



Global Environment Facility

GEF/ME/C.35/Inf. 5
May 28, 2009

GEF Council

June 22-24, 2009

GEF EVALUATION OFFICE: ANNUAL PERFORMANCE REPORT 2008

1. Background, Main Conclusions, and Recommendations

1.1 Background

1. This document is the fifth annual performance report (APR) that the Global Environment Facility (GEF) Evaluation Office has developed. The report presents a detailed account of some aspects of project results, of processes that may affect these results, and of monitoring and evaluation (M&E) arrangements in completed GEF projects. This APR also contains an assessment of the M&E arrangements in GEF projects at entry. This is a follow up to the first assessment presented in *APR 2005*. For the second time, a performance matrix is presented, which summarizes the performance of the GEF Agencies and the GEF Secretariat on various parameters tracked by the Office. This matrix will be a regular feature of future APRs. This year, the Evaluation Office also continued direct verification of terminal evaluations.

2. The *APR 2008* has been prepared to provide an input to the *Fourth Overall Performance Study (OPS4)* of the GEF. Therefore, much of the discussion focuses on the projects whose terminal evaluations were submitted after the *Third Overall Performance Study (OPS3)* period – submitted after FY 2004.

3. The *APR 2008* continues the annual presentation of assessment of project outcomes, project sustainability, project completion delays, materialization of cofinancing, and quality of monitoring. For the **assessment of project outcomes, project sustainability, and delays in project completion**, 210 projects, for which terminal evaluation reports were submitted by the GEF Agencies to the Evaluation Office since FY 2005, have been considered. Of these, 62 projects were submitted during FY 2008. The GEF investment in these 210 projects totaled \$989 million.

4. For **assessment of materialization of cofinancing**, all 285 projects for which terminal evaluation reports have been submitted since FY 2002 were considered. Of these, information on materialization of cofinancing was provided for 210 projects (74 percent), either in the terminal evaluation reports and/or by the relevant Implementing Agency (IA). The GEF has invested a total of \$987 million in these 210 projects; the IAs reported that an additional \$2.96 billion was leveraged in the form of cofinancing. These projects include 162 projects pertaining to OPS4, for which information on materialization of cofinancing is available. In these 162 projects GEF has invested \$ 705 million and an additional \$ 1.43 billion was leveraged in form of cofinancing.\

5. This year's **management action record (MAR)** tracks the level of adoption of 35 Council decisions based on 12 GEF Evaluation Office documents. The Evaluation Office was able to verify 32 of these decisions and plans to carry out thematic assessments in the future to evaluate adoption of Council decisions that have not been verified so far.

6. The **performance matrix** provides a summary of the performance of GEF Agencies and the GEF Secretariat on relevant parameters. Several of the parameters included in the matrix are assessed on an annual basis by the Evaluation Office. For the remainder, the Evaluation Office—in collaboration with the evaluation units of the GEF Agencies—is developing assessment approaches. Reporting on such parameters will be provided during OPS4.

7. The APR primarily involves review of the evidence presented in the terminal evaluation reports, with verification of performance ratings based primarily on desk reviews. Although this methodology enables the Evaluation Office to make the performance ratings more consistent with the evidence provided in the terminal evaluation reports, it does not allow it to establish the veracity of the evidence presented. Direct verification permits the Evaluation Office to address this concern. In addition, the field verification process provides the Evaluation Office with an opportunity to gather new information on post project completion impacts. Beginning FY 2007 the Evaluation Office has been piloting a methodology for field verification of terminal evaluation reports. So far the Evaluation Office has carried out seven field verifications. At least nine more field verifications will be completed for inclusion in OPS4.

8. One of the limitations noted in the earlier APRs has been that, on many issues, the number of projects for which data were available was not sufficient to allow for in-depth assessment of differences in performance and the underlying factors that affect performance. This constraint is mitigated gradually as each year the data pool increases. The second assessment of M&E arrangements in GEF projects at entry (FY 2008) permits the Evaluation Office to assess change compared to FY 2005, when the first assessment was carried out. While analysis of other issues remains constrained, this year the accumulated data allows the Evaluation Office to present a more in depth discussion on project outcomes and quality of terminal evaluation reports.

1.2 Findings and Conclusions

Results

Conclusion 1: Eighty percent of completed projects reviewed for OPS 4 were rated in the satisfactory range. This is significantly higher than the 75 percent target specified in the GEF-4 replenishment agreement. Fifty eight percent were rated moderately likely or above in sustainability of outcomes.

9. **Attainment of project outcomes.** The Evaluation Office rated the achievement of project outcomes on criteria of relevance, effectiveness, and efficiency. Of the 210 terminal evaluation reports submitted since FY 2005, five (2 percent) did not provide sufficient information to allow assessment of the level of attainment of project outcomes (see table 1.1). The key findings of this assessment are as follows:

- Of the 205 projects whose outcomes were rated by the Evaluation Office, 80 percent were rated moderately satisfactory or above.
- Of the total investment in the rated projects (\$983 million), 79 percent was allocated to projects that were rated moderately satisfactory or above.
- The outcomes of 78 percent of FSPs and 84 percent of MSPs were rated moderately satisfactory or above.

10. Compared to other agencies a greater proportion of the World Bank implemented projects were rated in the satisfactory range (85 percent), when other variables such as project size, geographical scope, focal area, and geographical region were controlled for.

11. National projects implemented in “Fragile” and in Small Island Developing Countries (SIDS) have a lower probability of being rated in the satisfactory range.

Table 1.1: Outcomes and Sustainability of Outcomes

Fiscal Year	FY 2005	FY 2006	FY 2007	FY 2008	OPS4
Number of TE submitted	41	66	41	62	210
Number of projects with outcomes rating	39	64	40	62	205
Percentage rated MS or above in outcomes rating	82%	84%	73%	81%	80%
Number of projects with sustainability of outcomes rating	39	54	39	60	192
Percentage rated ML or above in sustainability of outcomes rating	49%	65%	59%	57%	58%
Number of projects rated both on outcomes and sustainability of outcomes	39	54	39	60	192
Percentage of rated projects with MS/ML or above in both	44%	61%	51%	55%	54%

12. Project performance for the OPS4 cohort is better than the target set for the fourth GEF replenishment of 75 percent of projects having satisfactory outcomes (GEF 2006). Although the completed projects assessed since FY 2005 do not fall under the provisions of the fourth replenishment, their performance indicates that it is realistic to expect that the target for OPS4 will be achieved.

13. Of the terminal evaluations submitted during FY 2008, outcome achievements for 81 percent rated moderately satisfactory and above. These projects account for 74 percent of the total GEF investment in projects included in the FY 2008 cohort. These results are not substantially different from those for the projects from previous years.

14. **Sustainability of project outcomes.** The Evaluation Office rated sustainability based on its assessment of level of risks to sustainability of outcomes on four dimensions: financial, sociopolitical, institutional and governance, and environmental. Of the 210 terminal evaluation reports submitted since FY 2005, 18 (9 percent) did not provide sufficient information to allow assessment of sustainability of project outcomes (table 1.1). The key findings of this assessment are:

- Of the 192 projects rated, the sustainability of outcomes of 58 percent (111 projects) was deemed moderately likely or above;
- Of the total GEF investment in rated projects since FY 2005 (\$907 million), 60 percent (\$543 million) was invested in projects that were rated moderately likely or above in terms of the sustainability of their outcomes.

15. The Evaluation Office assessed the extent to which projects that were rated moderately satisfactory or above in achievement of outcomes were also rated moderately likely or above in sustainability of outcomes. The Evaluation Office found that of the rated projects, 54 percent were rated both moderately satisfactory or above in outcomes and moderately likely or above in sustainability. In terms of GEF investment, of the total investment in rated projects, 55 percent was invested in these projects.

16. Of the four risks to sustainability dimensions, financial risks were found to pose the most frequent threat to sustainability of outcomes. Thirty one percent of the examined projects were rated unlikely or moderately unlikely to sustain project outcomes in account of risks pertaining to this category.

17. Of the projects pertaining to FY 2008, sustainability of outcomes of 57 percent rated moderately likely or above and 55 percent were rated both moderately satisfactory or above in terms of outcome achievements and moderately likely or above on sustainability of outcomes.

Processes

Conclusion 2: On average, the materialization of cofinancing reported by the IAs is close to that promised at project approval.

18. There are great variations among projects in terms of level of GEF investment, promised cofinancing, and reported materialized cofinancing. The figures for a cohort could easily be skewed by a few projects. Consequently, the average figures for cohorts may fluctuate despite the absence of an underlying trend. The cofinancing figures reported by the IAs for the OPS-4 cohort (FY 2005 to FY 2008) need to be noted with this caveat.

19. The following analysis is based on the information available on 285 completed projects for which terminal evaluations were submitted since FY 2002. Terminal evaluations for 75 of these projects were submitted on or before FY 2004. The remainder (210) comprises the OPS4 cohort. Of the terminal evaluations submitted during or before FY 2004, for 48 projects information on materialization of cofinancing has been reported. For these projects \$5.9 cofinancing had been promised for every dollar of GEF financing and \$5.4 was reported to have materialized (see table 4.1). In comparison, of the 210 terminal evaluation reports submitted since FY 2005, 162 reported on materialization of cofinancing. For these projects, an average of \$2.1 had been promised per dollar of approved GEF grant. Of this, \$2.0 was reported to have materialized. The rate of

materialization of cofinancing is easily skewed by a few projects. Therefore, the difference – in terms of materialization of cofinancing – between the projects from the pre-OPS4 period and OPS4 period does not in itself indicate an underlying trend.

20. If all terminal evaluation reports submitted to the Evaluation Office to date are taken into consideration, information on cofinancing is available for 210 projects. For these, the Agencies promised an average of \$3.2 in cofinancing per \$1.00 of approved GEF grant. The actual cofinancing reported was slightly lower: \$3.0 per \$1.00 of approved GEF grant. Thus, an average of 95 percent of promised cofinancing was reported to have materialized.

21. Projects for which a lesser proportion of promised cofinancing actually materialized have a lower probability of being rated in the satisfactory range. However, it is not clear whether overall it is the lower materialization of cofinancing that is driving lower levels of achievement or it is the lack of progress towards achievement of results that leads to lower materialization of cofinancing. These issues will be explored in greater detail in the OPS4 report.

22. **On average, the projects ending during OPS4 were completed after a delay of 16 months; 22 percent were completed after a delay of at least two years.** The Evaluation Office began tracking project completion delays in FY 2005. The average project completion delay was 19 months for the FY 2005 cohort, 13 months for the FYs 2006 and 2007 cohorts and 18 months for the FY 2008 cohort. In comparison, 44 percent of the projects in the FY 2005 cohort, 17 percent of the projects in the FY 2006 cohort, 14 percent of the FY 2007 cohort, and 26 percent of the FY 2008 cohort had delays of two years or more. Thus, there is no definite trend that is visible.

Monitoring and Evaluation

Conclusion 3: There has been a significant improvement in the quality of M&E arrangements at the point of CEO endorsement, 76 percent of projects endorsed by the CEO during FY 2008 met the existing GEF M&E minimum requirements. In FY 2005, only 58 percent had met minimum requirements.

23. The GEF Evaluation Office carried out an assessment of the quality of M&E arrangements of FSPs that were endorsed by the CEO during FY 2008 as a follow up to the assessment carried out during FY 2005. An identical methodology was used in both assessments. These assessments included reviews for all FSPs approved in FY 2005 (74 FSPs) and in FY 2008 (82 FSPs).

24. The office rated M&E arrangements of 76 percent of FSPs that were CEO Endorsed in FY 2008 to be in compliance with the minimum requirements applicable in FY 2008. This is a significant improvement over the FY 2005 performance, when 58 percent of the projects were rated to be compliant with the requirements applicable in FY 2005. If the

requirements applicable in FY 2005 are applied to the FY 2008 cohort, then the percentage of projects in compliance increases to 80 percent. A greater proportion of World Bank and Climate Change projects were rated to be in compliance in FY 2008 than in FY 2005. For other agencies and focal areas the number of observations is too small to allow similar conclusions. However, for most, the direction of change is consistent with the overall trend.

Conclusion 4: There is a strong association between quality at entry of M&E arrangements and actual quality of monitoring during implementation.

25. Among the terminal evaluations submitted during OPS4, 77 percent provided sufficient information to allow the Evaluation Office to rate performance in this parameter. Among those that did provide sufficient information, 67 percent – including 72 percent from FY 2008 – were rated moderately satisfactory or above in terms of monitoring during implementation. This is significantly higher than the 55 percent that were so rated in the FY 2004 cohort (table 1.2).

26. The analysis of the information pertaining the terminal evaluations submitted for projects completed during OPS4 supports findings of previous APRs (APR 2006 and APR 2007) indicating that there is a correlation between quality of M&E arrangements and actual quality of monitoring during implementation. Of the completed projects from the OPS4 cohort that were rated both on quality of monitoring during implementation and quality of M&E at entry, only 39 percent of those rated in the unsatisfactory range on quality at entry were rated in the satisfactory range during implementation. In contrast, of those rated in the satisfactory range on quality of M&E at entry, 82 percent were also rated in the satisfactory range during implementation.

Table 1.2 Quality of monitoring during project implementation

FY of TE submission	FY 2004	FY 2005	FY 2006	FY 2007	FY 2008	OPS4
Terminal evaluation reports submitted	42	41	66	41	62	210
Terminal evaluations that reported on M&E	29	32	46	33	50	161
Percentage of projects rated moderately satisfactory or above	55%	66%	78%	61%	72%	67%

Conclusion 5: During the OPS4 period there has been significant improvement in the overall quality of terminal evaluation reports. However, further improvements are required in the reporting on M&E and financial information.

27. There were 210 terminal evaluation reports submitted to the Evaluation Office from FY 2005 to FY 2008. Of these reports, 89 percent were rated as moderately satisfactory or above by the office and 92 percent of those submitted in FY 2008 were similarly rated. This represents an important improvement from the baseline of FY 2004, when only 69 percent of the terminal evaluations was rated moderately satisfactory or above on quality.

28. For the OPS4 cohort there is little difference among agencies in terms of the terminal evaluations rated moderately satisfactory or above. Within agencies, for the World Bank the percentage of terminal evaluations for MSPs that are rated moderately satisfactory or above in quality is lower than for FSPs. This reflects the lower attention that World Bank accords to the terminal evaluations for MSPs.

29. When the OPS4 period is split in two periods of two years each the terminal evaluations submitted by UNEP show significant improvements (table 6.1). These gains are linked to the measures adopted by UNEP, including the transfer of the responsibility for the evaluation of GEF projects to the UNEP's Evaluation Office and a closer tracking of the quality of evaluations.

30. Despite improvements in overall quality of terminal evaluation reports, performance in the various quality dimensions remains uneven. Ratings on reporting on M&E and on financial information for the project are considerably lower than ratings for other dimensions.

31. Outcome ratings in terminal evaluations provided by the evaluation offices of the implementing agencies are generally consistent with those provided by the GEF Evaluation Office – on a binary scale there are negative disconnects between the Evaluation Office ratings and those provided by the World Bank's Independent Evaluation Group (IEG) and UNEP's Evaluation Office for only 4 percent of the projects (UNDP thus far has not provided ratings on outcomes). Disconnect with ratings given in terminal evaluations and Project Implementation Reports (PIRs) is significantly higher: 14 percent for terminal evaluations and 16 percent for PIR. This indicates that while there is much consistency in the ratings given by the GEF Evaluation Office and those given by the evaluation offices of World Bank and UNEP, significant differences remain with regards to the ratings given in terminal evaluations and in PIRs.

Management Action Record

Conclusion 6: While there is an overall high level of adoption of Council decisions by the GEF system, in few important cases progress has been slow.

32. This year's MAR tracks the level of adoption of 35 Council decisions based on 12 GEF Evaluation Office documents by presenting ratings from GEF management and verification of these ratings by the Evaluation Office. The Evaluation Office was able to verify the adoption of 32 of these 35 decisions.

33. The Evaluation Office has rated a total of 15 decisions as adopted since it introduced the MAR in FY 2005. These include 5 decisions rated as high in FY 2008 and 10 rated as high in prior years. The decisions adopted in FY 2008 include two decisions that come from the *Evaluation of the GEF Support to Biosafety*; and one decision each from the *Joint Evaluation of the SGP*, the *APR 2007* and the *RAF Midterm Review*.

34. Several Council decisions that require a more nuanced treatment have been deferred for further analysis and assessment during OPS4. These include:

1. From the *Country Portfolio Evaluation of Samoa*, the extent to which the higher transaction costs of working in the Pacific has been taken into account in the design of programmatic approaches for this area;
2. From the *Local Benefits Study*, the attention given in the GEF to social issues in project preparation and implementation;
3. From the Council decisions of various evaluations (*Costa Rica CPE, Philippines CPE, APR 2004, and APR 2006*), to give priority to the development of a management information system that is transparent and accurate; and
4. From the *Evaluation of the GEF Activity Cycle and Modalities*, the efficiency of the PIF process.

35. For the second consecutive year the percentage of Council decisions verified by the Evaluation Office that have been rated at a substantial or high level has increased; from 44 percent in FY 2007 to 59 percent in FY 2008. Also, while there was an increase in the number of Council decisions with no change in the rating, from 33 percent in FY 2007 to 55 percent in FY 2008; many of these ratings reflect the fact that even though the GEF has taken measures to address the decisions, these cannot yet be considered to demonstrate a high level of achievement. They also reflect the fact that some proposals to the GEF Council have not yet been approved. If, and when, the Council does approve these proposals, substantial adoption is likely to occur.

36. The Evaluation Office considers there has been little progress in the adoption of four Council decisions from the *Local Benefits Study* which point out the need to establish a system that ensures that local benefits are addressed in a more systematic way at all stages of the GEF project cycle. The GEF Secretariat currently cannot verify the quality of this aspect in project design or implementation because it still has no system in place to involve specialist social development expertise in its project review processes.

37. Two issues pertaining to Council decisions have resurfaced. The first one includes five decisions related to the improvement of the PMIS and to the need to ensure transparency of, and better access to, information on the GEF procedures, project approval criteria and the status of projects in the GEF project cycle. These decisions come from the *Costa Rica CPE, the Philippines CPE, the APR 2004 and the APR 2006*. Particularly important is the Council decision in the *APR 2006* that instructed the GEF Secretariat to make the development of a management information system a priority activity.

38. The other issue pertains to the Council decision from the *Samoa CPE*, which requested the Secretariat to take into account Samoa's experience with the GEF in its further development of the proposed GEF-Pacific Alliance for Sustainability. Adoption of

this decision was rated as high by both the GEF Secretariat and the GEF Evaluation Office in the last MAR. This year, the Evaluation Office considers that further analysis needs to be done on the inclusion of higher transaction costs in the Pacific Alliance for Sustainability project. The Fourth Overall Performance Study that is currently underway will take a more in depth look into programmatic approaches and will verify whether or not it was justified to include, or not include, transaction costs in the design of these types of projects.

Performance Matrix

39. The performance matrix provides a summary of the performance of the GEF Agencies and GEF Secretariat on 13 parameters, covering key areas such as results, processes affecting results, efficiency, M&E, and learning. Several of the parameters included in the matrix are already assessed by the Evaluation Office on an annual basis. Since performance ratings on these parameters fluctuate from year to year, running averages of two to four years, depending on the parameter, are used in the matrix (see chapter 9 and annex D for methodological details on the performance matrix). Of the 13 parameters included in the performance matrix, ratings have been provided on 10. The rating on the indicator on independence of terminal evaluations is being included for the first time. Also note that the information provided for parameter 13, improvement in performance, addresses only two of the parameter's 12 dimensions. In the future, as data for more years become available, it will be possible to track improvements on a greater number of dimensions.

40. Based on the review of terminal evaluation reports submitted to the Evaluation Office since FY 2005, the Evaluation Office rated outcome achievement (parameter 1) in 80 percent of the projects to be moderately satisfactory or above. As noted, this percentage is higher than the 75 percent target specified in the GEF-4 replenishment agreement, even though these projects are not subject to this provision. The percentage of World Bank projects with ratings in the satisfactory range (85 percent) has been significantly higher than the target.

41. On independence of terminal evaluations, those submitted by the UNDP and UNEP have been rated "satisfactory" on a six point scale, for both FSPs and MSPs. The terminal evaluations submitted by the World Bank have been rated "highly satisfactory" for the FSPs and "moderately unsatisfactory" for MSPs. The independence of the terminal evaluations for the World Bank FSPs is rated highly satisfactory because in addition to the desk reviews of the terminal evaluations, their evaluation office also conducts field verifications for a sample of terminal evaluations. The World Bank terminal evaluations for MSPs are rated moderately satisfactory because they do not undergo any review by their evaluation office.

42. For parameter 13, changes in performance have been assessed on two dimensions:

- Change in the quality of M&E arrangements at entry (parameter 10)

- Change in the quality of terminal evaluation reports (parameter 12)

43. The project appraisal documents submitted by the World Bank and UNDP showed improvement in terms of quality of M&E arrangements at entry. However, there is still considerable room for further improvement. The number of observations for UNEP is too small to draw conclusions. Although, the M&E arrangements in their project appraisal documents have made improvements on several parameters, the baseline information provided is not sufficient.

44. The ratings on changes in quality of terminal evaluation reports, provided in the performance matrix of *APR 2007*, have not changed. In recent years, the terminal evaluations submitted by UNEP and UNDP have significantly improved; most of the terminal evaluations submitted by the World Bank continue to be in the satisfactory range.

Chapter 2: Scope and Methodology

2.1 Scope

45. In the APR, each year the Evaluation Office presents an assessment of the results of completed GEF projects, an analysis of the processes that affect accomplishment of results, and the findings of its oversight of project monitoring and evaluation activities across the portfolio. Through the APR the Evaluation Office provides feedback to the GEF Council, other GEF institutions, and stakeholders to help improve the performance of GEF projects. Some issues are addressed in the APR annually, some biennially; others are addressed whenever such a need is felt.

46. One of the purposes of the APR 2008 is to provide inputs to the Fourth Overall Performance Study (OPS4) of the GEF. Therefore, the projects whose terminal evaluations were submitted since the Third Overall Performance Study period, i.e. from FY 2005 onwards, remain the focus of discussion. Special attention has been given to projects for which the terminal evaluations were submitted in FY 2008 and, therefore, not covered in the earlier APRs (see Annex B for the list of the projects). The APR 2008 includes the following:

- **An overview of the extent to which GEF projects are achieving their objectives (chapter 3).** This overview consists of an assessment of the extent to which the completed projects, for which terminal evaluation reports were submitted from FY 2005 to FY 2008 (the complete cohort for OPS4), achieved expected outcomes and the risks to sustainability of the achieved outcomes. These issues are covered in the APR annually.
- **Presentation of the materialization of project cofinancing and delays in project completion reported by the GEF Agencies (chapter 4).** This chapter reports on the extent to which cofinancing promised at the point of project endorsement has materialized and on delays in project completion. The

assessment on materialization of cofinancing is based on figures reported by the respective GEF Agencies. This issue is being tracked since FY 2002 and is reported on annually.

- **An assessment of the quality of project monitoring (chapter 5).** This chapter includes an assessment of the quality of M&E arrangements at entry for the full size projects that were CEO Endorsed during the FY 2008, and a comparison of its findings with that of a similar assessment conducted for the FY 2005 cohort that was included in the *APR 2005*. In addition, it includes an assessment of the quality of M&E during implementation in completed projects for which terminal evaluations were submitted during the OPS4 period.
- **An assessment of the quality of terminal evaluation reports submitted by the GEF Agencies to the Evaluation Office (chapter 6).** This chapter, provides information on the quality of terminal evaluation reports by focal area and by Agency and an assessment of trends in the quality of terminal evaluation reports. It also includes a comparison of the ratings on outcome achievements for assessed projects given by the Evaluation Office with those provided by others including the evaluation offices of the Implementing Agencies, in the terminal evaluations, and in project implementation reports submitted by the Implementing Agencies. The progress made by the Evaluation Office in piloting a methodology for direct verification of terminal evaluations has also been summarized.
- **A presentation of findings on management action records (chapter 7).** As part of this annual assessment, the Evaluation Office reviews and follows up on the implementation status of evaluation recommendations that have been accepted by management and/or the GEF Council.
- **A presentation of the performance matrix (chapter 8).** This assessment was first presented in *APR 2007* and is now a regular feature of the report. It summarizes the performance of the GEF Implementing Agencies and the GEF Secretariat on key performance parameters.

2.2 Methodology

2.2.1 Terminal Evaluation Reviews

Terminal Evaluations included in the review

47. Up to FY 2006 the terminal evaluations included in the review pertained to full size and medium size projects that had been closed after operational completion of project activities. This meant that the terminal evaluations for cancelled projects – in some of which a significant proportion of GEF grant had been utilized at the point of cancellation – and Enabling Activities did not form part of the project pool that was rated. From FY 2007 onwards the Evaluation Office has also started reviewing the terminal evaluations of cancelled projects, for which a GEF grant of more than US \$ 0.5 million had been utilized at the point of cancellation, as part of the APR terminal evaluation

review process. In addition, the Evaluation Office has started reviewing the terminal evaluations of those Enabling Activities that have a GEF investment of more than US \$ 0.5 m. The included Enabling Activities projects are reported as MSPs or FSPs based on the level of GEF investment.¹

Ensuring Reliability and Timeliness of Terminal Evaluation Reports

48. The project terminal evaluation reports submitted by the GEF Agencies to the Evaluation Office form the core information source for much of the APR, particularly for those topics that are reported on annually. Ensuring the reliability of these reports is therefore critical. The Evaluation Office seeks to assess and strengthen this reliability in several ways, as described below.

49. The Evaluation Office reviews terminal evaluation reports to determine the extent to which they address the objectives and outcomes set forth in the project document, to evaluate their internal consistency, and to verify that ratings are properly substantiated.

50. The reports are reviewed by Evaluation Office staff using a set of detailed guidelines to ensure that uniform criteria are applied (see annex A for these guidelines). When deemed appropriate, a reviewer may propose to upgrade or downgrade the project ratings presented in a terminal evaluation report.

51. A draft terminal evaluation review report is also examined by a peer reviewer with substantial experience in assessing terminal evaluations. The peer reviewer provides feedback on the report, which is incorporated by the primary reviewer in subsequent versions.

52. When projects are downgraded below moderately satisfactory (for outcomes), a senior evaluation officer in the GEF Evaluation Office also examines the review to ensure that the proposed ratings are justified. The reviews are then shared with the Agencies, and, after their feedback is taken into consideration, the reviews are finalized.

53. If a terminal evaluation report provides insufficient information to make an assessment or to verify the Agency ratings on outcomes, sustainability, or quality of project M&E systems, the Evaluation Office classifies the corresponding project as unable to assess and excludes it from any further analysis on the respective dimension.

54. The most pervasive limitation of this review process is that, ultimately, it is only based on a desk review of information provided by the terminal evaluation reports and other relevant documents. To address this weakness, the Evaluation Office has been piloting an approach to carry out verification of the terminal evaluation reports for a sample of completed FSPs. In all seven verifications have been conducted so far. Nine more will be completed in time for their results to be incorporated in the OPS4 – these

¹ The number of cancelled projects with over US \$ 0.5 m of GEF grant utilized and Enabling Activities with US \$ 0.5 m in GEF investment is too small to be reported as separate category.

additional field verifications are being conducted in countries where OPS4 country case studies are being undertaken.

55. Another way to address the reliability concerns pertaining to terminal evaluation reports is to work with the GEF Agencies to more fully engage their central evaluation groups in the process and, when necessary, to strengthen their independence. Presently, the World Bank's terminal evaluation process for FSPs meets most of the concerns of the GEF Evaluation Office. The Bank's Independent Evaluation Group (IEG) conducts desk reviews of all the terminal evaluation reports produced by the management for FSPs and conducts direct verification for a sample of these reports. The terminal evaluations for the World Bank implemented MSPs are, however, not reviewed by the IEG.

56. Beginning in FY 2006, the UNEP Evaluation Office started providing ratings and commentary on the quality of the terminal evaluation reports for the completed GEF projects implemented by UNEP. During FY 2007, it increased the scope of its commentaries by also assessing project outcomes, sustainability of outcomes, and implementation of M&E based on the evidence provided in the terminal evaluation reports. Although the ratings provided by UNEP Evaluation Office are consistent with those provided by the GEF Evaluation Office, since only a few assessments have been completed so far, it is still too early to determine their overall reliability.

57. During FY 2007 the UNDP Evaluation Office began to provide commentary on the quality of terminal evaluations for some of its completed GEF projects. The UNDP Evaluation Office has yet to begin reviewing project performance in terms of outcomes, sustainability of outcomes, and implementation of M&E, and provide ratings on these parameters.

58. The GEF Evaluation Office will continue its dialogue with the GEF Agencies so that they undertake independent review of the terminal evaluations and verify ratings. At the same time, the Evaluation Office will continue to review the terminal evaluation reports and verify their ratings.

59. The GEF Evaluation Office has been making efforts to ensure timely submission of terminal evaluation reports. Although the time lag between completion and submission has been declining, many reports are still being submitted after more than the two months that is the negotiated norm for this task. The World Bank has a system for automatic submission of an electronic version of terminal evaluation reports for its GEF-supported FSPs when these reports are completed. However, this system is effective only for projects that are identified in the World Bank database (warehouse) as GEF projects. There have been instances where projects, especially the blended ones, supported by GEF have not been identified as such. To address this lacuna the Evaluation Office now undertakes a time consuming process of identifying the projects that are expected to have been completed but for which terminal evaluations have not yet been submitted and then tracking the status of those projects in the World Bank intranet. The terminal evaluations for the MSPs are generally submitted to the Evaluation Office annually. Other GEF Agencies have not yet developed automatic submission systems. The Office will work

with the Agencies so that timely submission of all the terminal evaluation reports can be ensured.

Data Limitations

60. 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. During the past three fiscal years, 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.

61. Since data are now available for a greater number of completed projects and over a longer period of years, the GEF Evaluation Office is better able to predict trends and assess differences in performance. However, 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.

62. 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.

2.2.2 Quality of M&E at Entry

63. This assessment is a follow up on the assessment on quality at entry of M&E arrangements assessment presented in *APR 2005* and uses an identical methodology to facilitate comparisons over time. The Evaluation Office assessed the quality of M&E arrangements in full-size projects at the point of CEO endorsement. All full-size projects endorsed by the GEF CEO during FY 2008 were considered. The results of the review were then compared with those for the FY 2005 cohort (that were presented in *APR 2005*).

2.2.3 Management Action Records (MARs) Assessment

64. Management action records facilitate review and follow-up on the implementation status of evaluation recommendations that have been accepted by management (that is,

the GEF Secretariat and/or the GEF Agencies) and/or the GEF Council. For each MAR, the Evaluation Office completes the columns pertaining to recommendations, management responses, and Council decisions. The management is then invited to provide a self-rating of the level of adoption of Council decisions and add any comments as necessary. After the management's response is included in a MAR, the Evaluation Office verifies actual adoption and provides its own ratings, with comments, in time for presentation to the Council.

2.2.4 Performance Matrix

65. The performance matrix, which was presented for the first time in *APR 2007*, provides a summary of the performance of the three main GEF Implementing Agencies and the GEF Secretariat on relevant parameters. Performance on most of these parameters, such as project outcomes, implementation completion delays, materialization of cofinancing, quality of M&E during project implementation, and quality of project terminal evaluations, is already being assessed annually by the GEF Evaluation Office. Performance on other parameters, such as quality of supervision and adaptive management, realism of risk assessment, and quality of project M&E arrangements at entry, is being assessed and updated every two or three years through special appraisals. For assessing performance on project preparation elapsed time, assessments will be presented based on the GEF Project Management and Information System (PMIS) database and will be included from *APR 2009* onwards. Performance on the indicator on independence of terminal evaluations is being reported on for the first time in *APR 2008*. Improvement in performance on these indicators has been included as an indicator to measure institutional learning.

66. Two of the parameters included in the performance matrix – independence of GEF partner agency evaluation units; and, the assessment of robustness of program result indicators and tracking tools – will require development of new methodologies and approaches. Both these parameters will be addressed as part of OPS4.

2.2.5 Review of Findings

67. The preliminary findings of this report were presented to and discussed with the GEF Secretariat and GEF Agencies during an interagency meeting held in Washington, D.C., on the 12th of May 2009. Individual reviews of project terminal evaluation reports have been shared with the Agencies and GEF Secretariat for comments and their feedback incorporated.

Chapter 3: Outcomes and Sustainability

68. This chapter discusses verified ratings on outcomes and sustainability for 210 projects for which terminal evaluation reports have been submitted from FY 2005 to FY 2008. These completed projects constitute the cohort being assessed for OPS4 and account for a total of US \$ 989 million in GEF investments. Of these projects, 205 were rated on outcome achievements and 192 on sustainability of outcomes. Among those

rated, outcome achievements of 80 percent, accounting for 79 percent of the GEF investment, rated moderately satisfactory or above (table 3.1 and 3.2). The policy recommendations for GEF-4 contain a target of outcomes of 75% of finished projects rated in the satisfactory range (moderately satisfactory or above). Although the completed projects assessed since FY 2005 do not fall under the provisions of the fourth replenishment, their performance indicates that it is realistic to expect that the target for OPS4 will be achieved. Fifty eight percent of the projects rated for OPS4 rated moderately likely or above in terms of the sustainability of their outcomes. Fifty four percent rated both moderately satisfactory and above in their outcome ratings and moderately likely or above in their sustainability ratings.

69. Of the projects reviewed for OPS4, terminal evaluations for 62 of them were submitted in FY 2008. These include 32 full-size² and 30 medium-size projects. Of the projects rated for outcome achievements, 81 percent rated moderately satisfactory or above and these projects accounted for 74 percent of the GEF investment. Fifty seven percent of the projects rated moderately likely or above on sustainability of outcomes. Fifty five percent rated both moderately satisfactory and above in their outcome ratings and moderately likely or above in their sustainability ratings.

3.1 Rating Approach

70. The Evaluation Office rated project outcomes of the FY 2008 cohort of 62 projects based on level of achievement of project objectives and expected outcomes in terms of **relevance** on a two point scale and **effectiveness** and **efficiency** on a six-point scale. Up to *APR 2007*, relevance of outcomes had been rated on a six point scale. The assessment on relevance primarily focuses on determining whether the anticipated outcomes are relevant to the GEF mandate for generating global environmental benefits.

71. This year, only **satisfactory** or **unsatisfactory** rating was provided on relevance. Relevance ratings are considered critical: if relevance of outcomes rating is unsatisfactory then the overall outcome ratings may not be higher than unsatisfactory. Among the other criteria **effectiveness** continues to be a critical criterion; the overall rating on achievement of outcomes could not be higher than the rating on effectiveness. The modifications made in rating approach do not lead to changes in the outcome ratings presented in *APR 2007*, i.e. outcome ratings of the 210 projects reviewed for OPS4 are comparable.

72. During project implementation, the results framework of some projects had been modified. This poses a challenge because assessing actual outcomes for all projects based on original outcome expectations may discourage adaptive management. To address this challenge, for projects where modifications and improvements are made in the project objectives, outcomes and outputs, without a down scaling of their overall scope, the Office assesses outcome achievements based on the revised results framework. In instances where the scope of the project objectives, outcomes and outputs has been

² This includes one Enabling Activity project with GEF investment of over US dollar one million.

downscaled, the expected outcomes at the start of the project are used as a yard stick for performance assessment.

73. The GEF Evaluation Office rated sustainability of outcomes based on an assessment of four key risk dimensions: financial, sociopolitical, institutional framework and governance, and environmental. Based on the evidence presented in the terminal evaluation reports, **risks to sustainability of outcomes** were assessed on each of these dimensions. All risk dimensions were regarded as critical; overall ratings may not be higher than the lower rating on any of these dimensions.

74. Among the 210 terminal evaluations reviewed, five (2 percent) did not provide sufficient information to allow the Evaluation Office to rate outcome achievements, and 18 (9 percent) to rate risks to sustainability of outcomes. Among the 62 terminal evaluations reviewed in FY 2008, sufficient information has been provided in all to rate outcome achievements whereas it is not adequate to rate sustainability of outcomes for 2 projects (3 percent).

3.2 Outcomes

75. Of the projects reviewed for OPS4, outcome achievements of 80 percent were rated moderately satisfactory or above. This is better than the target set for the fourth GEF replenishment of 75 percent of projects having satisfactory outcomes (GEF 2006). Although the completed projects assessed since FY 2005 do not fall under the provisions of the fourth replenishment, their performance indicates that it is realistic to expect that the target for OPS4 will be achieved.

Table 3.1 Distribution of GEF Projects: by Outcome ratings

Outcome Rating	FY 2005	FY 2006	FY 2007	FY 2008	Total
By Number of Projects					
Number of terminal evaluations Submitted	41	66	41	62	210
Number of projects rated on outcomes	39	64	40	62	205
Highly Satisfactory	3%	6%	3%	5%	4%
Satisfactory	54%	44%	35%	52%	46%
Moderately Satisfactory	26%	34%	35%	24%	30%
Moderately Satisfactory or Above	82%	84%	73%	81%	80%
Moderately Unsatisfactory	10%	14%	8%	13%	12%
Unsatisfactory	8%	2%	18%	5%	7%
Highly Unsatisfactory	0%	0%	3%	2%	1%
Moderately Unsatisfactory or Below	18%	16%	27%	19%	20%

76. Seventy nine percent of GEF investment in rated projects is in those that rated moderately satisfactory or above. Overall, although the proportion of projects and of investment in the projects rated in the satisfactory range is similar, fluctuations across the annual cohorts are higher for as it tends to get skewed due to a few projects that have high investment (Table 3.1 and 3.2). Chapter 4 presents a detailed discussion on drivers of project outcome achievements.

Table 3.2 Distribution of GEF Investment: by Outcome Ratings

Outcome Rating	FY 2005	FY 2006	FY 2007	FY 2008	Total
Total GEF investment in reviewed projects	258.3	255.8	199.3	275.3	988.7
Total GEF investment in rated projects	255.3	254.3	198.3	275.3	983.2
Highly Satisfactory	0%	6%	5%	8%	5%
Satisfactory	64%	30%	18%	55%	43%
Moderately Satisfactory	20%	53%	46%	12%	31%
Moderately Satisfactory or Above	84%	88%	69%	74%	79%
Moderately Unsatisfactory	15%	11%	14%	13%	13%
Unsatisfactory	1%	1%	12%	10%	6%
Highly Unsatisfactory	0%	0%	5%	3%	2%
Moderately Unsatisfactory or Below	16%	12%	31%	26%	21%

77. Table 3.3 provides information on the outcome ratings of different categories of GEF projects based on Implementing Agency; the agency type for project execution; focal areas; project size; geographical scope; geographical region; and, important country groupings. Outcomes of 78 percent of the FSPs³ projects and 84 percent of the MSPs were rated moderately satisfactory or above. Among the agencies, 85 percent of projects implemented by the World Bank, 78 percent by UNDP and 72 percent by UNEP, were rated moderately satisfactory or above. Among the agencies that execute GEF projects, 82 percent of the projects executed by the government agencies (including quasi governmental agencies), 73 percent by multilateral and bilateral agencies, 79 percent by NGOs and Foundations, and 92 percent by others including private sector organizations, rated moderately satisfactory or above. In terms of geographical regions, 83 percent of the national and 74 percent of the regional and global projects were rated moderately satisfactory or above. Among the geographical regions, outcomes of 84 percent of the projects implemented in Latin America and the Caribbean and 73 percent of those implemented in Africa rated moderately satisfactory or above.

78. Analysis of outcome ratings on a binary scale for different categories of projects shows that:

- Compared to other agencies a greater proportion of the World Bank implemented projects were rated in the satisfactory range (85 percent), when other variables such as project size, geographical scope, focal area, and geographical region were controlled for.
- National projects implemented in “Fragile” and in Small Island Developing Countries (SIDS) have a lower probability of being rated in the satisfactory range.

³ This includes one Enabling Activity project with GEF investment of over US dollar one million.

Table 3.3 Outcome performance: by categories of projects

Category	Number of Projects Reviewed	Number of Projects Rated	Percentage Rated Projects that are rated MS or Above
All projects	210	205	80%
Implementing Agencies			
World Bank	95	93	85%
UNDP	85	83	78%
UNEP	30	29	72%
Agencies that execute projects			
Government and parastatal agencies	109	107	82%
NGOs and Foundations	49	48	79%
Bilaterals and Multilaterals	39	37	73%
Others including private sector organizations	13	13	92%
Focal Areas			
Biodiversity	117	116	81%
Climate Change	51	49	84%
International Waters	23	23	78%
Other focal areas	19	17	71%
Project Size			
Full Size Projects	116	114	78%
Medium Size Projects	94	91	84%
Geographic Scope of Projects			
National Projects (Single Country Projects)	150	147	83%
Regional and Global Projects	60	58	74%
Geographical Areas			
Africa	45	45	73%
Asia	57	55	85%
Europe and Central Asia	36	36	78%
Latin America and the Caribbean	52	51	84%
Country Groups⁴			
Fragile	11	11	64%
Small Island Developing States (SIDS)	14	14	71%
Least Developed Countries (LDCs)	23	23	78%
Land Locked Countries	24	24	83%

- Although a greater proportion of MSPs, Climate Change projects, national projects, projects executed by ‘others including private sector’, and project implemented in Asia are rated in the satisfactory range, the difference is not statistically significant.
- Between the national projects and regional & global projects, although there is a higher probability of the former being rated in the satisfactory range the difference (owing to the small number of observations for the latter) is marginal⁵.

⁴ The figures pertain to projects that were either national projects or regional and global projects where all the participating countries were members of the relevant group.

⁵ Statistically significant at 85 percent confidence but not at 90 percent.

79. Differences in outcome achievement ratings among various project categories become more pronounced on the six point scale. On a six point scale – after controlling for other variables project size, geographical scope, focal area, and geographical region – the MSPs are rated higher than other projects. Projects implemented by UNEP and those implemented in Africa are rated lower.

80. Of the projects from the FY 2008 cohort, outcome achievements of 81 percent rated moderately satisfactory and above. These projects account for 74 percent of GEF investments in total rated projects. Thus, the results are not different from those for projects from other annual cohorts (table 3.1).

3.3 Sustainability of Project Outcomes

81. The rating on sustainability of project outcomes assesses the likelihood of continued benefits after the GEF project ends. To assess outcome sustainability, the criticality and probability of risks affecting continuation of benefits at the point of project completion are assessed. Of the 210 terminal evaluation reports submitted from FY 2005 to FY 2008, 18 did not provide sufficient information to allow assessment of sustainability of project outcomes. Of 192 that were rated, sustainability of outcomes of 58 percent rated moderately likely or above; while 54 percent were rated both moderately satisfactory or above in terms of outcome achievements and moderately likely or above in terms of sustainability. In terms of GEF investments, 60 percent was in projects with sustainability of outcomes rated moderately likely or above; 54 percent was in those that rated both moderately satisfactory or above in terms of outcome achievements and moderately likely or above in terms of sustainability (table 3.4).

82. Among the various categories of risks, financial risks pose threat to the sustainability of outcomes for the largest proportion of projects: the outcomes of 10 percent of projects were unlikely, and of an additional 21 percent were moderately unlikely, to be sustained due to risks pertaining to this category. The outcomes achievements of 23 percent of the projects were unlikely or moderately unlikely to be sustained due to institutional and governance related risks and for 19 percent due to socio-political risks (figure 3.1).

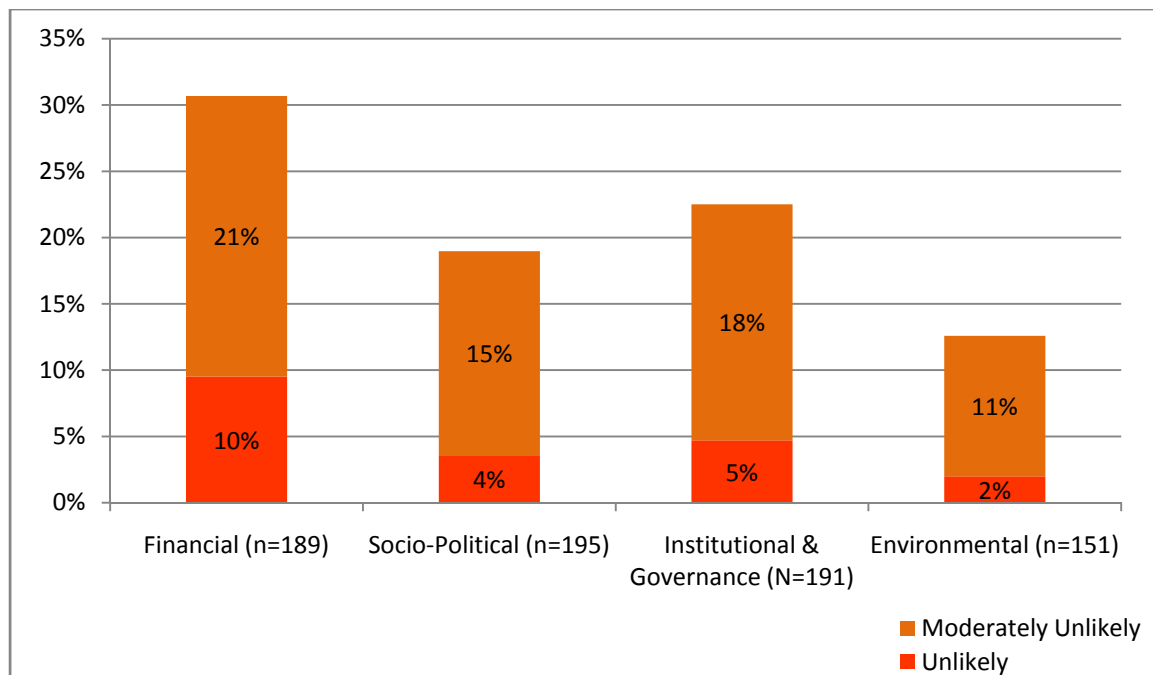
Table 3.4 Sustainability and Outcome Achievements

Category	FY 2005	FY 2006	FY 2007	FY 2008	Total
Number of Projects					
Sustainability of Outcomes rated ML or above	49%	65%	59%	57%	58%
Outcomes rated MS or above and Sustainability of Outcomes rated ML or above	44%	61%	51%	55%	54%
GEF Investment in Projects					
Sustainability of Outcomes rated ML or above	65%	60%	55%	58%	60%
Outcomes rated MS or above and Sustainability of Outcomes rated ML or above	60%	56%	44%	56%	55%

83. Of the projects pertaining to FY 2008, sustainability of outcomes of 57 percent rated moderately likely or above; while 55 percent were rated both moderately

satisfactory or above in terms of outcome achievements and moderately likely or above in terms of sustainability of outcomes. In terms of GEF investments, 58 percent of the investment was in projects that rated moderately likely or above for sustainability of outcomes; while 56 percent was in those that rated both moderately satisfactory or above in terms of outcome achievements and moderately likely or above in terms of sustainability of outcomes.

Figure 3.1: Risks to Sustainability Ratings: OPS4 Projects



84. As also for the OPS4 cohort, for the project reviewed for FY 2008 financial risks were the most important risk category: outcomes sustainability of 5 percent rated unlikely and that of 22 percent rated moderately unlikely due to this category of risk. Lack of financial support for follow up activities posed considerable risk to outcomes sustainability of “Management of Indigenous Vegetation for the Rehabilitation of Degraded Rangelands in the Arid Zone of Africa” (GEF ID 504, UNEP) and “Conservation and Sustainable Use of Biodiversity in the Andes Region” (GEF ID 774, World Bank). For “Improving the Energy Efficiency of the Heat and Hot Water Supply” (GEF ID 983, UNDP) the project management was unable to identify and leverage additional sources of funding from government and other donors for the follow up activities. Outcomes of a substantial proportion of projects were unlikely or moderately unlikely to sustain due to socio-political risks (FY 2008: 22 percent), institutional and governance risks (FY 2008: 22 percent) and environmental risks (FY 2008: 13 percent).

4. Factors Affecting Attainment of Project Results

85. Project outcomes may be affected by factors such as project design, quality of project implementation and execution, the operational context in which projects are implemented and executed, and exogenous factors beyond the control of project management. Given the wide range of variables and their interactions, it is difficult to isolate variables and determine their specific effects on a project's results. However, associations among variables and results can be determined. The Evaluation Office has been reporting on some of these variables such as cofinancing and delays in project completion annually. This year the Evaluation Office also reports on changes made in the results framework during implementation and the factors that are reported to have led to lower outcome achievements of reviewed projects.

4.1 Changes in the Results Framework

86. The results framework included in the project appraisal document submitted to the GEF for approval/endorsement by the CEO establishes project outcome expectations. At the time of project completion, these ex-ante expectations generally form a yard stick for assessment of outcome achievements. However, in some instances during the course of project implementation the implementing agency may make changes to the results framework. As discussed in Chapter 3 (section 3.1) of this report, the Office takes into account the nature of these changes to determine whether using the original results framework is more appropriate to assess project's outcome achievements.

87. The changes in the results framework are generally indicative of adaptive management by the implementing agencies. In some instances they are reflective of the attempts to adjust the results framework to unexpected exogenous changes taking place in a project's operational environment. In others, they may be required to protect the GEF investments from continued losses in certain project activities. This year the Evaluation Office presents an assessment of the levels at which changes in the results framework of a project are made and the drivers for these changes. The discussion in this section is based on the evidence presented in the terminal evaluations submitted in FY 2008.

88. Of the 62 terminal evaluation reports submitted, 23 (37 percent) informed that changes had been made in the results framework during project implementation. Although the changes made vary across projects they may be broadly categorized as changes made in:

- Development objectives and outcomes
- Project outputs
- Outcome indicators

89. The **development objectives and outcomes** of 8 projects were changed during implementation. In three instances, the change in development objectives and outcomes reoriented the scope of the project activities. The drivers for change include exogenous

factors such as shift in political priorities of participating countries⁶ and large scale migration of the intended beneficiary community from the project area⁷; and endogenous factors such as newer problem analysis suggesting another approach to be more effective in enhancing the financial viability of a key project activity.⁸ In two instances the scope of the development objectives was reduced because of slow and inefficient project implementation.⁹ In another three, the original project development objectives were modified because they were not well articulated¹⁰. Therefore, in such cases these changes did not lead to a substantive change in the scope and orientation of objectives and activities of the respective projects.

90. The expected **outputs** of 13 projects were changed during project implementation. Of these, in 5 instances some of the project outputs were dropped due to poor progress made during implementation. In the remainder, original outputs were replaced or modified to reflect the experience and knowledge gained during project implementation. For example for the ‘Capacity Building for Sustainable Land Management (SLM)’ project in Bulgaria (GEF ID 2726, UNDP), some of the original activities were dropped because they had already been undertaken by other organizations by the time project implementation started. The resources saved from the dropped activities were used to undertake a greater number of demonstration activities. The expected outputs of the ‘Methane Gas Capture and Use at a Landfill – Demonstration Project’ in Mexico (GEF ID 784, World Bank) were modified to support installation of 88 additional wells to ensure sufficient supply for the generators. The resultant additional cost of US \$ 0.27 million was met through savings in construction of the methane capture plant.

91. For two projects the outcome indicators were modified. For the ‘Indigenous Management of Protected Areas in the Peruvian Amazon’ (GEF ID 651, World Bank) these were modified to reflect additional activities covered by the project and to also address the measurability and attribution related issues. Similarly, indicators for ‘Conservation and Sustainable Use of the Mesoamerican Barrier Reef’ (GEF ID 837, World Bank) were modified to better account for the activities being undertaken by the project.

92. To summarize, of the 23 projects with reported changes in objectives, outcomes, outputs, and/or indicators, in seven – a relatively smaller number – changes appear to be driven by factors that reduce efficiency in project implementation. Furthermore, for only

⁶ ‘Establishment of a Program for the Consolidation of the Meso-American Biological Corridor’ (GEF ID 243, UNDP)

⁷ ‘Conservation of Biodiversity in Pastaza’ in Ecuador (GEF ID 1301, World Bank)

⁸ ‘Egyptian Engineered Wetlands - Construction of Wetland Project Components’ (GEF ID 395, UNDP)

⁹ ‘Optimization of Biodiversity in Game Ranching Systems’ (GEF ID 359, UNDP) and ‘Conservation Management of Eritrean Coastal, Marine and Island Biodiversity Project’ (GEF ID 411, UNDP).

¹⁰ ‘Management of Indigenous Vegetation for the Rehabilitation of Degraded Rangelands in the Arid Zone of Africa’ (GEF ID 504, UNEP), ‘Caribbean Archipelago Biosphere Reserve: Regional Marine Protected Area (MPA) System’ (GEF ID 773, World Bank), and ‘Conservation and Sustainable Use of Biodiversity in Dibeen Nature Reserve’ (GEF ID 1438, UNDP).

two projects the inefficiencies were considered to be pervasive enough to warrant changes at the development objective and outcome levels.

4.2 Factors associated with lower outcome achievements

93. In earlier APRs lower outcome achievements have been linked with factors that could be broadly classified as weakness in project design, weakness in project implementation, or exogenous changes in a project's operational environment that affect its ability to achieve intended outcomes. This year, the Office presents an assessment of the drivers of lower outcome performance for the completed projects reviewed for OPS4. Of the 210 projects reviewed, the outcome rating of 40 (20 percent) projects was in the unsatisfactory range.¹¹ This discussion pertains to these 40 projects.

94. For 30 projects (75 percent) weakness in project design were reported to be the key driver of low outcome achievements. Twenty six projects had design weaknesses related to problem analysis, choice of activities, implementation and execution arrangements, and project's theory of change.¹² Eleven projects were reported to be overambitious as they allocated inadequate resources in terms of finance and timeframe to the problems being addressed. Seven of these also had other weaknesses related to project design.

95. Of the projects that had lower outcomes due to weaknesses in project design, in seven cases the project theory of change was weak. For these projects even though project components and activities were completed in a timely manner, and the project did not face any exogenous change that could have affected its ability to achieve intended outcomes, the expected outcomes did not materialize because the activities chosen and assumptions made did not eventually lead to the expected outcomes.

96. For 24 projects (60 percent) lower outcome achievements were linked with implementation and execution related problems. These include weak technical capacity of hired staff,¹³ high staff turnover,¹⁴ delays in implementation of critical project activities such as hiring of staff,¹⁵ weak institutional capacity of the chosen executing agency,¹⁶ financial mismanagement and weak oversight,¹⁷ poor project supervision by the

¹¹ The ratings in the unsatisfactory range are Moderately Unsatisfactory, Unsatisfactory and Highly Unsatisfactory.

¹² Example: 'Conservation and Sustainable Use of Biodiversity in the Andes Region' (GEF ID, World Bank)

¹³ Example: Conservation of Biodiversity in the Lake Titicaca Basin (GEF ID 202, UNDP)

¹⁴ Example: Sustainable Management of Mt. Isarog's Territories (GEF ID 798, UNDP)

¹⁵ Example: Improving the Energy Efficiency of the Heat and Hot Water Supply (GEF ID 983, UNDP)

¹⁶ Examples: Conserving Mountain Biodiversity in Southern Lesotho (GEF ID 245, UNDP); Improving the Energy Efficiency of the Heat and Hot Water Supply (GEF ID 983, UNDP)

¹⁷ Examples: Conservation of Biodiversity through Integrated Collaborative Management in Rekawa, Ussangoda, and Kalametiya Coastal Ecosystems (GEF ID 802, UNDP); Community-based Coastal and Marine Conservation in the Milne Bay Province (GEF ID 1261, UNDP); Barriers and Best practices in Integrated Management of Mountain Ecosystems (GEF ID 1328, UNEP); Dry Forest Biodiversity Conservation (GEF ID 815, World Bank).

implementing agency,¹⁸ etc. Due to these problems the project activities were either not complete at the time of project closure or had been completed after considerable delays¹⁹ leading to lower outcome achievements. Of the projects for which problems related to project implementation and execution were reported, for 15 of them problems related to project design were also reported.

97. For four projects (10 percent) lower outcome achievements were linked with the exogenous factors that were beyond the control of GEF partnership involved in project implementation. For three projects outcome achievements were reported to be lower because of the political instability and civil strife in the project area.²⁰ Activities of one project had to be curtailed because of a natural disaster.²¹

4.3 Materialization of Cofinancing

98. The Office reports on materialization of cofinancing in completed projects every year based on the information presented in the terminal evaluation reports. However, many key issues such as the level of global environmental benefits that accrue from cofinancing, the processes through which these benefits accrue, and the costs incurred by the GEF partnership in mobilizing cofinancing, have not been assessed in detail so far. These issues are presently being assessed by the Evaluation Office and will be reported on in the OPS4 report.

99. The analysis presented in this section is based on the information available on 285 completed projects for which terminal evaluations were submitted since FY 2002. Terminal evaluations for 75 of these projects had been submitted during or before FY 2004; the remainder (210 projects) comprise the OPS4 cohort. Information on materialization of cofinancing is available for 48 (64 percent) projects from the pre-OPS4 period and for 162 (77 percent) from the OPS4 period. Among the projects reviewed all but one project had at least some cofinancing promised and for all but three projects some cofinancing was reported to have materialized.

100. For the projects from the pre-OPS4 period, for every dollar of approved GEF grant on average cofinancing of US \$ 5.9 was promised and US \$ 5.4 was reported to have materialized. In comparison, for the projects from the OPS4 period, for a dollar of approved GEF grant, on average, cofinancing of US \$ 2.1 was promised and US \$ 2.0 was reported to have materialized (table 4.1).

¹⁸ Example: Dry Forest Biodiversity Conservation (GEF ID 815, World Bank); Optimizing Biological Diversity within Wildlife Ranching systems; A Pilot Demonstration in A Semi-arid Zone (GEF ID 359, UNDP)

¹⁹ Relevant completed activities were time critical and where delays could have affected achievement of intended outcomes.

²⁰ Examples: Aceh Elephant Landscape Project (GEF ID 26, World Bank); West Africa Pilot Community-Based Natural Resource Management Project (GEF ID 55, World Bank); Forestry and Conservation Project (GEF ID 513, World bank).

²¹ Dry Forest Biodiversity Conservation (GEF ID 815, World Bank)

101. Although on average a higher amount of cofinancing was promised and was reported to have materialized for the projects from the pre-OPS4 period, the rate of materialization (cofinancing materialized vis-à-vis promised) is better for the OPS4 projects (98 percent compared to 92 percent). However, both average amount of cofinancing per dollar of approved GEF grant and rate of materialization of cofinancing tend to get skewed by a few projects. For example, two projects ‘Sichuan Gas Transmission and Distribution Rehabilitation’ in China (GEF ID 75, World Bank) and ‘Renewable Resources Development Project’ in India (GEF ID 76, World Bank) from the pre-OPS4 cohort, account for 77 percent of the cofinancing raised by the projects pertaining to that cohort. If these projects are dropped from consideration, the average materialized cofinancing for the cohort drops from US \$ 5.4 to US \$ 1.4 per dollar of GEF grant. In comparison, figures for the OPS4 cohort are not as sensitive: when the top two projects in terms of cofinancing are dropped from consideration, the average materialized cofinancing drops from US \$ 2.0 to US \$ 1.7 per dollar of GEF grant.

Table 4.1 Materialization of Cofinancing

Period	FY 2002 to FY 2004	FY 2005 to FY 2008 (OPS 4)	FY 2008	All Years
TE reports submitted	75	210	62	285
Cofinancing data available for	48	162	53	210
Approved GEF grant per project in m \$	5.9	4.4	3.9	4.7
Actual GEF grant per project in m \$	5.0	4.1	3.7	4.3
Promised cofinancing per project in m \$	35.0	9.0	7.4	14.9
Promised cofinancing per dollar of approved GEF grant	5.9	2.1	1.9	3.2
Reported materialized cofinance per project in m \$	32.0	8.8	8.5	14.1
Reported materialized cofinance per dollar approved of GEF grant	5.4	2.0	2.2	3.0
Materialized cofinance per dollar of promised cofinance (%)	92	98	114	95

102. Compared to other agencies, projects implemented by the World Bank had a greater amount of cofinancing promised and a greater amount of cofinancing materialized per dollar of GEF approved grant (table 4.2). However, for the OPS 4 cohort the difference among agencies on this count has reduced.

103. Overall within the OPS4 cohort, the projects with a lesser proportion of promised cofinancing materialized have a lower probability of being rated in the satisfactory range. For example, of projects where less than 80 percent of promised cofinancing materialized outcomes of 73 percent were rated in the satisfactory range. In comparison among the projects with 80 percent or more realization of promised cofinancing outcomes of 83 percent had been rated in the satisfactory range. The assessment on materialization of cofinancing vis-à-vis outcome ratings is, however, affected by reverse causality. As reported previously in *APR 2007*, inability of projects to show progress in earlier stages of their implementation is one of the major reasons that lead to withdrawal of support by

non-GEF donors.²² In some instances, it is reported that the expected project outcomes were achieved despite cofinancing commitments not being fully met.²³ The overall direction of causality remains unclear because other pathways, which demonstrate negative effect of lower materialization of cofinancing on project outcomes, have also been reported. For example, in instances where cofinancing and the activities supported through cofinancing were well integrated in the project design, non-materialization of a significant proportion of promised cofinancing²⁴ or delays in materialization of cofinancing²⁵ led to lower outcome achievements or stalled project progress.²⁶ These issues will be explored in greater detail in the OPS4 report.

Table 4.2 Materialization of Cofinancing: By Agency²⁷

Period	World Bank	UNDP	UNEP
Pre OPS4 Period (FY 2002 to FY 2004)			
Number of projects for which Cofinancing data is available	31	11	6
Promised cofinancing per dollar of approved GEF grant	6.5	3.6	1.1
Reported materialized cofinance per dollar approved of GEF grant	6.0	2.8	1.0
Materialized cofinance per dollar of promised cofinance	93 %	78%	87%
OPS 4 Period (FY 2005 to FY 2008)			
Number of projects for which Cofinancing data is available	80	59	23
Promised cofinancing per dollar of approved GEF grant	2.4	1.6	1.5
Reported materialized cofinance per dollar approved of GEF grant	2.3	1.6	1.5
Materialized cofinance per dollar of promised cofinance (%)	98%	99%	101%

4.4 Delays in Project Completion

104. The Evaluation Office tracks the time difference between expected closing at project start and actual closing to know the extent projects supported by GEF are being completed in a timely manner. Important causes of delay, as reported in *APR 2007*, include internal factors such as overly optimistic project design, communication problems among project partners, delays in transfer of money to the partners, cumbersome implementation procedures, and delays in hiring key staff; and external

²² Examples: Renewable Energy and Forest Conservation project in Nicaragua (GEF ID 847, World Bank); Pilot Production and Commercial Dissemination of Solar Cooker project in South Africa (GEF ID 1311, UNDP); Solar Development Group Project (GEF ID 595, World Bank).

²³ Example: ‘Capacity Building for GHG Emission Reduction through Energy Efficiency’ in Romania (GEF ID 284, UNDP).

²⁴ Examples: Conservation of Globally Significant Forest Ecosystems in Suriname’s Guayana Shield (GEF ID 661, UNDP) , Conservation of Biodiversity through Integrated Collaborative Management in Rekawa, Ussangoda, and Kalametiya Coastal Ecosystems (GEF ID 802, UNDP).

²⁵ Example: Management of Indigenous Vegetation for the Rehabilitation of Degraded Rangelands in the Arid Zone of Africa (GEF ID 504, UNEP).

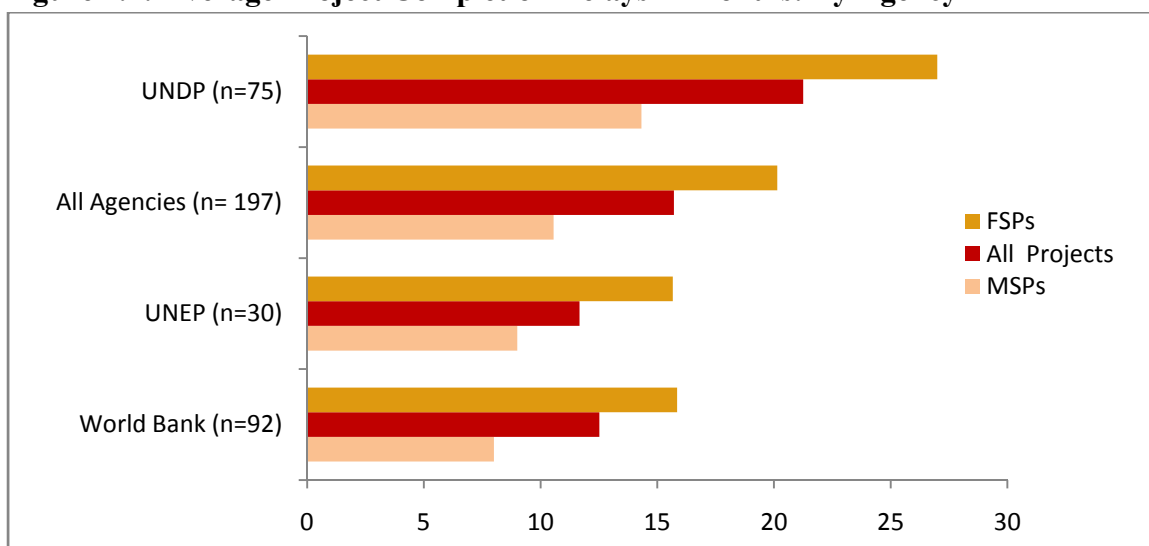
²⁶ Examples: ‘Coastal and Marine Biodiversity Management Project’ in Mozambique (GEF ID 648, World Bank); Biodiversity Conservation in the Azov-Black Sea Ecological Corridor (GEF ID 412, World Bank).

²⁷ Joint projects have been attributed to the lead implementing agency.

causes such as unanticipated political instability, changes in project processing requirements on the part of the host government, and financial crisis in the host country.

105. An assessment of the relationship between delays in completion of projects and their respective outcome ratings for the OPS4 projects shows little correlation. Similar results have been reported in *APR 2006*. Nonetheless, the evidence from the reviewed terminal evaluations shows that in some instances extension of project completion date facilitates achievement of the expected project outcomes.²⁸ In such cases insistence on timely completion would mean that some activities are either not completed or completed in haste, thus resulting in lower outcome achievements.

Figure 4.1: Average Project Completion Delays in months: By Agency



106. Although relationship between delays in project completion and project outcome ratings remains unclear, the evidence from terminal evaluations shows that delays often indicate management inefficiency. Extended duration of implementation often translates into higher management costs. Consequently, the project’s management may down scale scope of some activities²⁹ or may look for additional funding to meet the resultant shortfall in funding.³⁰

Table 4.3: Distribution of Projects: Completion Delay by Agency

Parameter	Percentage of Projects completed within one year of expected completion date			Percentage of Projects completed after a delay of more than two years		
	MSPs	FSPs	All Projects	MSPs	FSPs	All Projects
UNDP	53	39	45	12	49	32
UNEP	67	58	63	6	17	10
World Bank	79	57	66	8	26	18
All Agencies	67	50	58	9	34	22

²⁸ ‘Environmental Protection of the Rio de la Plata and Its Maritime Front’ (GEF ID 613, UNDP)

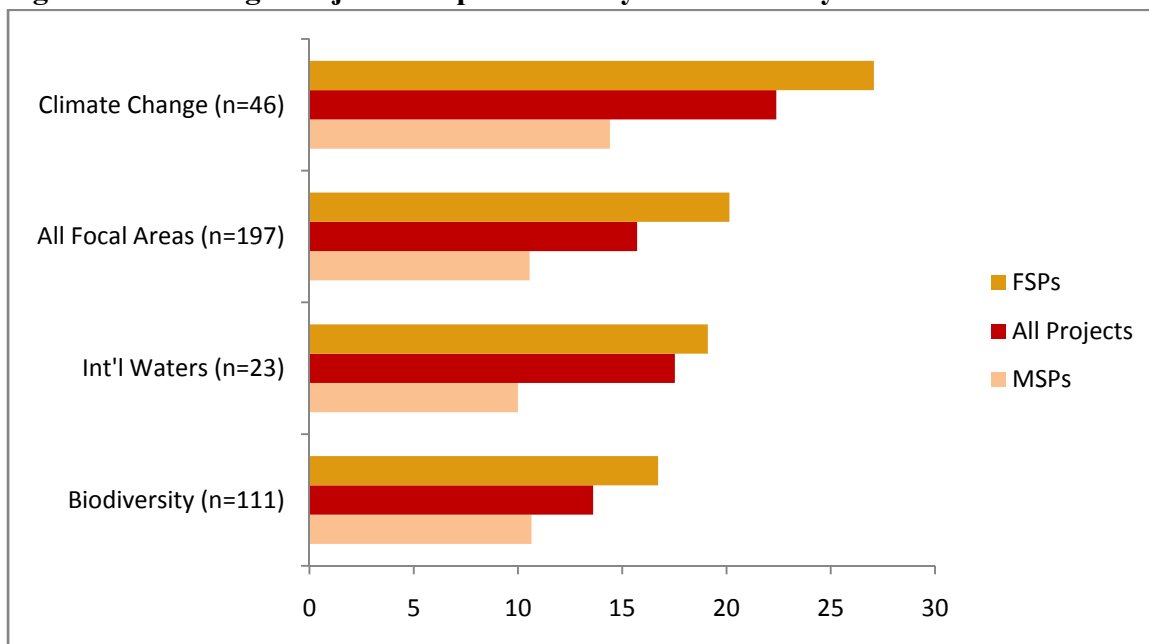
²⁹ ‘Coastal and Marine Biodiversity Management’ in Mozambique (GEF ID 648, World Bank)

³⁰ ‘Establishment of the Nuratau-Kyzylkum Biosphere Reserve as a Model for Biodiversity Conservation’ in Uzbekistan (GEF ID 855, UNDP)

107. The average delay in project completion for the OPS4 projects was 16 months (figure 4.1 and 4.2). On average the FSPs were completed with a 20 months delay whereas the MSPs were completed with a significantly smaller delay of 11 months. Overall 58 percent of the projects were completed within a year of their expected completion date at the point of project start, whereas 22 percent were completed after a delay of more than two years (table 4.3). Consistent with the level of complexities involved, a greater proportion of FSPs are completed with a delay of more than two years than MSPs (table 4.3).

108. On average, the World Bank implemented projects were completed with a delay of 13 months (FSPs 16 months and MSPs 8 months). The completion delays for World Bank implemented projects are substantially lower than that for projects implemented by other agencies.³¹ Projects implemented by UNDP are completed with a significantly greater time lag of 21 months (FSPs 27 months and MSPs 13 months). While average delay for UNEP is lower than for other agencies the difference is not statistically significant.

Figure 4.2: Average Project Completion Delay in months: By Focal Area



³¹ The difference is significant when other factors such as geographical scope, geographical region, project size and focal area are controlled for.

109. Among the focal areas, on average the projects pertaining to the biodiversity focal area were completed with significantly smaller delays, whereas those from the climate change focal area had significantly greater time lags.

110. Within the OPS4 cohort, the average project completion delay was 19 months for the FY 2005 cohort, 13 months for the FYs 2006 and 2007 cohorts and 18 months for the FY 2008 cohort. In addition, 44 percent of the projects in the FY 2005 cohort, 17 percent of the projects in the FY 2006 cohort, 14 percent of the FY 2007 cohort, and 26 percent of the FY 2008 cohort had delays of two years or more.

5. Quality of Project Monitoring

111. A project's M&E system provides early information on progress towards achievement of its intended results. It also helps in identification of issues that warrant corrective measures in order to facilitate progress. The Council approved *Monitoring and Evaluation Policy* (2006) lays minimum requirements for M&E relevant to different stages of the project life cycle. The requirements relevant to the design stage form a basis for assessment of quality of M&E arrangements of projects at CEO approval or endorsement, whereas those relevant to project implementation and completion stages form a basis for assessment of quality of M&E implementation in completed projects.

112. The assessment on 'Quality of M&E Arrangements at Entry' is a follow up to a similar assessment presented in the *APR 2005*. The office rated M&E arrangements of 76 percent of FSPs that were CEO Endorsed in FY 2008 to be in compliance with the minimum requirements applicable in FY2008. This is a significant improvement over the FY 2005 performance, when 58 percent of the projects were rated to be compliant with the requirements applicable in FY 2005. If the requirements applicable in FY 2005 are also applied to the FY 2008 cohort, the percentage of projects of FY 2008 cohort in compliance increases to 80 percent.

113. Sixty seven percent of the OPS4 projects were rated in the satisfactory range on quality of M&E implementation. For FY 2008, 72 percent were rated in the satisfactory range.

5.1 Rating Approach

114. To assess the quality of M&E arrangements at Entry, the Office reviewed the M&E plans presented in the project appraisal documents of all the FSPs that were endorsed by the CEO in FY 2005 and FY2008. In all, M&E plans of 156 FSPs, including 74 from FY 2005 and 82 from FY 2008, were reviewed. The results for the FY 2005 cohort have already been presented in *APR 2005*. The M&E plans of the additional projects, i.e. the FY 2008 cohort, were reviewed following an identical methodology. The M&E plans were assessed on 13 parameters that have been classified as either "critical" – where non compliance indicates serious deficiencies in M&E arrangements – or "other" (see 'Annex C' for detailed descriptions). To be in compliance with the GEF M&E expectations at entry, a project needs to be in compliance with all the critical parameters

and needs to perform sufficiently well on all the parameters together. To be classified as compliant, projects were required to score at least a 2 (on a scale of 1 to 3, with 3 being the highest) on each of the critical parameters and to have an aggregate score of 26 out of a maximum of 39. Here it should be noted that for the parameter on baseline information – a critical parameter – the GEF requirements have changed. In FY 2005 to be in compliance with the GEF M&E arrangements at entry requirements for baseline information a promise to conduct baseline survey in the first year of implementation was sufficient. Now, other than rare situations where an exception may be made to allow the implementing agency to conduct a baseline survey within a year of start of project start, baseline information is required at the point of CEO Endorsement. To account for this change the minimum level of expected performance on this parameter has been adjusted for the FY 2008 cohort.

115. Quality of project monitoring in completed projects was assessed on a six point scale. All the 252 terminal evaluation reports submitted since FY 2004 were considered. These include 42 reports submitted in FY 2004 that are not part of the OPS4 cohort. Of these reviewed terminal evaluation reports 191 provided sufficient information to rate quality of monitoring during implementation. To rate quality of monitoring in completed projects, the Office assesses whether:

- an M&E system was in place and facilitated timely tracking of results and progress toward project objectives by collecting information on chosen indicators continually throughout the project implementation period
- annual project reports were complete and accurate, with well-justified ratings
- the information provided by the M&E system was used for project management
- the parties responsible for M&E activities were properly trained to ensure that correct procedures are followed and quality is maintained in data collection

5.2 Quality of M&E at Entry

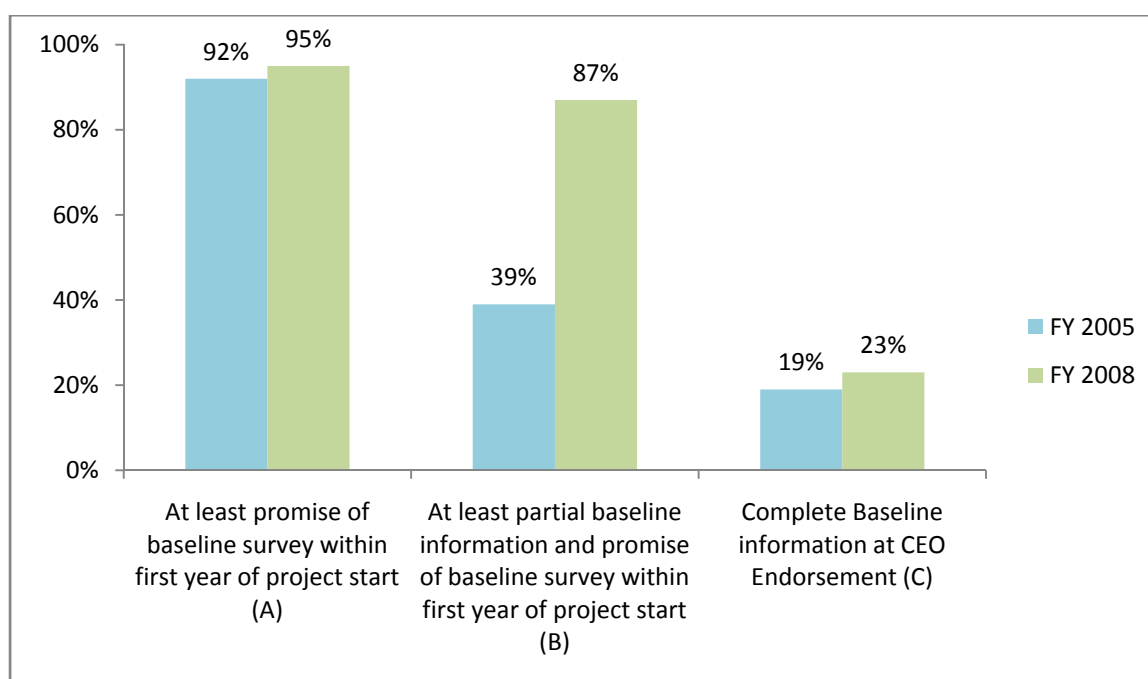
116. The office rated M&E arrangements of 76 percent of FSPs that were CEO Endorsed in FY 2008 to be in compliance with the minimum requirements applicable in FY2008. This is a significant improvement over the FY 2005 performance, when 58 percent of the projects were rated to be compliant with the requirements applicable in FY 2005. If the requirements applicable in FY 2005 are also applied to the FY 2008 cohort, then the percentage of projects in compliance increases to 80 percent (table 5.1). A greater proportion of projects to be implemented by the World Bank and UNDP and those pertaining to the Biodiversity and Climate Change focal area were rated to be in compliance in FY 2008 than in FY 2005. For other agencies and focal areas similar conclusions may not be drawn. However, for most, the direction of change is consistent with the overall trend.

Table 5.1: Quality of M&E at Entry: Compliance with M&E requirements

Year	FY 2005		FY 2008		
	Projects Reviewed	Percentage of Compliant projects (2005 Standards)	Projects Reviewed	Percentage of Compliant projects (2005 Standards)	Percentage of Compliant projects (2008 Standards)
GEF Portfolio	74	58%	82	80%	76%
By Agency					
UNDP	25	68%	36	83%	83%
UNEP	5	40%	5	60%	20%
World Bank	30	50%	25	80%	80%
Joint/Others	14	64%	16	81%	69%
By Focal Area					
Biodiversity	28	50%	21	95%	95%
Climate Change	21	76%	25	88%	88%
International Waters	11	55%	10	70%	70%
Land Degradation	3	33%	10	80%	67%
Multi Focal Area	8	50%	7	71%	57%
Persistent Pollutants	2	50%	10	50%	33%

117. Among the parameters on which compliance was assessed, M&E plans of a significantly larger percentage of projects specify at least one indicator for each expected outcome (95 percent of FY 2008 versus 57 percent of FY 2005) and for a greater percentage the specified indicators were assessed to be sufficient or largely sufficient to allow measurement of progress towards achievement of expected outcomes (94 percent of FY 2008 versus 76 percent of FY 2005). Although compared to FY 2005 (95 percent) a lower percentage of projects from FY 2008 (85 percent) provide output targets in M&E plans. This is in line with the expectations because shift to results based management has led to increased focus on outcome indicators vis-à-vis output targets.

Figure 5.1: Percentage of projects in compliance with different levels of baseline information requirements



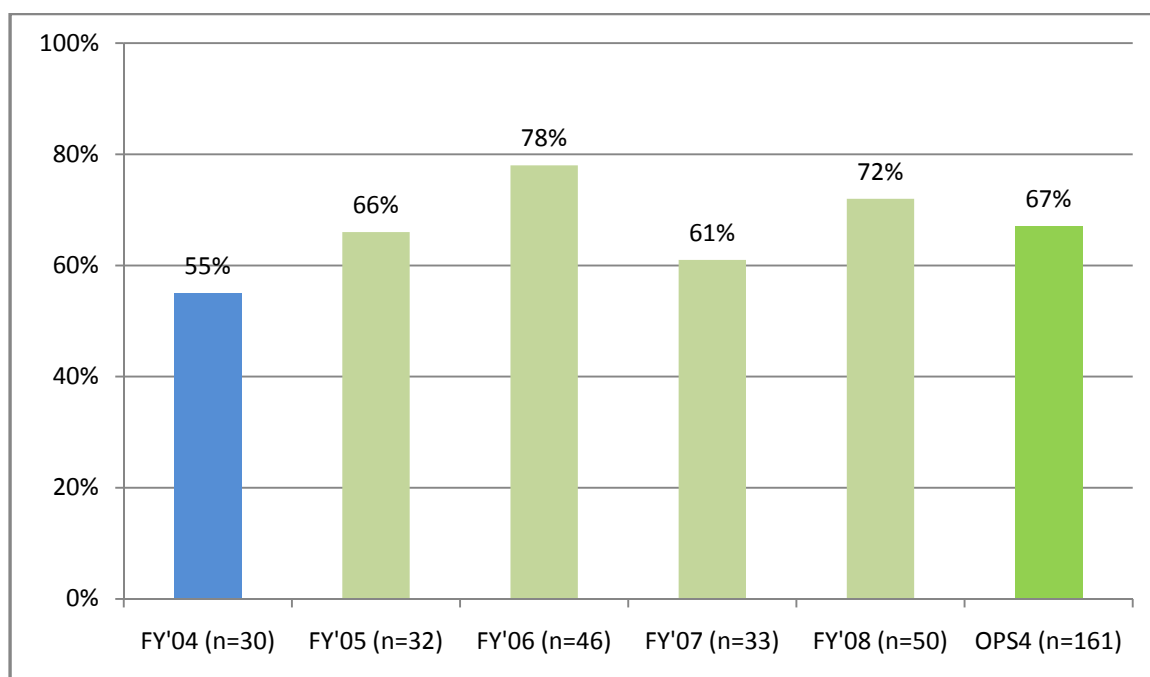
118. Figure 5.1 presents the percentage of projects in compliance at three different levels of baseline information requirements. At the time the first M&E quality at entry assessment was carried out (FY 2005), the GEF M&E requirements for baseline information stipulated that at the point of CEO Endorsement a project should provide baseline information or should commit to conducting a baseline survey within a year of project start. When projects are assessed on this criterion then 95 percent of those from FY 2008 and 92 percent of those from FY 2005 met this requirement (A). The difference between the two cohorts increases when they are assessed based on the requirement that at least partial baseline information on important indicators along with a promise of baseline survey in the first year of project start is provided (B)³². The M&E policy of 2006, however, requires projects to provide baseline information at the point of CEO Endorsement except in those rare cases when such is not possible. If this more stringent interpretation was applied, assuming those rare cases were absent from the pool of projects that were reviewed, only 23 percent of the projects from FY 2008 were in compliance (C). Thus, this figure encapsulates both: the improvements made by the system in providing baseline information at the CEO Endorsement (B), and the difficulties in achieving the most desirable state of performance (C).

5.3 Quality of monitoring during Implementation

119. It is difficult to make robust conclusions on performance among agencies and focal areas because a significant proportion of the terminal evaluations (23 percent of OPS4 projects) did not provide sufficient information to allow the Evaluation Office to rate performance on this parameter. Among those that did provide this information, 67 percent – including 72 percent in FY 2008 – were rated moderately satisfactory or above in terms of quality of monitoring during implementation (figure 5.2). The proportion of projects from FY 2004 that were rated in this range was a significantly lower 55 percent.

³² Requirement “B” has been taken as a standard to assess compliance for the FY 2008 cohort.

Figure 5.2: Percentage of Projects with M&E rated MS or above



120. As also discussed in *APR 2006* and *APR 2007*, the projects that have a weak M&E plan at entry are unlikely to be able to make sufficient corrections to improve their quality of project monitoring during implementation. Among the completed projects from OPS4 cohort that were rated both on quality of monitoring during implementation and on quality of M&E at entry, only 39 percent of those rated in the unsatisfactory range on quality at entry were rated in the satisfactory range for quality of M&E during implementation. In contrast, of those that were rated in the satisfactory range on quality of M&E at entry, 82 percent were also rated in the satisfactory range during implementation. Thus, quality at entry seems to be an important driver of the eventual quality of monitoring during implementation.

6. Quality of Terminal Evaluations

121. Terminal evaluations provide an assessment of project accomplishments and shortcomings. They form the building blocks for the assessment of performance of completed projects presented in the APR. The effectiveness of the terminal evaluations as a learning tool for the GEF partnership may be compromised if the information provided by them is inaccurate, incomplete, and biased. The Evaluation Office reviews terminal evaluations to provide verified ratings on project performance and on the quality of terminal evaluation reports. Through the process of assessing quality of the terminal evaluation reports, the Evaluation Office also identifies the areas where the reports could be improved.

122. To date, 285 terminal evaluation reports have been submitted to the GEF Evaluation Office. This is the fifth year the Evaluation Office has rated the quality of project terminal evaluation reports; 252 have been rated thus far including the 210 that have been considered for OPS4. In FY 2008, 62 terminal evaluation reports were submitted by the GEF Agencies.

123. Although, the incidence of delays in submission, or non submission of terminal evaluation reports, has decreased many reports continue to be submitted after considerable delays. The Office is working with the GEF Secretariat and GEF Agencies so that timely submission of terminal evaluation reports for completed projects can be reliably tracked.

124. Overall, the quality of 89 percent of the terminal evaluation reports considered for OPS4 including 92 percent of those submitted in FY 2008 was rated moderately satisfactory or above. As also mentioned in the earlier APRs (FY 2005, FY 2006 and FY 2007), the overall quality of terminal evaluations has improved substantially from the baseline of FY 2004, when only 69 percent of the terminal evaluations were rated moderately satisfactory or above on quality.

125. Up to FY 2006, the ratings verified by the Office had been based on desk review of the terminal evaluation reports, which limited the reliability of these reviews. Since FY 2007 the Office is piloting a methodology for field verification of terminal evaluations to more rigorously confirm the results of completed projects. Since field verification is time consuming and costly, such verifications are being undertaken only for a sample drawn from the completed FSPs for which terminal evaluation reports have been submitted to the Office. So far seven field verifications have been completed. In addition, at least nine more field verification will be conducted for inclusion in OPS4.

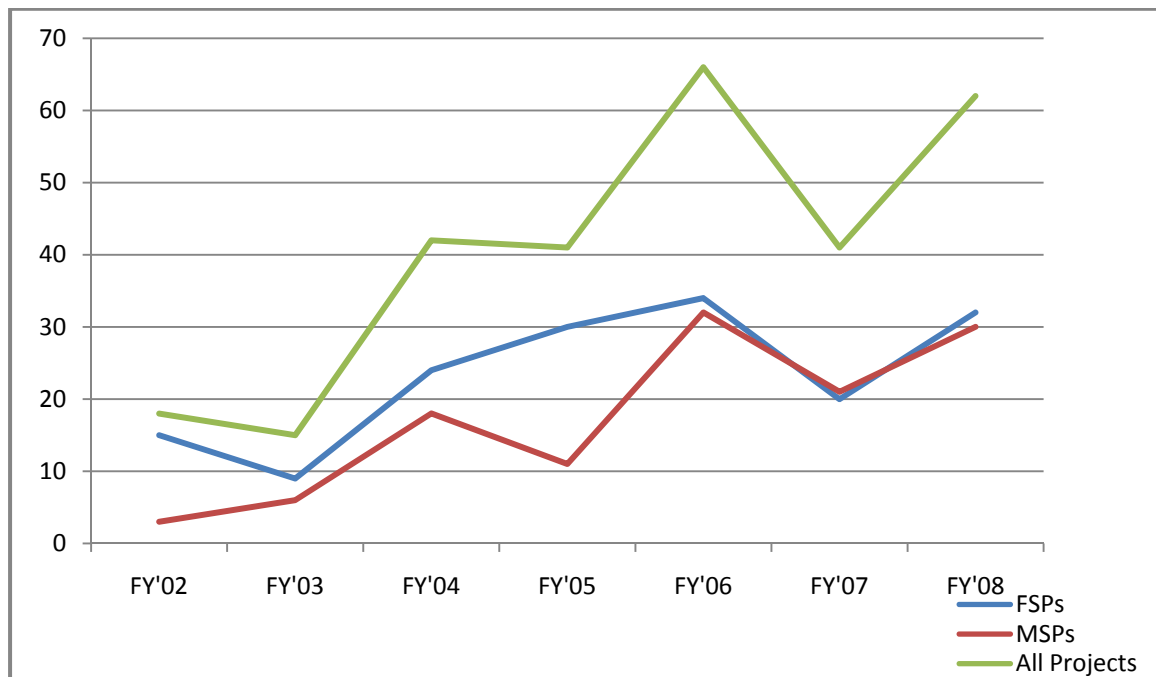
6.1 Rating Approach

126. The assessment approach adopted for the terminal evaluation reports submitted from FY 2005 to FY 2008 has remained the same; those submitted in FY 2004 were assessed using a slightly different but comparable methodology. The reports submitted since FY 2005 have been assessed based on the following criteria:

- Did the report present an assessment of relevant outcomes and achievement of project objectives in the context of the focal area program indicators, if applicable?
- Was the report consistent, the evidence complete and convincing, and the ratings substantiated when used?
- Did the report present a sound assessment of sustainability of outcomes?
- Were the lessons and recommendations supported by the evidence presented?
- Did the report include the actual project costs (total and per activity) and actual cofinancing used?
- Did the report include an assessment of the quality of the project M&E system and its use in project management?

127. Performance on each of these criteria is rated on a six-point scale. The overall rating is a weighted average of these ratings: the first two criteria are given a weight of 0.3 each, and the remainder a weight of 0.1 each.

Figure 6.1: Number of Terminal Evaluation Reports Submitted



128. The Evaluation Office also tracks consistency between the verified outcome rating provided by it and those provided in the last project implementation report (PIR) that was submitted to the GEF Secretariat, in the terminal evaluation report, and by the evaluation offices of the implementing agencies. Since not all ratings have been provided on a six point scale used by the Office, to make comparisons possible ratings are converted to a binary scale.

129. To field verify findings of terminal evaluations, the Office gathers information on project performance through:

- interviews of key stakeholders
- verification of the physical evidence of the reported accomplishments
- desk review of the project related literature including, but not restricted to, terminal evaluation reports, audits, progress reports, and other reviews.

130. The rating approach followed by the Evaluation Office for piloting field verifications is identical to the rating approach followed for terminal evaluation reviews. In addition, elements of the other approaches being adopted by the Evaluation Office to assess impacts of GEF activities are being incorporated in the field verification approach.

6.2 Findings

131. As noted, of the 210 terminal evaluation reports submitted for review during the period considered for OPS4, **the quality of 89 percent (187 projects) was rated moderately satisfactory or above** (see table 6.1). The terminal evaluation reports of 92 percent of FSPs and 85 percent of MSPs were so rated.

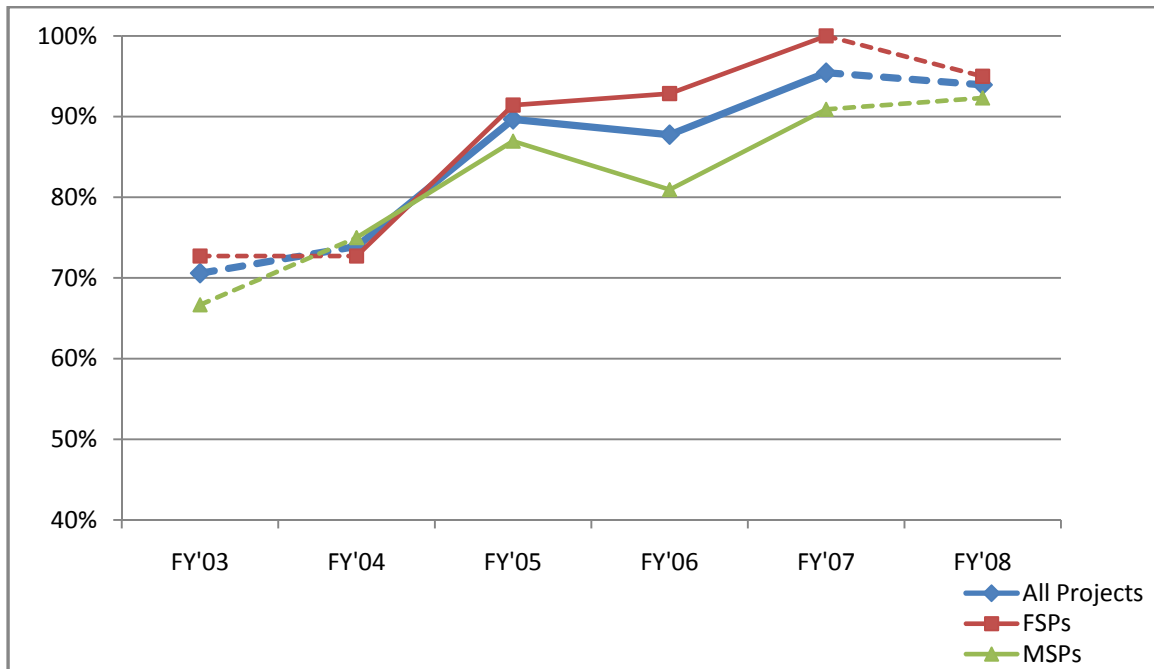
132. By way of comparison, the proportion of reports rated moderately satisfactory or above was 92 percent in FY 2008, 95 percent in FY 2007, 84 percent in FY 2006, 88 percent in FY 2005. For the terminal evaluations that were submitted in FY 2004 and, therefore do not pertain to the OPS4 cohort, only 69 percent had been so rated.

Table 6.1: Quality of Terminal Evaluations (OPS4) – Percentage rated MS or Above

Period	FY 2005 to 2006			FY 2007 to 2008			OPS4 (FY 2005 to 2008)		
	FSPs	MSPs	All Projects	FSPs	MSPs	All Projects	FSPs	MSPs	All Projects
UNDP	92%	83%	89%	96%	91%	94%	94%	89%	92%
UNEP	63%	57%	60%	100%	100%	100%	75%	83%	80%
World Bank	93%	83%	89%	96%	82%	90%	94%	83%	89%
All Agencies	89%	79%	85%	96%	90%	93%	92%	85%	89%

133. Table 6.1 presents the percentage of reports (pertaining to the OPS4) submitted by the Implementing Agencies that were rated moderately satisfactory or above in terms of quality. For the OPS4 period, there is little difference among agencies in terms of percentage of terminal evaluation reports rated moderately satisfactory or above on quality. However, within agencies, for the World Bank the percentage of terminal evaluation for MSPs that rated moderately satisfactory or above in quality is lower than for its FSPs. This is reflective of the lower attention that World Bank accords to the terminal evaluations for MSPs. When the OPS4 period is split into two halves – first half (FY 2005 to FY 2006) and second half (FY 2007 to FY 2008) – the improvements made in quality of terminal evaluations across agencies can be assessed. Compared to the first half the terminal evaluations submitted in the second half were more likely to be rated moderately satisfactory or above. Among the agencies, across these two periods there have been significant improvements in the quality of terminal evaluations submitted by UNEP. Gains made by UNEP are linked to the transfer of responsibility for conducting terminal evaluations to its evaluation office and a closer tracking of the quality of evaluations.

Figure 6.2: Terminal evaluations with MS or above quality ratings (based on year of terminal evaluation report completion)

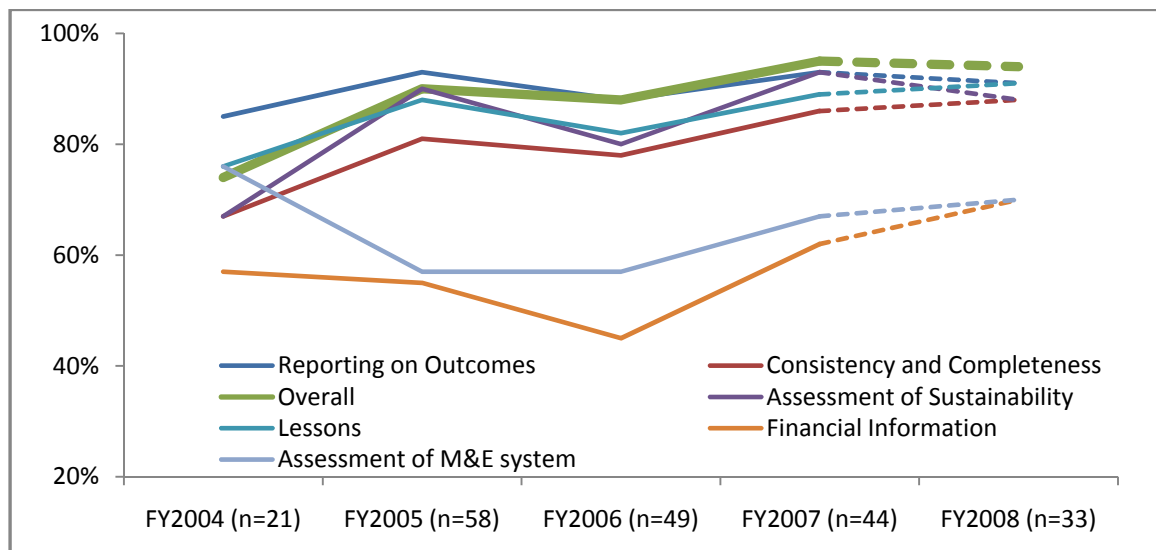


134. To assess the underlying trends in change in quality of terminal evaluation reports, the Evaluation Office has, since FY 2006, compared ratings for quality of terminal evaluation reports based on the year of report completion rather than the year of report submission. However, efficacy of using year of completion as a basis for comparison is limited by time lag in submission of terminal evaluation reports. Since all reports for some cohorts—especially the most recent one—have not yet been submitted, ratings for as-yet-unsubmitted reports are not known. To address this weakness, in Figure 6.2 – which presents trends in quality of terminal evaluation reports based on year of completion – the periods for which a significant proportion of terminal evaluations have not been assessed are represented by dotted lines. The findings confirm that the quality of terminal evaluations has indeed improved.

6.2.1 Performance by Quality Dimension

135. As discussed in section 6.1 on the rating approach, to arrive at the overall rating on the quality of a terminal evaluation report, the report is rated on six quality dimensions. Figure 6.3 shows the trends in percentage of terminal evaluation reports rated moderately satisfactory or above on individual quality dimensions. Despite improvements in overall quality of terminal evaluation reports, improvements in performance on various quality dimensions have been uneven. The ratings on dimensions such as reporting on project outcomes, sustainability of outcomes, quality of lessons, and consistency and completeness of reporting continue to be high, whereas the ratings on reporting on M&E and on financial information about the project are considerably lower.

Figure 6.3: Percentage with MS or above ratings on quality dimensions (based on Year of terminal evaluation report completion)



136. Among the Implementing Agencies, a greater percentage of the terminal evaluation reports submitted by World Bank are rated moderately satisfactory or above on quality of financial information. In recent years, reports submitted by UNEP also show marked improvement in quality of financial information provided.

6.2.2 Comparison of Ratings

137. The Office compares its verified project outcome ratings, with those provided in:

- the reviews conducted by the evaluation offices of implementing agencies
- the terminal evaluation reports
- the last PIR submitted by the implementing agency before project completion

138. Despite recent attempts by the agencies to harmonize the rating scales, the scale used to rate outcome achievements (or progress towards achieving expected outcomes in case of PIRs) for a significant number of OPS4 projects are not directly comparable to those used by the GEF Evaluation Office. Therefore, in order to make the ratings comparable all the ratings were converted into a binary scale.³³

139. Disconnects between the outcome ratings given by the GEF Evaluation Office and by others could potentially be in two directions: positive – when GEF Evaluation Office rates outcome achievements of a project to be satisfactory but the others rate it to be unsatisfactory; and, negative – when the Evaluation Office rates it to be unsatisfactory but the others rate it to be satisfactory. For the OPS4 projects there was no positive disconnect between the outcome ratings provided by the Evaluation Office and by others.

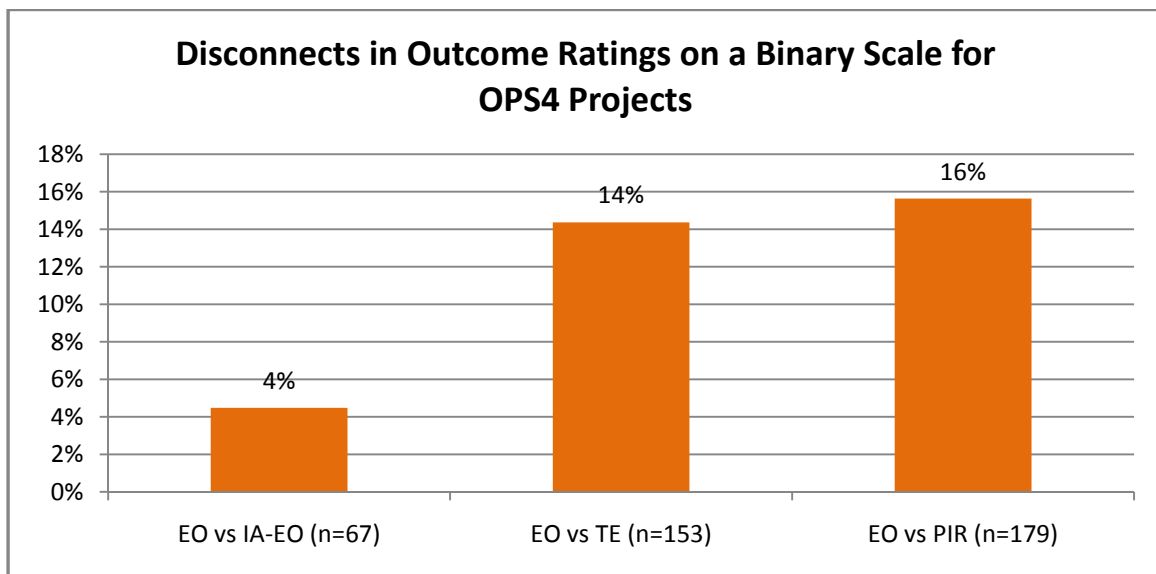
³³ The moderately satisfactory or above on the GEF EO scale becomes “satisfactory” and moderately unsatisfactory or below becomes “unsatisfactory.”

However, several negative disconnects – others rated it in the satisfactory range but the Evaluation Office rated it in the unsatisfactory range – were observed.

140. Among the implementing agencies, the evaluation office of the World Bank provides ratings on outcome achievements of all FSPs – it does not provide such ratings for MSPs. The evaluation office of UNEP has started providing ratings on outcome achievements since FY 2007 for both FSPs and MSPs. The evaluation office of UNDP started rating quality of terminal evaluation reports in FY 2007. However, it is yet to start providing outcome achievement ratings. Overall, whenever provided, the outcome ratings given by the evaluation offices of the Implementing Agencies are consistent with those provided by the GEF Evaluation Office. Negative disconnects were found for only 4 percent of the projects. However, the ratings given in the terminal evaluations and those given in the last PIRs show negative disconnects for a comparatively higher 14 percent and 16 percent of projects, respectively (figure 6.4).

141. The extent of disconnects between the Evaluation Office outcome ratings and those provided in the PIRs are masked by the ‘missing’ PIR ratings. For example of the 175 projects for which last PIR submitted included a rating on progress towards achievement of outcomes, the Evaluation Office rated outcome achievements of only 16 percent in the unsatisfactory range. However, for the 26 projects for which the last PIR had not included such ratings, it rated a comparatively higher 31 percent in the unsatisfactory range. Such a bias was not observed for the ratings provided in the terminal evaluations and in the ratings provided by the evaluation offices of the Implementing Agencies. Thus, this underscores the inability of the PIRs to report with candor on progress towards achievement of project outcomes and objectives.

Figure 6.4: Negative disconnects between the ratings by GEF EO and others



6.2.3 Piloting of Field Verification Approach

142. As mentioned before, in FY 2007 the Evaluation Office initiated piloting of an approach to field verification of terminal evaluations. The experience gained through this process will be used to develop the field verification approach further. So far seven terminal evaluations have been field verified. The observations are still too few to draw conclusions. In addition to the seven field verifications conducted so far, at least nine more will be conducted for inclusion in OPS4.

Chapter 7: Management Action Records

143. The Management Action Records (MAR) keep track of the level of adoption of Council's decisions on the basis of evaluations findings and recommendations. It also seeks to increase the accountability of GEF management regarding Council decisions on monitoring and evaluation issues. The GEF Council approved the format and procedures for the GEF MAR at its November 2005 meeting and requested the GEF Evaluation Office prepare updated MAR to be presented to the Council for review and follow up on an annual basis.

7.1 Rating Approach

144. The rating categories for the progress of adoption of Council decisions were agreed upon in the consultative process of the Evaluation Office with the GEF Secretariat and the GEF Agencies and are as follows:

- **High:** Fully adopted and fully incorporated into policy, strategy or operations.
- **Substantial:** Decision largely adopted but not fully incorporated into policy, strategy or operations as yet.
- **Medium:** Adopted in some operational and policy work, but not to a significant degree in key areas.
- **Negligible:** No evidence or plan for adoption, or plan and actions for adoption are in a very preliminary stage.
- **N/A:** Non-applicable
- **Not possible to verify yet:** verification will have to wait until more data is available or proposals have been further developed.

145. This year's MAR present ratings of GEF management and the verification of these ratings by the Evaluation Office. They track management actions on Council decisions based on twelve GEF Evaluation Office documents. These include nine evaluations that were already presented in last year's MAR:

- Role of Local Benefits in Global Environmental Programs (GEF/ME/C.27/4, October 2005)
- Annual Performance Report 2005 (GEF/ME/C.28/2/Rev.1, May 2006)
- Evaluation of Incremental Cost Assessment (GEF/ME/C.30/2, November 2006)
- Evaluation of the GEF Activity Cycle and Modalities (GEF/ME/C.30/6, November 2006)

- Evaluation of the GEF Support to Biosafety (GEF/ME/C.28/Inf.1, May 2006)
- Annual Performance Report 2006 (GEF/ME/C.31/1, May 2007)
- Country Portfolio Evaluation: Philippines (1992-2007) (GEF/ME/C.31/3, May 2007)
- Joint Evaluation of the Small Grants Programme – Executive Version (GEF/ME/C.32/2, October 2007)
- GEF Annual Report on Impact 2007 – Executive Version (GEF/ME/C.32/4, October 2007)

And three new evaluations presented to Council in 2008:

- Annual Country Portfolio Evaluation Report 2008 (GEF/ME/C.33/4, Mar 2008)
- Annual Performance Report 2007 (GEF/ME/C.33/2, Mar 2008)
- Mid-term Review of the Resource Allocation Framework (GEF/ME/C.34/2, Oct 2008)

7.2 Findings

146. This year the MAR tracks management actions on 35 Council decisions (see Table 7.1). The GEF Evaluation Office was able to verify 32 of these. Two of the decisions that could not yet be verified are currently being assessed by the GEF Evaluation Office and will be incorporated in the Fourth Overall Performance Study. These include a review of the efficiency of the PIF process, and an assessment of the involvement of social and institutional expertise in project supervision.

7.2.1 Decisions rated at a substantial or high level

147. **For the second consecutive year the percentage of verified Council decisions that have been rated at a substantial or high level by the Evaluation Office has increased, from 44 last year to 59 percent.** This year, most of the verified council decisions coming from the new evaluations included in the MAR have been rated as substantial or high. For example, in response to the Council recommendations from the RAF Mid-term review, the GEF Secretariat has placed substantial efforts in preparing options for a new and improved system for allocations in GEF-5. In addition, following the Council Decision requesting the development of “specific, proactive and more flexible engagement approaches with countries in Africa, particularly LDCs that have limited capacity to access and implement GEF funding” (Annual Country Portfolio Evaluation Report 2008), the GEF Secretariat has proactively engaged LDCs from Africa and other regions by establishing programs such as the *Programmatic approach on Sustainable Forest Management (SFM) in the Congo Basin*, and the *West Africa Program*.

148. According to the GEF Evaluation Office, five Council decisions reached a high level of adoption and will be graduated from next year’s MAR. Two of the decisions that were rated at a high level come from the Evaluation of the GEF Support to Biosafety. This is a result of Council approving the *Program for GEF Support to Biosafety in GEF-4*, which has incorporated the recommendations made in the evaluation and is now under

implementation. As a result, a total of 18 projects have been approved totaling \$21.6 million in funding, and benefiting 28 countries. The other three decisions that were rated high come from three different evaluations. The decision from the 2007 Annual Performance Report invited the GEF Agencies to provide information on progress in June 2009. The decision from the Joint Evaluation of the Small Grants Programme requested the SGP Steering Committee to report for decision of the Council on the actions taken to implement the recommendations at the April 2008 Council Meeting. And the decision from the Mid-term Review of the Resource Allocation Framework, which allowed reallocation of unused funds in the last year of GEF-4.

Table 7.1: Ratings of GEF Progress in Adopting Council Decisions³⁴

Management ratings	GEF EO ratings					Sum of management ratings
	High	Substantial	Medium	Negligible	Not possible to verify yet	
High	5	6	5	-	-	16
Substantial	-	7	-	-	2	9
Medium	-	1	8	-	1	10
Negligible	-	-	-	-	-	0
Not possible to verify yet	-	-	-	-	-	0
Sum of GEF EO ratings	5	14	13	0	3	35

7.2.2 Decisions that have shown no change in ratings

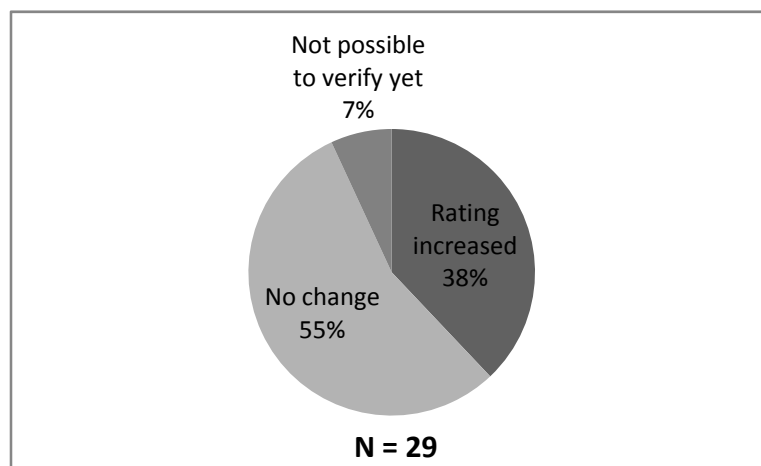
149. **This year MAR finds a noticeable increase in the number of Council decisions that have showed no change in their rating with respect to last year, from 33 to 55 percent.** It is very important to note, however, that adoption of these decisions is divided between two different sets, one set in which progress has taken place and higher ratings are imminent, and a second set in which the Evaluation Office considers that little progress has been achieved.

150. Decision with no change in progress but with imminent higher ratings. On fourteen decisions, the GEF Secretariat has done substantial work towards achieving their adoption, and lack of progress in the rating only reflects the fact that proposals to the Council are yet to be approved, or that it is still too early for the GEF Evaluation Office to properly verify the level of adoption. For example, five of seven decisions coming

³⁴ Highlighted fields show agreement between management and GEF EO; fields to the right of the diagonal represent higher rating by the management than by GEF EO (except in the case of “not possible to verify yet”). The last column shows the sum of ratings in each category by management; the last row shows the sum of ratings by GEF EO.

from the Joint Evaluation of the Small Grants Programme had the same rating as last year. These decisions requested the Small Grants Programme Steering Committee to propose a new level of management cost, to start a process to change SGP's central management system, and to propose a revision of the current criteria for access to SGP resources, among others. The GEF Secretariat and the Small Grants Programme have carried out consultations and are currently preparing a paper for the consideration of the GEF Council in June 2009. A similar situation occurs with the Evaluation of Incremental Cost Assessment. The ratings for the adoption of Council decisions coming from this evaluation have not improved mainly because the strategies for GEF-5 are currently under development and will be discussed at the coming replenishment.

Figure 7.1: Change in Ratings on Adoption of Council Decisions since APR 2007 Assessment



151. Decisions with slow meaningful progress. On the other set, the GEF Evaluation Office considers there has been little progress in the adoption of the decisions. This includes four Council decisions related to the Local Benefits Study. These Council decisions, requesting that local benefits become more systematically addressed in all stages of the project cycle in GEF activities, have maintained the same rating as last year, medium. The GEF Secretariat rated these decisions as fully adopted, commenting that the Secretariat has carried out studies on Indigenous Peoples involvement (April 2008) and Gender Mainstreaming (Nov 2008) in GEF projects, and that a recent study showed that 80% of recently approved projects have included social assessments. In addition the GEF Secretariat notes that for the ongoing focal area strategy development for GEF-5, each focal area's technical advisory group has articulated linkages between global environmental benefits and local benefits, and that they will hire a consultant in 2010 to help integrate social impacts and benefits in Focal Area tracking tools. Although the GEF Evaluation Office finds these measures encouraging, it still considers that socio-economic issues, such as poverty and gender, cannot be effectively mainstreamed into GEF approaches without appropriate in-house expertise. The hiring of temporary consultants does not solve this structural weakness.

7.3 Issues that have resurfaced from last MAR

152. The issue pertaining to the PMIS and accuracy and of access to information has been raised in several evaluations. It was first noted in the 2004 Annual Performance Report and in the Costa Rica Country Portfolio Evaluation. Later, the Philippines Country Portfolio Evaluation and the 2006 Annual Performance Report included Council decisions requesting GEF Secretariat to “ensure transparency of, and better access to, information on GEF procedures and the status of projects in the GEF project cycle”, and to “make the development of a management information system a priority activity”, respectively. Last year, the GEF Secretariat rated the adoption of all Council decisions requesting increased transparency in the GEF approval process through an improved PMIS as high, and the MAR highlighted the noticeable progress in the adoption of these decisions. This year, however, the GEF Evaluation Office finds that there still are several issues that need to be addressed with regard to the PMIS. For example, the Egypt CPE noted the lack of a comprehensive and updated database of GEF activities in the country. Project information coming from the GEF Secretariat, the GEF operational focal point, and the implementing agencies, did not always match, and there was confusion regarding the status of some projects.

153. While the GEF Evaluation Office acknowledges that the PMIS is still a “work in progress”, the fact that it is not considered to be a reliable source even within the Secretariat (GEF Secretariat managers still rely on their own databases for obtaining project information) raises the question as to the degree in which the PMIS has been adopted into the GEF system. The GEF Evaluation Office will continue to closely assess the issue of the access to and quality of information available in the PMIS, both at the country and the project level, thru its various evaluations.

154. The other issue pertains to the Council decision from the Samoa Country Portfolio Evaluation, which requested the Secretariat to take into account Samoa’s experience with the GEF in its further development of the proposed GEF-Pacific Alliance for Sustainability. Adoption of this decision was rated as high by both the GEF Secretariat and the GEF Evaluation Office in last MAR. This year, the Evaluation Office considers that further analysis needs to be done on the inclusion of higher transaction costs in the Pacific Alliance for Sustainability project. The Fourth Overall Performance Study that is currently underway will take a more in depth look into programmatic approaches and will verify whether or not it was justified to include, or not include, transaction costs in the design of these types of projects.

7.4 Factors that influence the speed by which Council decisions are adopted

155. Levels of adoption of Council decisions vary depending on the type of the decision itself. Some Council decisions are very straight forward, and require simple and specific actions from the GEF Secretariat or the GEF Agencies. On a general basis, these decisions are adopted fairly quickly. An example is the Council decision coming from the 2007 Annual Performance Report inviting the GEF Agencies to “provide information

in June 2009 on progress regarding reduction of greenhouse gas emissions in their internal operations”. This decision was made during the March 2008 Council and was rated as high in this MAR.

156. On the other hand, adoption of many Council decisions require substantial changes in strategies and processes, and therefore take longer to be fully adopted and integrated into the GEF system. Council decisions coming from the Evaluation of the GEF Activity Cycle and Modalities (presented to Council in November 2006) requesting GEF Secretariat to “expedite the project cycle”, or to “move the work program from being project-based to being program-based in line with GEF strategies and policies” are still too early in the process for the Evaluation Office to assess their level of adoption.

157. Another example are the different decisions coming from the Annual Performance Reports that deal with the issue of setting up and improving the system for monitoring and evaluating projects. This issue has been raised since the 2004 Annual Performance Report and, while there has already been substantial progress, there are some aspects that still need to be improved. In addition, these will be reconsidered in light of the possible changes that the GEF Evaluation Office might introduce to the GEF M&E Policy and the Terminal Evaluation Guidelines in 2010.

7.5 Lessons related to the MAR

158. The development and use of the MAR has been a valuable learning process for the GEF Evaluation Office. Not only has it allowed the office to track the adoption of the recommendations presented to Council, but it has also served as a platform for reflection on the quantity and quality of these recommendations.

159. Regarding the quantity of recommendations included in each evaluation that is presented to Council, the GEF Evaluation Office is looking into possible options for consolidating these recommendations before presenting them to Council in order to improve the efficiency of communication with Council and of the MAR process. A clear step forward into this direction was presenting a summary of the findings and recommendations coming from the different Country Portfolio Evaluations into one Annual Country Portfolio Evaluation Report. As a result, last year the number of recommendations related to Country Portfolio Evaluations presented to Council was two, instead of twenty had each country portfolio evaluation been presented individually to Council.

160. Finally, the GEF Evaluation Office notices that although the difference between its ratings and the ones given by the GEF Secretariat have decreased (from 49 to 37 percent), a significant discrepancy still persists. In many cases, the ratings made by the Evaluation Office and the ones made by the Secretariat, differ due to their respective interpretations of Council decisions. These differences are more frequent in cases in which Council decisions do not provide specific targets, or describe progress using terms such as “improve”, “expedite”, etc, instead of providing precise guidance.

161. A complete version of the MAR is available at the GEF EO website (www.gefeo.org).

8. Performance Matrix

162. The performance matrix provides a summary of the performance of the GEF Implementing Agencies and GEF Secretariat on a variety of parameters (see table 8.1). Although several of these parameters are assessed by the Evaluation Office on an annual basis, to mitigate fluctuations in performance ratings due to differences in project mix or other idiosyncratic factors, the values presented in the matrix are, depending on the parameter, running averages of two to four years. Of the 13 parameters included in the performance matrix, information has been provided for ten parameters. Also note that the information provided for parameter 13, improvement in performance, addresses only 2 of that parameter's 12 dimensions. In the future, as data for more years become available, it will be possible to track improvement in performance on a greater number of dimensions. See annex D for methodological notes for the performance matrix data reported for this APR.

8.1 Rating Approach

163. Reporting methodology varies by parameter:
- Four performance parameters—project outcomes, implementation completion delays, materialization of cofinancing, and quality of monitoring and evaluation during project implementation—are being reported as four-year running averages, as improvements on these parameters are expected to be gradual. The figures listed in this year's matrix, however, are not four-year averages for all the parameters.
 - The figures reported for performance on quality of project terminal evaluation reports are two-year running averages, as meaningful changes can be attained in the short run.
 - Changes in performance are also likely to be gradual for a second set of parameters: quality of supervision and adaptive management, realism of risk assessment and risk management, and quality of project M&E arrangements. Moreover, assessment of performance on these parameters requires intensive thematic appraisals. For the sake of efficiency, the Evaluation Office will take up such appraisals as part of the APR after a two- to three-year interval.
 - Performance on the parameter on independence of terminal evaluations (parameter 7), which is being reported on for the first time, is rated on a six point scale: from 1, highly unsatisfactory, to 6, highly satisfactory. The findings on this parameter will be updated annually.
164. Several items cannot yet be reported on.
- For assessing performance on **project preparation elapsed time**, information will be analyzed from the GEF database. An assessment on time taken on preparation and processing of Project Information Forms (PIFs) is being carried

out by the Evaluation Office for OPS4. Findings of this assessment will be included in the matrix once they become available.

- Two parameters – independence of GEF partner Agency evaluation units and the assessment of robustness of program result indicators and tracking tools – included in the performance matrix require development of new methodologies and approaches. This is being addressed as part of the work being carried out for OPS4.

Table 8.1: GEF Agency and Institutions Performance Matrix (for OPS4 or FY 2008?)

Parameter ³⁵	UNDP	UNEP	World Bank	GEF Secretariat	Overall GEF Performance
Results					
1. Project Outcomes: percentage of completed projects with outcomes rated moderately satisfactory or above	78	72	85	—	80
Processes affecting results					
2. Quality of supervision and adaptive management: percentage rated moderately satisfactory and above.	88	36	87	—	81
Efficiency					
3. Project preparation elapsed time: average number of months required to prepare projects.	—	—	—	—	—
4. Implementation completion delays: average delay in completion of projects in months	21	12	13	—	16
5. Materialization of Co-financing:					
a. Reported materialization of cofinancing per dollar of approved GEF financing	1.6	1.5	2.3	2.0	2.0
b. Reported materialization of cofinancing as percentage of promised cofinancing	99	101	98	98	98
Quality of M&E					
6. Independence of agency central evaluation units	—	—	—	—	—
7. Independence of terminal evaluations or independent review of terminal evaluations for FSP/MSP	S/S	S/S	HS/MU	—	MS
8. Realism of risk assessment (Robustness of project-at-risk systems): percentage of projects rated moderately satisfactory or above in candor and realism in supervision reporting	75	29	80	71	71
9. Robustness of program result indicators and tracking Tools	—	—	—	—	—
10. Quality assurance of project M&E arrangements at entry	83	—	80	—	76
11. Quality of project M&E during implementation	59	68	79	—	67
12. Quality of project terminal evaluation	94	100	90	—	89
Quality of Learning					
13. Improvement in performance					
(i) Project Outcomes.	—	—	—	—	—
..					
(x) Quality of M&E arrangements at entry	3	2	3	—	3
..					
(xii) Improvement in quality of terminal evaluations: on a scale of 4 (high performance) to 1 (low performance)	4	4	3	—	3

³⁵ See Annex C for detailed notes on rating methodology for each of the parameters included in the table.

8.2 Findings

165. Information has been provided on nine of the 13 parameters included in the performance matrix. However, only three findings, which are more important than the others, are highlighted here.

- The first finding pertains to ratings on outcome achievements. Based on the review of terminal evaluation reports submitted to the Evaluation Office from FY 2005 to FY 2008, the Evaluation Office rated outcome achievement (parameter 1) in 80 percent of the projects to be moderately satisfactory and above. Although, the rated projects do not fall under the provisions of the fourth GEF replenishment, the percentage is higher than the 75 percent target specified for GEF-4. Among the Implementing Agencies, the percentage of World Bank projects with ratings in the satisfactory range has been significantly higher than the target (85 percent).
- The second finding pertains to improvement in quality at of M&E arrangements at entry. A review was carried out in FY 2008 to assess level of compliance of FSPs that were CEO Endorsed in FY 2008 with the GEF M&E requirements at entry. This review was a follow up to a similar review that was presented in FY 2005. The Evaluation Office rated 76 percent of the reviewed projects of FY 2008 cohort to be in compliance with the existing requirements. Had the requirements prevalent in FY 2005 been used to assess compliance, 80 percent of the projects of the FY 2008 cohort would have been in compliance. In either case, there has been a marked improvement compared to FY 2005 when 58 percent had been rated to be in compliance. Among agencies, World Bank and UNDP have shown considerable improvements. Although performance of UNEP is lower, the number of observations for UNEP is not sufficient to allow robust conclusions.
- The third finding relates to the independence of terminal evaluations or the independent review of terminal evaluations. This year's ratings are based on the extent to which Agencies have put in place a system to independently recruit evaluators, and track and report on the quality of terminal evaluations. In this regards the terminal evaluations submitted by UNDP and UNEP are rated 'satisfactory' on independence for FSP and MSPs. Both organizations have transferred the terminal evaluation process to their evaluation offices, which are now responsible for hiring evaluators, developing terms of references of evaluations, ensuring quality control, and reporting on the quality of terminal evaluations and of quality of M&E during implementation. In the case of the World Bank, independence of terminal evaluations of FSP is rated "highly satisfactory". While conducting terminal evaluations is a responsibility of the task managers, these reports are desk reviewed by the evaluation office of the World Bank which verifies outcome ratings, assesses the quality of the terminal evaluation reports and project M&E during implementation. In addition to the desk reviews, the World Bank evaluation office also conducts field verifications for a sample of terminal evaluations. Terminal evaluations for World Bank MSPs, on

the other hand, are rated as “moderately unsatisfactory” as these terminal evaluations are not reviewed by their evaluation office and the World Bank does not have a process in place to systematically track, control quality or report on the quality of terminal evaluations of MSPs.

Annex A. Terminal Evaluation Report Review Guidelines

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

A.1 Criteria for Outcome Ratings

Based on the information provided in the terminal evaluation report, the terminal evaluation review will make an assessment of the extent to which the project's major relevant objectives were achieved or are expected to be achieved, relevance of the project results, and the project's cost-effectiveness³⁶. The ratings on the outcomes of the project will be based on performance on the following criteria³⁷:

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

An overall rating will be provided according to the achievement and shortcomings in the three criteria ranging from highly satisfactory, satisfactory, moderately satisfactory, moderately unsatisfactory, unsatisfactory, highly unsatisfactory, and unable to assess. The reviewer of the terminal evaluation will provide a rating under each of the three criteria (relevance, effectiveness, and efficiency). Relevance of outcomes will be rated on a binary scale: a 'satisfactory' or an 'unsatisfactory' rating will be provided. If an 'unsatisfactory' rating has been provided on this criterion, the overall outcome

³⁶ *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).

³⁷ *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.

achievement rating may not be higher than “unsatisfactory”. Effectiveness and Efficiency will be 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 will consider all three criteria, of which relevance criterion will be applied first - the overall outcome achievement rating may not be higher than “unsatisfactory”. The second constraint that is applied is that the overall outcome achievement rating may not be higher than the “effectiveness” rating. The third constraint that is applied is that the overall rating may not be higher than the average score of effectiveness and efficiency criteria calculated using the following formula:

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

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

A.2 Impacts

Has the project achieved impacts, or is it likely that outcomes will lead to the expected impacts?

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

A.3 Criteria for Sustainability Ratings

Sustainability will be understood as the likelihood of continuation of project benefits after completion of project implementation (GEF 2000). To assess sustainability, the terminal evaluation reviewer will identify and assess the key risks that could undermine continuation of benefits at the time of the evaluation. Some of these risks might include the absence of or inadequate financial resources, an enabling legal framework,

commitment from key stakeholders, and enabling economy. The following four types of risk factors will be assessed by the terminal evaluation reviewer to rate the likelihood of sustainability of project outcomes: financial, sociopolitical, institutional frameworks and governance, and environmental.

The following questions provide guidance to assess if the factors are met:

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

The reviewer will provide a rating under each of the four criteria (financial resources, sociopolitical, institutional, and environmental) as follows:

- 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 number rating 1–4 will be provided in each category according to the achievement and shortcomings with likely = 4, moderately likely = 3, moderately unlikely = 2, unlikely = 1, and not applicable = 0. A rating of unable to assess will be used if the reviewer is unable to assess any aspect of sustainability. In such instances, it may not be possible to assess the overall sustainability.

All the risk dimensions of sustainability are critical. Therefore, the overall rating will not be higher than the rating of the dimension with the lowest rating. For example, if the

project has an unlikely rating in either of the dimensions, then its overall rating cannot be higher than unlikely, regardless of whether higher ratings in other dimensions of sustainability produce a higher average

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

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

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?

Other questions. This includes questions on funding and whether the M&E system was a good practice.

- i. Was sufficient funding provided for M&E — in the budget included in the project document?
- ii. Was sufficient and timely funding provided – for M&E during project implementation?
- iii. Can the project M&E system be considered – a good practice?

A number rating 1–6 will be provided for each criterion according to the achievement and shortcomings with highly satisfactory = 6, satisfactory = 5, moderately satisfactory = 4, moderately unsatisfactory = 3, unsatisfactory = 2, highly unsatisfactory = 1, and unable to assess = no rating. The reviewer of the terminal evaluation will provide a rating under each of the three criteria (M&E design, M&E plan implementation, and M&E properly budgeted and funded) as follows:

- **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 will be the overall rating of the M&E system: Rating on the Quality of the Project Monitoring and Evaluation System = b

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

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

- a. The report presents an assessment of all relevant outcomes and achievement of project objectives in the context of the focal area program indicators if applicable.
- b. The report was consistent, the evidence presented was complete and convincing, and ratings were well substantiated.
- c. The report presented a sound assessment of sustainability of outcomes.
- d. The lessons and recommendations are supported by the evidence presented and are relevant to the portfolio and future projects.
- e. The report included the actual project costs (totals, per activity and per source) and actual 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.

A number rating 1–6 will be provided for each criterion according to the achievement and shortcomings with highly satisfactory = 6, satisfactory = 5, moderately satisfactory = 4, moderately unsatisfactory = 3, unsatisfactory = 2, highly unsatisfactory = 1, and unable to assess = no rating. Each criterion to assess the quality of the terminal evaluation will be rated as follows:

- **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 will be calculated by 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 will be rounded and converted to the scale of highly satisfactory to highly unsatisfactory.

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

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

- Cofinancing and project outcomes and sustainability. If there was a difference in the level of expected cofinancing and actual cofinancing, what were the reasons for it? To what extent did materialization of cofinancing affect project outcomes and/or sustainability? What were the causal linkages of these effects?
- Delays and project outcomes and sustainability. If there were delays, what were the reasons for them? To what extent did the delay affect project outcomes and/or sustainability? What were the causal linkages of these effects?

Annex B

List of projects for which terminal evaluation reviews were conducted in FY 2008³⁸

GEF ID	Project Title	Country	Region	Project Size	IA
22	Commercialization of Super Insulated Building Technology	Mongolia	Asia	MSP	UNDP
98	Energy Conservation	China	Asia	FP	WB
129	Biodiversity Conservation Management Project	Romania	ECA	FP	WB
243	Establishment of a Programme for the Consolidation of the Meso-American Biological Corridor	Regional	LAC	FP	UNDP
280	Reduction of Methane Emissions and Utilization of Municipal Waste for Energy in Amman	Jordan	Asia	FP	UNDP
359	Optimizing Biological Diversity within Wildlife Ranching systems; A Pilot Demonstration in A Semi-arid Zone	Burkina Faso	AFR	FP	UNDP
395	Lake Manzala Engineered Wetlands	Egypt	AFR	FP	UNDP
411	Conservation Management of Eritrea's Coastal, Marine and Island Biodiversity	Eritrea	AFR	FP	UNDP
412	Biodiversity Conservation in the Azov-Black Sea Ecological Corridor	Ukraine	ECA	FP	WB
448	Industrial Energy Efficiency Improvement Project	Malaysia	Asia	FP	UNDP
497	Conservation of Biodiversity and Protected Areas Management	Syria	Asia	MSP	WB
504	Management of Indigenous Vegetation for the Rehabilitation of Degraded Rangelands in the Arid Zone of Africa	Regional	AFR	FP	UNEP
513	Forestry and Conservation Project	Papua New Guinea	Asia	FP	WB
520	Experimental Validation of Building Codes and Removal of Barriers to Their Adoption	Tunisia	AFR	FP	UNDP
571	Low-Cost/Low-Energy Buildings in the Czech Republic	Czech Republic	ECA	MSP	UNDP
613	Environmental Protection of the Rio de la Plata and Its Maritime Front: Pollution Prevention and Control and Habitat Restoration	Regional	LAC	FP	UNDP
622	Energy Conservation and GHG Emission Reduction in Chinese Township and Village Enterprises (TVE), Phase II	China	Asia	FP	UNDP
631	Conservation and Sustainable Use of Medicinal Plants	Ethiopia	AFR	FP	WB
645	Oaxaca Sustainable Hillside Management Project	Mexico	LAC	MSP	WB
648	Coastal and Marine Biodiversity Management Project	Mozambique	AFR	FP	WB

³⁸ Lists of projects for which terminal evaluations were reviewed in earlier years have been considered for review have been presented in earlier APRs: APR 2004 – Annex D; APR 2005 – Annex F; APR 2006 – Annex B; APR 2007 – Annex B.

651	Indigenous Management of Protected Areas in the Amazon	Peru	LAC	FP	WB
653	Coastal and Marine Biodiversity Conservation in Mindanao	Philippines	Asia	FP	WB
655	Ozone Depleting Substance Consumption Phase-out Project : Tranche III - Small Grant Program (SGP) - Residual ODS Phase Out Management Component	Russian Federation	ECA	FP	WB
773	Caribbean Archipelago Biosphere Reserve: Regional Marine Protected Area System	Colombia	LAC	MSP	WB
774	Conservation and Sustainable Use of Biodiversity in the Andes Region	Colombia	LAC	FP	WB
780	Development of Mnazi Bay Marine Park	Tanzania	AFR	FP	UNDP
784	Methane Capture and Use (Landfill Demonstration Project)	Mexico	LAC	FP	WB
802	Conservation of Biodiversity through Integrated Collaborative Management in Rekawa, Ussangoda, and Kalametiya Coastal Ecosystems	Sri Lanka	Asia	MSP	UNDP
815	Dry Forest Biodiversity Conservation	Grenada	LAC	MSP	WB
837	Conservation and Sustainable Use of the Mesoamerican Barrier Reef	Regional	LAC	fp	WB
855	Establishment of the Nuratau-Kyzylkum Biosphere Reserve as a Model for Biodiversity Conservation	Uzbekistan	ECA	MSP	UNDP
864	Multi-agency and Local Participatory Cooperation in Biodiversity Conservation in Yunnan's Upland Ecosystem	China	Asia	MSP	UNDP
874	Assessments of Impacts and Adaptation to Climate Change in Multiple Regions and Sectors (AIACC)	Global	CEX	FP	UNEP
876	Partnership for Natural Ecosystem Management Program (PAGEN)	Burkina Faso	AFR	FP	WB
906	Landscape-scale Conservation of Endangered Tiger and Rhinoceros Populations in and Around Chitwan National Park	Nepal	Asia	MSP	UNDP
922	Baltic Sea Regional Project, Tranche 1	Regional	ECA	FP	WB
925	Conservation of Montane Forest and Paramo in the Colombian Massif, Phase I	Colombia	LAC	FP	UNDP
945	National Protected Areas System	Ecuador	LAC	FP	WB
957	Conservation and Sustainable Use of Biodiversity in the Amarakaeri Communal Reserve and Adjoining Indigenous Lands	Peru	LAC	MSP	UNDP
983	Improving the Energy Efficiency of the Heat and Hot Water Supply	Turkmenistan	ECA	MSP	UNDP
1067	Integrated Coastal and Marine Biodiversity Management	Gambia	AFR	MSP	WB
1086	Developing an Integrated Protected Area System for the Cardamom Mountains	Cambodia	Asia	MSP	UNDP
1159	Agricultural Pollution Control Project - under WB-GEF Strategic Partnership for Nutrient Reduction in the Danube River and Black Sea	Romania	ECA	FP	WB
1301	Conservation of Biodiversity in Pastaza	Ecuador	LAC	MSP	WB

1307	In-situ Conservation of Native Landraces and their Wild Relatives in Vietnam	Vietnam	Asia	MSP	UNDP
1318	Reducing Greenhouse Gas Emissions through the Use of Biomass Energy in Northwest Slovakia	Slovak Republic	ECA	MSP	UNDP
1340	Promoting Industrial Energy Efficiency through a Cleaner Production/Environmental Management System Framework	Global	CEX	MSP	UNEP
1344	Conservation of Gramineae and Associated Arthropods for Sustainable Agricultural Development in Africa	Regional	AFR	MSP	UNEP
1409	Galapagos Oil Spill : Environmental Rehabilitation and Conservation	Ecuador	LAC	MSP	UNDP
1416	Community -based Conservation and Sustainable Use of the Atiquipa and Taimara Lomas Ecosystems	Peru	LAC	MSP	UNDP
1438	Conservation and Sustainable Use of Biodiversity in Dibeen Nature Reserve	Jordan	Asia	MSP	UNDP
1604	Sustainable Conservation of Globally Important Caribbean Bird Habitats: Strengthening a Regional Network for a Shared Resource	Regional	LAC	MSP	UNEP
1611	Developing a Model Conservation Programme-Conservation of the Gobi Desert Using Wild Bactrian Camels as an "Umbrella Species".	Mongolia	Asia	MSP	UNDP
1650	Enabling Activities Leading to the Second National Communication of the Argentine Government to the Conference of the Parties to UNFCCC	Argentina	LAC	EA	WB
1694	Development of the Econet for Long-term Conservation of Biodiversity in the Central Asia Ecoregions	Regional	ECA	MSP	UNEP
1702	Rehabilitation and Expansion of Small Hydro-Plants on the River Raba in Hungary	Hungary	ECA	MSP	WB
1794	Removing Obstacles to Direct Private-Sector Participation in In-situ Biodiversity Conservation	Bolivia	LAC	MSP	WB
1952	Support for World Parks Congress, September 8-17, 2003, Durban, South Africa	Global	CEX	MSP	UNEP
2042	Strengthening the Implementation Capacities for Nutrient Reduction and Transboundary Cooperation in the Danube River Basin (Tranche 2)	Regional	ECA	FP	UNDP
2167	Global Support to Facilitate the Early Development & Implementation of Land Degradation Programs & Project Under the GEF Operational Programme N 15	Global	CEX	MSP	UNEP
2263	Control of Eutrophication, Hazardous Substances and Related Measures for Rehabilitating the Black Sea Ecosystem: Tranche 2	Regional	ECA	FP	UNDP
2726	Capacity Building for Sustainable Land Management (SLM) in Bulgaria	Bulgaria	ECA	MSP	UNDP

Annex C

C.1 Quality of M&E at Entry

In this annex the instrument used to assess quality of M&E plans and the percentage of projects in compliance with the minimum expected level of performance has been provided. For more details also refer to Annex D and Annex E of APR 2005.

Instrument for Assessment of M&E Plans

S. No	Parameters	Response and Raw Score
1	Is there at least one specific indicator in the log frame for each of the project objectives and outcomes?	Yes.....3 No.....1
2	Are the indicators in the log frame relevant to the chosen objectives and outcomes?	Yes.....3 Yes, but only some are relevant.....2 No.....1
3	Are the indicators in the log frame sufficient to assess achievement of the objectives and outcomes?	Sufficient.....3 Largely Sufficient.....2 Some important indicators are missing.....1
4	Are the indicators for project objectives and outcomes quantifiable?	Yes.....3 Some of them are.....2 No, or else it has not been shown how the indicators could be quantified.....1
5	Has the complete and relevant baseline information been provided?	Yes, complete baseline info provided.....3 Partial info but baseline survey in 1 st year.....2.5 No info but baseline survey in 1 st year.....2 Only partial baseline information.....1.5 No info provided.....1
6	Has the methodology for determining the Baseline been explained?	Yes.....3 No.....1
7	Has a separate budget been allocated to M&E activities?	Yes.....3 No.....1
8	Have the responsibilities been clearly specified for the M&E activities?	Yes, and clearly specified.....3 Yes, broadly specified.....2 No.....1
9	Have the time frames been specified for the M&E activities?	Yes, for all the activities.....3 Yes, but only for major activities2 No.....1
10	Have the performance standards (targets) been specified in the log frame for the project outputs?	Yes, for all the outputs.....3 Yes, but only for major outputs.....2 No.....1
11	Have the targets been specified for the indicators for project objectives and outcomes in the log frame?	Yes, for most.....3 Yes, but only for some indicators2 No1
12	Are the specified targets for indicators of project objective and outcomes based on initial conditions?	Yes, for most.....3 Yes, but only for some of the indicators.....2 No.....1
13	Do the project documents mention having made a provision for mid term and terminal evaluation?	Yes, both mid term and terminal evaluation....3 Only terminal evaluation.....2.5 Only mid term evaluation.....1.5 No information provided.....1

Projects that scored two or more on a parameter were rated to be in compliance with the requirements on that parameter. The percentage of projects in compliance with minimum requirements on respective parameters is provided in the following table. For each parameter the responses that were considered to be in compliance have been provided in parenthesis in front of the corresponding parameter. Also note that since the GEF has shifted to results based management, targets for outcome achievements are now accorded lesser importance in M&E plans.

C.2: Percentage of projects meeting the minimum criteria

S. No	Parameters	FY 2005	FY 20008
1	Is there at least one specific indicator in the log frame for each of the project objectives and outcomes? (Yes)	57%	95%
2	Are the indicators in the log frame relevant to the chosen objectives and outcomes? (Yes, all or at least some are relevant)	100%	100%
3	Are the indicators in the log frame sufficient to assess achievement of the objectives and outcomes? (Sufficient or largely sufficient)	76%	94%
4	Are the indicators for project objectives and outcomes quantifiable? (Yes or some of them are)	97%	99%
5	Has the complete and relevant baseline information been provided? (Yes, complete baseline info provided, or Partial info but baseline survey in 1 st year, or no information but baseline survey in 1 st year)	92% (36%) ³⁹	95% (76%)
6	Has the methodology for determining the Baseline been explained? (Yes)	84%	99%
7	Has a separate budget been allocated to M&E activities? (Yes)	92%	95%
8	Have the responsibilities been clearly specified for the M&E activities? (Yes, and clearly or broadly specified)	99%	100%
9	Have the time frames been specified for the M&E activities? (Yes, and clearly or broadly specified)	99%	100%
10	Have the performance standards (targets) been specified in the log frame for the project outputs? (Yes, for all outputs or for major outputs)	95%	85%
11	Have the targets been specified for the indicators for project outcomes in the log frame? (Yes, for all outcomes or for major outcomes)	89%	99%
12	Are the specified targets for indicators of project outcomes based on initial conditions? (Yes, for all outcomes or for major outcomes)	82%	94%
13	Do the project documents mention having made a provision for mid-term and terminal evaluation? (Yes, both mid-term and terminal evaluation, or at least the terminal evaluation)	81%	100%

³⁹ The figures in parentheses indicate the project in compliance when requirements for baseline information as specified in the 2006 M&E policy (which was not operational at the time projects of the FY 2005 cohort were approved) are applied.

C.3 List of FY 2008 Projects Reviewed for Quality of M&E at Entry

GEF ID	ProjectName	IA	Country	Focal	Project Size	Compliant on '08 requirements
967	Private Sector Led Development of On-Grid Wind Power in Tunisia	UNDP	Tunisia	CC	FP	Yes
1017	Partnership Interventions for the Implementation of the Strategic Action Programme (SAP) for Lake Tanganyika	UNDP	Regional	IW	FP	Yes
1027	Strengthening Governance and Financial Sustainability of the National Protected Area System	UNDP	Ukraine	BD	FP	Yes
1028	Mainstreaming Conservation of Migratory Soaring Birds into Key Productive Sectors along the Rift Valley/Red Sea Flyway (Tranches 1 and 2)	UNDP	Regional	BD	FP	Yes
1032	Sustainable Management of the Shared Marine Resources of the Caribbean Large Marine Ecosystem (CLME) and Adjacent Regions	UNDP	Regional	IW	FP	No
1040	Solar Thermal Hybrid Project	World Bank	Egypt	CC	FP	No
1156	Mainstreaming Conservation and Sustainable Use of Medicinal Plant Diversity in Three Indian States	UNDP	India	BD	FP	Yes
1197	Enhancing the Effectiveness and Catalyzing the Sustainability of the W-Arly-Pendjari (WAP) Protected Area System	UNDP	Regional	BD	FP	Yes
1207	Regional System of Protected Areas for Sustainable Conservation and Use of Valdivian Temperate Rainforest	UNDP	Chile	BD	FP	Yes
1234	Community-based Coastal and Marine Biodiversity Management Project	World Bank	Benin	BD	FP	Yes
1239	Sustainable Development of the Protected Area System	UNDP	Ethiopia	BD	FP	Yes
1252	Bay of Bengal Large Marine Ecosystem	FAO / World Bank	Regional	IW	FP	Yes
1331	Demonstrating Cost-effectiveness and Sustainability of Environmentally-sound and Locally Appropriate Alternatives to DDT for Malaria Control in Africa	UNEP	Regional	PP	FP	No
1420	Reducing Dependence on POPs and other Agro-Chemicals in the Senegal and Niger River Basins through Integrated Production, Pest and Pollution Management	UNEP	Regional	MF	FP	No
1520	Development of a National Implementation Plan in India as a First Step to Implement the Stockholm Convention on Persistent Organic Pollutants (POPs).	UNIDO	India	PP	FP	No
1607	Increased Access to Electricity Services	World Bank	Zambia	CC	FP	Yes
1614	Demonstrating the Development and Implementation of a Sustainable Island Resource Management Mechanism in a Small Island Developing State	UNDP	Antigua And Barbuda	MF	FP	No
1620	Mainstreaming Biodiversity Management into Production Sector Activities	UNDP	Seychelles	BD	FP	Yes
1802	Demonstrating and Promoting Best Techniques and Practices for Reducing Health-care Waste to Avoid Environmental Releases of Dioxins and Mercury	UNDP	Global	PP	FP	No

2035	SFM Strengthening Protected Area System of the Komi Republic to Conserve Virgin Forest Biodiversity in the Pechora River Headwaters Region	UNDP	Russian Federation	BD	FP	Yes
2096	Development of a National Implementation Plan in Brazil as a First Step to Implement the Stockholm Convention on Persistent Organic Pollutants (POPs)	UNEP	Brazil	PP	FP	No
2127	Conservation and Adaptive Management of Globally Important Agricultural Heritage Systems (GIAHS)	FAO	Global	BD	FP	Yes
2129	Demonstrating and Capturing Best Practices and Technologies for the Reduction of Land-sourced Impacts Resulting from Coastal Tourism	UNEP	Regional	IW	FP	Yes
2133	Lake Skader-Shkoder Integrated Ecosystem Management	World Bank	Regional	IW	FP	Yes
2261	Building Partnerships to Assist Developing Countries to Reduce the Transfer of Harmful Aquatic Organisms in Ships' Ballast Water (GloBallast Partnerships)	UNDP	Global	IW	FP	Yes
2268	SIP: Integrated Ecosystem Management in Four Representative Landscapes of Senegal, Phase 2	UNDP	Senegal	LD	FP	Yes
2329	Global Programme to Demonstrate the Viability and Removal of Barriers that Impede Adoption and Successful Implementation of Available, Non-Combustion Technologies for Destroying Persistent Organic Pollutants (POPs)	UNIDO	Philippines	PP	FP	No
2355	Agricultural Productivity and Sustainable Land Management	World Bank	Kenya	LD	FP	Yes
2368	Hanoi Urban Transport Development	World Bank	Vietnam	CC	FP	Yes
2372	Forest and Mountain Protected Areas Project	World Bank	Bosnia-Herzegovina	BD	FP	Yes
2377	Sustainable Land Management in the High Pamir and Pamir-Alai Mountains - and Integrated and Transboundary Initiative in Central Asia Phase I	UNEP	Regional	LD	FP	No
2499	Productive Uses of Renewable Energy in Guatemala	UNDP	Guatemala	CC	FP	Yes
2509	Sustainable Land Management for Combating Desertification (Phase I)	UNDP	Pakistan	LD	FP	Yes
2511	Groundnut Basin Soil Management and Regeneration	UNDP	Senegal	LD	FP	Yes
2517	Sustainable Environmental Management for Sixaola River Basin	IADB	Regional	MF	FP	Yes
2545	Catalyzing the Implementation of Uruguay's National Protected Area System	UNDP	Uruguay	BD	FP	Yes
2555	Promotion of a Wind Power Market	World Bank	Jordan	CC	FP	Yes
2584	Nile Transboundary Environmental Action Project (NTEAP), Phase II	UNDP	Regional	IW	FP	Yes
2589	Institutionalizing Payments for Ecosystem Services	UNDP	Global	BD	FP	Yes
2600	Strategic Partnership for the Mediterranean Large Marine Ecosystem-Regional Component: Implementation of Agreed Actions for the Protection of the Environmental Resources of the Mediterranean Sea and Its Coastal Areas	UNEP / UNIDO	Regional	MF	FP	Yes
2604	Sustainable Public Transport and Sport: A 2010 Opportunity	UNDP	South Africa	CC	FP	Yes
2609	GEF-World Bank-China Urban Transport Partnership	World	China	CC	FP	Yes

	Program (CUTPP)	Bank				
2611	Integrated Energy Services for Small Localities of Rural Mexico	World Bank	Mexico	CC	FP	Yes
2614	Adaptation to Climate Change - Responding to Shoreline Change and its human dimensions in West Africa through integrated coastal area management.	UNDP	Regional	CC	FP	Yes
2615	National Grasslands Biodiversity Program	UNDP	South Africa	BD	FP	Yes
2625	Energy Efficiency Project	World Bank	Argentina	CC	FP	Yes
2633	Mainstreaming and Sustaining Biodiversity Conservation in Three Productive Sectors of the Sabana Camaguey Ecosystem	UNDP	Cuba	BD	FP	Yes
2689	Latin America: Multi-country Capacity-building for Compliance with the Cartagena Protocol on Biosafety	World Bank	Regional	BD	FP	No
2700	Implementation of Sustainable Development Strategy for the Seas of East Asia (SDS-SEA)	UNDP	Regional	IW	FP	No
2703	Effective Conservation and Sustainable Use of Mangrove Ecosystems in Brazil	UNDP	Brazil	BD	FP	Yes
2720	Regional Project to Develop Appropriate Strategies for Identifying Sites Contaminated by Chemicals listed in Annexes A, B and/or C of the Stockholm Convention	UNIDO	Regional	PP	FP	No
2753	Participatory Coastal Zone Restoration and Sustainable Management in the Eastern Province of Post-Tsunami Sri Lanka	IFAD	Sri Lanka	MF	FP	Yes
2764	National Biodiversity Mainstreaming and Institutional Consolidation Project	World Bank	Brazil	BD	FP	Yes
2774	Community-based Adaptation (CBA) Programme	UNDP	Global	CC	FP	Yes
2777	Barrier Removal to the Cost-Effective Development and Implementation of Energy Standards and Labeling Project (BRESL)	UNDP	Regional	CC	FP	Yes
2794	SIP-Country Program for Sustainable Land Management (ECPSLM)	World Bank	Ethiopia	LD	FP	No
2801	Promotion of Environmentally Sustainable Transport in Metropolitan Managua	UNDP	Nicaragua	CC	FP	Yes
2889	Zambezi Valley Market Led Smallholder Development	World Bank	Mozambique	MF	FP	Yes
2902	Design and Implementation of Pilot Climate Change Adaptation Measures in the Andean Region	World Bank	Regional	CC	FP	Yes
2903	Tanzania Energy Development and Access Project (TEDAP)	World Bank	Tanzania	CC	FP	Yes
2911	West African Regional Biosafety Program	World Bank	Regional	BD	FP	Yes
2927	Environmentally Sustainable Management of Medical Waste in China	UNIDO	China	PP	FP	Yes
2931	Adaptation to Climate Change through Effective Water Governance	UNDP	Ecuador	CC	FP	No
2932	Alternatives to DDT Usage for the Production of Anti-fouling Paint	UNDP	China	PP	FP	Yes
2935	Micro-turbine Cogeneration Technology Application Project (MCTAP)	UNDP	Indonesia	CC	FP	Yes
2949	Critical Ecosystems Partnership Fund (CEPF), Phase 2	World Bank	Global	BD	FP	Yes

2950	Lighting the "Bottom of the Pyramid"	World Bank / IFC	Regional	CC	FP	Yes
2951	Energy Efficiency Financing	World Bank	China	CC	FP	Yes
2970	Integrated Nutrient Pollution Control Project-under the WB-GEF Investment Fund for Nutrient Reduction in the Danube River and Black Sea	World Bank	Romania	IW	FP	Yes
2996	Portfolio Approach to Distributed Generation Opportunity (PADGO) (Phase 1)	World Bank / IFC	Sri Lanka	CC	FP	Yes
3148	Agricultural Pollution Control Project - under the Strategic Partnership Investment Fund for Nutrient Reduction in the Danube River and Black Sea	World Bank	Croatia	IW	FP	No
3219	Reducing Climate Change-induced Risks and Vulnerabilities from Glacial Lake Outbursts in the Punakha-Wangdi and Chamkhar Valleys	UNDP	Bhutan	CC	FP	Yes
3227	Conservancy Adaptation Project	World Bank	Guyana	CC	FP	Yes
3228	Small Grants Programme, 4th Operational Phase	UNDP	Global	MF	FP	No
3232	CACILM Partnership Framework - Land Improvement Project	ADB	Uzbekistan	LD	FP	Yes
3233	CACILM: Southern Agriculture Area Development Project--under CACILM Partnership Framework, Phase 1	ADB	Kyrgyzstan	LD	FP	Yes
3234	CACILM: Rural Development Project under CACILM Partnership Framework, Phase I	ADB	Tajikistan	LD	FP	No
3254	Mainstreaming Prevention and Control Measures for Invasive Alien Species into Trade, Transport and Travel Across the Production Landscape	UNDP	Seychelles	BD	FP	Yes
3263	Strengthening Institutions, Regulations and Enforcement Capacities for Effective and Efficient Implementation of the National Implementation Plan (NIP) in China	UNIDO	China	PP	FP	Yes
3265	Mainstreaming Adaptation to Climate Change Into Water Resources Management and Rural Development	World Bank	China	CC	FP	Yes
3296	Geothermal Power Generation Development Program	World Bank	Indonesia	CC	FP	No
3567	CPP Burkina Faso: Sub-programme of the Northern Region-under Partnership Programme for Sustainable Land Management	IFAD	Burkina Faso	LD	FP	No

Annex D. Methodological Notes on the Performance Matrix

This annex briefly describes the considerations taken into account for each of the performance matrix's 13 parameters.

D.1 Project Outcomes

Although the figures on project outcomes are projected to be four-year moving averages based on the terminal evaluation reports submitted in the preceding years, including the fiscal year for which the APR is being presented, the figures presented in this year's APR are based on the terminal evaluation reports submitted during FYs 2005, 2006, 2007, and 2008. The aggregate figures are weighted averages, with each project considered to have equal weight.

D.2 Quality of Supervision and Adaptive Management

The figures presented on quality of supervision and adaptive management are based on the findings of the pilot assessment of project supervision presented in the FY 2006 APR. The projects considered for this assessment were under implementation during FY 2005 and/or FY 2006. The figures will be updated and included in OPS4.

D.3 Project Preparation Elapsed Time

The figures presented for project preparation elapsed time will indicate average number of months required to prepare projects. The data on this parameter will be provided by the Agencies and the GEF Secretariat database. These figures will be updated biennially. This year, no figures are provided for this parameter yet. However, figures on preparation and approval of Project Information Forms (PIFs) will be presented in OPS4.

D.4 Project Implementation

Completion Delays

The information presented in the terminal evaluation reports is the primary source for this parameter. The figures for implementation completion delays are projected to be four-year averages and are based on the information provided in the terminal evaluation reports. The figures presented in this year's APR are based on the terminal evaluation reports submitted during FYs 2005, 2006, 2007, and 2008..

D.5 Materialization of Cofinancing

The figures for materialization of cofinancing pertain to projects whose terminal evaluation reports were submitted to the Office during FYs 2005, 2006, 2007, and 2008. The analysis is based on the information provided by the Agencies in the terminal evaluation reports or through other communications. These figures have not been verified.

D.6 Independence of Agency Evaluation Units

Last year, the GEF Evaluation Office started a consultation process with the evaluation units of the GEF Agencies to define an appropriate way forward in assessing their independence. Broadly, the assessment provided on this parameter will be based on self-reporting by the Agencies and peer reviews carried out in the context of the Evaluation Cooperation Group of the Banks and the United Nations Evaluation Group. The charter and mandate of the various evaluation units will also provide evidence of their degree of independence. No ratings are provided on this parameter this year.

D.7 Independence of Terminal Evaluations

Independence of terminal evaluations will be appraised through the assessment of the process followed for conducting terminal evaluations through field verification and will be based on interviews with relevant staff and consultants of the partner Agencies. This will allow the Office to assess the extent to which systems in the partner Agencies are conducive to unbiased and candid terminal evaluations. Independence of terminal evaluations is rated in a six point scale on which a 1 is “highly unsatisfactory” and a 6 is “highly satisfactory”. The following dimensions are assessed to provide ratings on this parameter:

- Extent to which the drafting of the terms of reference is independent of the project management team
- Extent to which the recruitment of the evaluator was independent of the project management team
- Extent to which the Agency recruited the appropriate evaluator for the project
- Extent to which the evaluator had adequate resources (budget and time) to carry out the evaluation
- Extent to which the M&E system provides access to timely and reliable information
- Extent to which there was any undue pressure from management on the evaluators regarding the evaluation process (for example, in terms of site selection, selection of informants, confidentiality during interviews, information disclosure, and ratings)
- Extent to which the evaluation was subjected to an independent review process

D.8 Realism of Risk Assessment

The figures for realism of risk assessment are based on the findings of the pilot assessment of project supervision for candor and realism of supervision reporting presented in the 2006 APR. The projects considered for this assessment were under implementation during FY 2005 and/or FY 2006. The updated figures for projects under implementation during FY 2007 and FY 2008 will be included in OPS4.

D.9 Robustness of Program Result Indicators and Tracking Tools

The assessment of robustness of program result indicators and tracking tools will remain unreported in the 2008 APR. Given the highly specialized and technical nature of this assessment, the Evaluation Office is taking up this exercise as part the work being done for the OPS4, enlisting the assistance of the appropriate technical experts.

D.10 Quality Assurance of Project M&E Arrangements at Entry

An assessment of quality assurance of project M&E arrangements at entry was carried out in the 2005 APR. It was based on a review of the M&E plans of the project appraisal documents that were endorsed by the GEF Chief Executive Officer in that fiscal year. In FY 2008, the Evaluation Office updated the ratings on this parameter based on the findings of a follow-up assessment.

D.11 Quality of Project M&E during Implementation

Figures on quality of project M&E during implementation are based on review of the terminal evaluation reports submitted to the Evaluation Office. The figures need to be four-year running averages of the percentage of projects rated moderately satisfactory or above in M&E during implementation. The figures reported in the matrix are a weighted average, with each project having an equal weight, of the data from the review of the reports submitted during FYs 2006, 2007, and 2008. In due course, as data for subsequent cohorts become available, the figures presented will shift to a four-year running average.

D.12 Quality of Project Terminal Evaluation

Figures on quality of terminal evaluation reports are based on the ratings provided by the Evaluation Office after their review. For this parameter, twoyear running averages are used, with each project having an equal weight. The figures presented in the matrix pertain to FYs 2007 and 2008.

D.13 Quality of Learning – Improvement in Performance

The performance matrix presents an assessment of the improvement demonstrated by GEF Agencies and entities on the other 12 parameters included in the performance matrix.

This section of the matrix will be accompanied by a narrative that explains the areas in which learning has taken place and will identify the specific changes or factors that have contributed to improved performance. Ratings on improvement in performance on individual parameters will be provided using the following scale:

- **4**—significantly improved or maintained excellent performance
- **3**—marginally improved or maintained good performance
- **2**—marginally deteriorated or maintained mediocre performance
- **1**—significantly deteriorated or maintained poor performance

In this performance matrix, only the changes in quality of M&E arrangements at entry (parameter 10) and in quality of terminal evaluations (parameter 12) have been reported on. For assessing changes in quality of M&E arrangements at entry the findings of the pilot assessment in FY 2005 were updated and compared with those of a follow up assessment in FY 2008. For assessing changes in quality of terminal evaluation reports, a two-year running average of quality ratings for the terminal evaluation reports submitted during FYs 2007 and 2008 has been compared with the baseline for the FYs 2005 and 2006 cohort.