

OPS4

PROGRESS TOWARD IMPACT



FOURTH OVERALL PERFORMANCE STUDY OF THE GEF

APPROACH TO TERMINAL EVALUATION REVIEW

METHODOLOGICAL PAPER #9

Methodology for assessment of performance of completed GEF projects

Review of terminal evaluation reports submitted to the Evaluation Office forms the basis of the assessment on performance of completed GEF projects. OPS3 has already reported on projects for which terminal evaluations had been submitted up till FY 2004. From FY 2005 to FY 2008, 210 more terminal evaluations – of which 116 are for full size projects and 94 for medium size projects – were submitted. The submitted terminal evaluations were reviewed by the Evaluation Office using a two step approach: desk verification of the all the terminal evaluation reports¹ and field verification of a randomly selected sample of the completed full size projects.

The Evaluation Office reviewed the submitted terminal evaluation reports to verify the terminal evaluation performance ratings and to assess the overall quality of the terminal evaluation report. The focus of the terminal evaluation reviews was on verification of the following ratings: Outcomes; Sustainability; Project M&E; Quality of Implementation and Execution; and, Quality of Terminal Evaluation Report. The reports were reviewed by a team comprised of Evaluation Office staff and consultants using a set of detailed guidelines to ensure that uniform criteria are applied. When a reviewer found the ratings presented in the terminal evaluation to be inconsistent with the available evidence, the reviewer proposed to upgrade or downgrade project ratings in the terminal evaluation report. In such instances, the draft reviews were then examined by a peer reviewer who provided feedback on the review and this was incorporated by the primary reviewer in the subsequent versions of the review. The review reports were then shared with the respective implementing agencies for comments and finalized after taking their comments into account. On average this exercise takes about two to three person days of work.

The Evaluation Office conducts direct verification of the terminal evaluation reports to verify the validity of the evidence presented in them and to assess at the systemic level the extent to which the verified ratings given through the terminal evaluation review process are reflective of the actual performance indicated during the direct field verification. Such verifications usually involve field visit of 5 to 10 days accompanied with review of available secondary literature on the project. The parameters on which ratings that are verified through this process are the same as those verified through the desk review of the terminal evaluation reports. For OPS4 the field verifications have been carried out for the terminal evaluations for 14 GEF projects.

Criteria for Outcome Ratings

Based on the information provided in the terminal evaluation report, an assessment was made of the extent to which the project's major relevant objectives were achieved or are expected to be achieved,

¹ These include full size projects, medium size projects, and those enabling activities in which GEF has invested more than 0.5 million dollars.

relevance of the project results, and the project's cost-effectiveness². The ratings on the outcomes of the project are based on performance on the following criteria³:

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

An overall rating was provided according to the achievement and shortcomings in the three criteria ranging from highly satisfactory, satisfactory, moderately satisfactory, moderately unsatisfactory, unsatisfactory, highly unsatisfactory, and unable to assess. A rating was provided for each of the three criteria (relevance, effectiveness, and efficiency). Relevance of outcomes was rated on a binary scale: 'satisfactory' or 'unsatisfactory'. If an 'unsatisfactory' rating was provided on this criterion, the overall outcome achievement rating was not higher than "unsatisfactory". Effectiveness and Efficiency was rated as following:

- Highly satisfactory: The project had no shortcomings.
- Satisfactory: The project had minor shortcomings.
- Moderately satisfactory: The project had moderate shortcomings.
- Moderately unsatisfactory: The project had significant shortcomings.
- Unsatisfactory: The project had major shortcomings.
- Highly unsatisfactory: The project had severe shortcomings.
- Unable to assess: The reviewer was unable to assess outcomes on this dimension.

The calculation of the overall outcomes score of projects considered all three criteria, of which relevance criterion was applied first - the overall outcome achievement rating may not be higher than "unsatisfactory" if the relevance rating is unsatisfactory. The second constraint that was applied is that the overall outcome achievement rating may not be higher than the "effectiveness" rating. The third constraint that was applied was that the overall rating may not be higher than the average score of effectiveness and efficiency criteria calculated using the following formula:

$$\text{Outcomes} = (b + c) \div 2$$

² *Objectives* are the intended physical, financial, institutional, social, environmental, or other development results to which a project or program is expected to contribute (OECD DAC 2002).

³ *Outcomes* are the likely or achieved short-term and medium-term effects of an intervention's outputs. Outputs are the products, capital goods, and services that result from a development intervention; these may also include changes resulting from the intervention that are relevant to the achievement of outcomes (OECD DAC 2002). For the GEF, environmental outcomes are the main focus.

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

Criteria for Sustainability Ratings

Sustainability is understood as the likelihood of continuation of project benefits after completion of project implementation (GEF 2000). To assess sustainability, the key risks that could undermine continuation of benefits at the time of the evaluation were identified and assessed. To rate the likelihood of sustainability of project outcomes following risks were assessed:

- a. Financial resources. What is the likelihood that financial resources will be available to continue the activities that result in the continuation of benefits (income-generating activities, and trends that may indicate that it is likely that in future there will be adequate financial resources for sustaining project outcomes)?
- b. Sociopolitical. Are there any social or political risks that can undermine the longevity of project outcomes? What is the risk that the level of stakeholder ownership is insufficient to allow for project outcomes/benefits to be sustained? Do the various key stakeholders see in their interest that the project benefits continue to flow? Is there sufficient public/stakeholder awareness in support of the long-term objectives of the project?
- c. Institutional framework and governance. Do the legal frameworks, policies, and governance structures and processes pose any threat to the continuation of project benefits? While assessing this parameter, consider if the required systems for accountability and transparency, and the required technical know-how, are in place.
- d. Environmental. Are there any environmental risks that can undermine the future flow of project environmental benefits? The terminal evaluation should assess whether certain activities in the project area will pose a threat to the sustainability of project outcomes. For example, construction of dam in a protected area could inundate a sizable area and thereby neutralize the biodiversity-related gains made by the project.

A rating was provided for each of the four criteria using the following scale:

- Likely. There are no risks affecting that criterion of sustainability.
- Moderately likely. There are moderate risks that affect that criterion of sustainability.
- Moderately unlikely. There are significant risks that affect that criterion of sustainability.
- Unlikely. There are severe risks affecting that criterion of sustainability.
- Unable to assess. Unable to assess risk on this dimension.
- Not applicable. This dimension is not applicable to the project.

A rating of unable to assess was used when a reviewer is unable to assess sustainability on any given risk category. If ratings were not possible for two or more categories, then the overall rating on sustainability was not provided. All the risk categories of sustainability were considered to be critical. Therefore, the overall rating could not be higher than the lowest rating risk category/categories.

Criteria for Assessment of Quality of Project M&E Systems

GEF projects are required to develop M&E plans by the time of work program inclusion, to appropriately budget M&E plans, and to fully carry out the M&E plan during implementation. Project managers are

also expected to use the information generated by the M&E system during project implementation to improve and adapt the project to changing situations. Given the long-term nature of GEF projects, project proponents are also encouraged to include long-term monitoring plans that measure results (such as environmental results) after project completion. Terminal evaluation reviews include an assessment on the achievement and shortcomings of M&E systems of the assessed project.

M&E design. Project should have a sound M&E plan to monitor results and track progress in achieving project objectives. An M&E plan should include a baseline (including data, methodology, and so on), SMART (specific, measurable, achievable, realistic, and timely) indicators and data analysis systems, and evaluation studies at specific times to assess results. The time frame for various M&E activities and standards for outputs should have been specified. Questions to guide this assessment include: In retrospect, was the M&E plan at entry practicable and sufficient (sufficient and practical indicators identified; timely baseline; targets created; effective use of data collection; analysis systems including studies and reports; practical organization and logistics in terms of what, who, and when for M&E activities)?

M&E plan implementation. The M&E system was in place and allowed the timely tracking of results and progress toward project objectives throughout the project. Annual project reports were complete, accurate, and with well-justified ratings. The information provided by the M&E system was used to improve and adapt project performance. An M&E system should be in place with proper training for parties responsible for M&E activities to ensure that data will continue to be collected and used after project closure. Question to guide this assessment include: Did the project M&E system operate throughout the project? How was M&E information used during the project? Did it allow for tracking of progress toward project objectives? Did the project provide proper training for parties responsible for M&E activities to ensure data will continue to be collected and used after project closure?

A rating is provided for the M&E design and M&E plan implementation criteria using the following scale:

- Highly satisfactory. There were no shortcomings in that criterion of the project M&E system.
- Satisfactory. There were minor shortcomings in that criterion of the project M&E system.
- Moderately satisfactory. There were moderate shortcomings in that criterion of the project M&E system.
- Moderately unsatisfactory. There were significant shortcomings in that criterion of the project M&E system.
- Unsatisfactory. There were major shortcomings in that criterion of the project M&E system.
- Highly unsatisfactory. There was no project M&E system.

The rating for M&E during implementation constitutes the overall rating of the M&E system whereas the rating on M&E design is used as an explanatory variable in the analysis.

Criteria for Assessment of Quality of Terminal Evaluation Reports

The ratings on quality of terminal evaluation reports are assessed using the following criteria:

- a. The report presents an assessment of all relevant outcomes and achievement of project objectives in the context of the focal area program indicators if applicable.

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

Each criterion to assess the quality of the terminal evaluation is rated on the following scale:

- Highly satisfactory. There were no shortcomings in the terminal evaluation on this criterion.
- Satisfactory. There were minor shortcomings in the terminal evaluation on this criterion.
- Moderately satisfactory. There were moderate shortcomings in the terminal evaluation on this criterion.
- Moderately unsatisfactory. There were significant shortcomings in the terminal evaluation on this criterion.
- Unsatisfactory. There were major shortcomings in the terminal evaluation on this criterion.
- Highly unsatisfactory. There were severe shortcomings in the terminal evaluation on this criterion.

The first two criteria (of all relevant outcomes and achievement of project objectives and report consistency and substantiation of claims with proper evidence) are more important and have therefore been assigned a greater weight. The quality of the terminal evaluation reports is calculated using the following formula:

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

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

Data Limitations

The Evaluation Office uses statistical tests to assess differences among groups of projects, and the findings reported here are significant at the 90 percent confidence level. Regression analysis was used to assess the magnitude and direction of change associated with different variables. Over the past four fiscal years (FY2005 to FY2008), there has been an improvement in the overall quality of information provided in terminal evaluation reports. However, information on financial issues, including materialization of cofinancing, and on M&E related issues remains below expectations. When sufficient information on a performance parameter for a project has not been provided in its terminal evaluation report, that project has not been included in the portfolio-level assessment for that parameter.

Assessing the performance of completed projects in terms of their outcomes, sustainability of outcomes, and implementation of M&E reflects actions that are now long past, limits the extent to which information gathered from analysis of these data are useful in making real-time corrections in

operations. Notwithstanding this limitation, this assessment provides a long-term perspective on the extent to which GEF projects are performing vis-à-vis expectations.

Some of the limitations are related to the use of results based framework on which outcome achievements of a project are assessed. Outcome achievements are generally assessed through comparison of actual achievements with the commitments made at inception. While this allows an assessment of the extent a project meets the ex ante expectations, it does not facilitate a direct inter-project and inter-period comparison because the deliverables promised for projects that are otherwise comparable may be different.

wb269966

M:\M&E\EO Staff\Neeraj\OPS4\Methodological notes\Approach to Terminal Evaluation Review.docx

9/15/2009 5:07:00 PM

