprise: A Knowledge and Service Based Paradigm for Industr.* Simon and Schuster, 1992.

Ramaprasad, Arka, and Thant Syn. "Design thinking and evaluation using an ontology." In *European Design Science Symposium*, pp. 63-74. Springer, Cham, 2013.

Reagans, Ray, and Bill McEvily. "Network structure and knowledge transfer: The effects of cohesion and range." *Administrative science quarterly* 48, no. 2 (2003): 240-267.

Reich, Robert B. "The work of nations. 1992." NY: Vintage. (1992).

Ross, Stephen A. "The economic theory of agency: The principal's problem." *The American economic review* 63, no. 2 (1973): 134-139.

Scriven, Michael. *Evaluation bias and its control*. Kalamazoo, MI: Evaluation Center, Western Michigan University, 1975.

Scriven, Michael. *Evaluation thesaurus*. Sage, 1991.

Segone, Marco. "Bridging the gap. The role of monitoring and evaluation in evidence-based policy making." (2008).

Spender, J-C. "Making knowledge the basis of a dynamic theory of the firm." *Strategic management journal* 17, no. S2 (1996): 45-62.

Szulanski, Gabriel, Rossella Cappetta, and Robert J. Jensen. "When and how trustworthiness matters: Knowledge transfer and the moderating effect of causal ambiguity." *Organization science* 15, no. 5 (2004): 600-613.

Szulanski, Gabriel. "Exploring internal stickiness: Impediments to the transfer of best practice within the firm." *Strategic management journal* 17, no. S2 (1996): 27-43.

Tamer Cavusgil, S., Roger J. Calantone, and Yushan Zhao. "Tacit knowledge transfer and firm innovation capability." *Journal of business & industrial marketing* 18, no. 1 (2003): 6-21.

Taut, Sandy. "Studying self-evaluation capacity building in a large international development organization." *American Journal of Evaluation* 28, no. 1 (2007): 45-59.

The World Bank Independent Evaluation Group (WB IEG). 2016. *Behind the Mirror: A Report on the Self-Evaluation Systems of the World Bank Group*. Available at: <http://documents.worldbank.org/curated/en/902331469736885125/pdf/107274-WP-REVISED-PUBLIC.pdf>

United Nations. 2018. Regulations and Rules Governing Programme Planning, the Programme Aspects of the Budget, the Monitoring of Implementation and the Methods of Evaluation. Secretary-General's bulletin. ST/SBG/2018/3. 1 June 2018.

United Nations Development Programme (UNDP). 2016. UNDP Evaluation Policy. Available at:

http://web.undp.org/evaluation/documents/policy/2016/Evaluation_policy_EN_2016.pdf

United Nations Evaluation Group (UNEG). 2016. Norms and Standards for Evaluation. Available at: <http://www.uneval.org/document/detail/1914>

United Nations Industrial Development Organization (UNIDO) 2018. UNIDO Evaluation Policy. Available at: https://www.unido.org/sites/default/files/files/2018-06/Evaluation_Policy_DGB-2018-08.pdf

Yin, Robert K. Case study research and applications: Design and methods. Sage publications, 2017.

Zall Kusek, Jody, and Ray Rist. Ten steps to a results-based monitoring and evaluation system: a handbook for development practitioners. The World Bank, 2004.