GEF Annual Performance Report 2006

MARCH 2008

GEF

Global Environment Facility
Evaluation Office

GEF Annual Performance Report 2006

March 2008

(The main findings and recommendations of this evaluation were presented to the GEF Council in June 2007.)

Evaluation Report No. 38

1818 H Street, NW
Washington, DC 20433
Internet: www.gefeo.org
Email: gefevaluation@thegef.org

ISBN-10: 1-933992-10-7
ISBN-13: 978-1-933992-10-5

Credits
Director of the GEF Evaluation Office: Robert D. van den Berg
Task Manager: Aaron Zazueta, Senior Evaluation Officer, GEF Evaluation Office
Evaluation Team: Nadia Moriah Ahlsten, Ines Angulo, Neeraj Kumar Negi, and Frederick Swartzendruber, consultants

Editing and design: Nita Congress
Printing: Graphic Communications
Cover photo: China, Curt Carnemark, World Bank

Evaluation Report No. 38

A FREE PUBLICATION

Contents

Tables

Abbreviations

APR	annual performance report
CEO	Chief Executive Officer
ExA	Executing Agency
FSP	full-size project
FY	fiscal year
GEF	Global Environment Facility
IA	Implementing Agency
IBRD	International Bank for Reconstruction and Development
IDA	International Development Association
IEG	Independent Evaluation Group
M&E	monitoring and evaluation
MAR	management action record
MSP	medium-size project
PIR	project implementation report
QAG	Quality Assurance Group
QSA	quality of supervision assessment
UNDP	United Nations Development Programme
UNEP	United Nations Environment Programme
WB	World Bank

Foreword

This is the third annual performance report (APR) of the Global Environment Facility (GEF) Evaluation Office on the performance of the GEF portfolio, as reported through the monitoring and evaluation systems of the GEF Agencies. It provides an independent assessment of project outcomes and sustainability, delays in project completion, materialization of cofinancing, and quality of monitoring and evaluation arrangements, continuing the work that was initiated in the fiscal year 2005 APR. This year's APR also looks for the first time at the quality of supervision of GEF projects by the three Implementing Agencies. The pilot assessment of supervisory quality was conducted in close collaboration with the United Nations Development Programme, the United Nations Environment Programme, and particularly with the Quality Assurance Group of the World Bank.

The GEF Council reviewed the 2006 APR and the management response (included as annex F) at its 31st meeting in June 2007. Based on this review, it requested that the United Nations Development Programme and the United Nations Environment Programme involve social and institutional expertise in project supervision where appropriate. Second, it noted that, because special attention is required to ensure continued and improved supervision by the GEF Agencies during project implementation, adequate funding should be provided for this supervision from the project fees. Third, it requested that the United Nations Environment Programme develop a systemic approach to supervision of its GEF portfolio. Finally, it requested that all GEF Agencies ensure that terminal evaluation reports include adequate information on sustainability of outcomes, quality of monitoring and evaluation systems, and cofinancing, in line with the minimum requirements for project evaluation set forth in the GEF Monitoring and Evaluation Policy.

The Council also noted that negligible progress has been made in developing a management information system and asked that the Secretariat make this a priority activity for completion before the end of the calendar year. It encouraged the GEF Evaluation Office to include in future APRs an agency performance matrix as required in the GEF-4 policy recommendations.

Various intermediary products were discussed with the GEF Secretariat and Implementing Agencies, which has led to significant enrichments of the work performed. I would like to thank all those involved for their support and useful criticism.

Rob van den Berg
Director, Evaluation Office

Acknowledgments

The task manager for development of this year's annual performance report was Aaron Zazueta, Senior Evaluation Officer in the Evaluation Office of the Global Environment Facility. Frederick Swartzendruber, consultant, with the assistance of Nadia Moriah Ahlsten, consultant, carried out the pilot assessment of project supervision, and drafted chapter 5. Neeraj Kumar Negi, a consultant to the Evaluation Office, drafted the overall report. Ines Angulo, a consultant to the Evaluation Office, drafted chapter 8. Antonio del Monaco, Evaluation Specialist, coordinated the review of the terminal evaluation reports. These reports were reviewed by colleagues in the Evaluation Office, including Lee Risby, Evaluation Officer; Dyvia Nair, Evaluation Specialist; and consultants Anna Viggh, Neeraj Kumar Negi, Ines Angulo, and Tarek Soueid.

The Evaluation Office thanks the staff of the Implementing Agencies for their comments and contributions to this evaluation, and the World Bank Quality Assurance Group for collaborating with the Office in the assessment of project supervision of World Bank–implemented Global Environment Facility projects. The final responsibility for this report remains firmly with this office.

1. Background, Main Conclusions, and Recommendations

1.1 Background

This document is the third annual performance report (APR) that the Global Environment Facility (GEF) Evaluation Office has presented since the GEF Council approved the transfer of responsibility for project monitoring to the Implementing Agencies (IAs) and GEF Secretariat. This transfer has allowed the Evaluation Office to sharpen its focus on assessing the results of GEF activities and overseeing monitoring and evaluation (M&E) operations. The report presents a detailed account of some aspects of project results, of processes that may affect these results, and of M&E arrangements in completed projects. Last year, the Evaluation Office presented an assessment of the M&E arrangements across the GEF partners in considerable detail. This year's APR looks, for the first time, at the quality of supervision of GEF projects by the respective IA.

The APR also continues the assessment of project outcomes, project sustainability, project completion delays, materialization of cofinancing, and quality of monitoring initiated in the fiscal year (FY) 2005 APR.[1] For the **assessment of project outcomes, project sustainability, and delays in project completion**, 107 projects were considered for which terminal evaluation reports were submitted by the IAs to the Evaluation Office in FY 2005 (41 projects) and FY 2006 (66 projects). Altogether, the GEF had invested $514 million in these 107 projects.[2] The focus of this APR's reporting is on the 66 projects for which terminal evaluation reports were submitted in FY 2006 and in which the GEF has invested $255 million.

For **assessment of materialization of cofinancing**, all 182 terminal evaluation reports submitted after January 2001 were considered. Of these, 118 (65 percent) provided information on actual materialization of cofinancing. The GEF has invested a total of $593 million in these 118 projects; it was reported that an additional $2.16 billion was leveraged for these projects in the form of cofinancing.

For **assessing the quality of supervision**, a stratified random sample of 49 GEF projects under implementation during FY 2005 and 2006 was examined in detail. These projects were implemented by the World Bank (15), the United Nations Development Programme (UNDP) (18), and the United Nations Environment Programme (UNEP) (16). The GEF has invested $215 million in these projects. No terminal evaluation reports had yet been received from the Executing Agencies (ExAs) of the GEF.

The findings presented here have several limitations. Most findings are based on terminal evaluation reviews, which are in turn based on information provided by IA terminal evaluation reports. This reliance introduces uncertainty into the verification process, which the Evaluation Office seeks to mitigate by incorporating in its reviews any pertinent information it has independently gathered through other evaluations. The Office is also seeking to improve the independence of terminal evaluation reports by more fully involving the central evaluation units of its partner Agencies in the process.

For many issues on which performance is being reported in this APR, information is presently available only for FY 2005 and 2006. For assessment of quality of terminal evaluation reports, data are available for FY 2004, 2005, and 2006. Although this allows for comparison of performance in these years, it does not permit analysis of long-term trends. Further, the number of projects for some partner Agencies is too small to draw meaningful conclusions for these IAs. These limitations will be mitigated in the future with the accumulation and availability of data for more cohorts.

In November 2005, the GEF Council approved a procedure and format for management action records (MARs) capturing the rate of adoption of Council decisions on evaluation reports. The purpose of the MARs is to facilitate systematic follow-up on and periodic review of the implementation of evaluation recommendations accepted by management and/or the GEF Council. The Evaluation Office and GEF Secretariat, in consultation with the appropriate partner organizations, report annually to the Council on the follow-up of the Council decisions compiled in a MAR (GEF EO 2006a). MARs are published and updated on the GEF Evaluation Office Web site (www.gefeo.org).

1.2 Main Conclusions

Results

Conclusion 1: Completed GEF projects remain on target to achieve the 75 percent satisfactory outcomes as agreed upon in the GEF-4 replenishment agreement.

Among the completed GEF projects assessed and rated this year, 84 percent were rated moderately satisfactory or above in achievement of outcomes, and 65 percent were rated moderately likely or above in sustainability of outcomes.

Attainment of project outcomes. The Evaluation Office rated the achievement of project outcomes on criteria of relevance, effectiveness, and efficiency. Of the 66 terminal evaluation reports submitted in FY 2006, 64 reports (97 percent) provided 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 64 projects whose outcomes were rated by the Evaluation Office, 84 percent were rated moderately satisfactory or above. After controlling for differences in rating methodologies, this performance is similar to that for projects whose terminal evaluation report was submitted in FY 2005.
- Of the total investment in the rated projects ($254 million), 88 percent ($224 million) was allocated to projects that were rated moderately satisfactory or above.
- The outcome ratings of full-size projects (FSPs) and medium-size projects (MSPs) were similar: the outcomes of 85 percent of FSPs and 83 percent of MSPs were rated moderately satisfactory or above.

Project performance for the FY 2006 cohort is on track with the target set for the fourth GEF replenishment (2006–10) of 75 percent of projects having satisfactory outcomes (GEF 2006). Although the completed projects assessed for FY 2006 do not fall under the provisions of the fourth replenishment, their performance does indicate that the target outcome achievement is realistic.

Table 1.1

Summary of Project Outcomes and Sustainability Ratings

Factor	FY 2005	FY 2006	FY 2006
	Old methodology[a]		New methodology[a]
Number of terminal evaluation reports submitted	41	66	66
Projects for which no outcome rating was possible	2	2	2
Number of projects with outcome rating	39	64	64
% rated moderately satisfactory or above in outcome ratings[b]	92	92	84
Projects for which no sustainability of outcomes rating was possible	5	9	12
Number of projects with sustainability of outcomes ratings	36	57	54
% rated moderately likely or above in sustainability of outcomes ratings	76	84	65
Number of projects with moderately satisfactory/moderately likely or above for both	26	47	35
% of rated projects with moderately satisfactory/moderately likely or above for both	72	82	61

a. Different methodologies were used to rate project outcomes in 2005 and 2006; these are explained at the beginning of chapter 3.

b. In this APR, the overall ratings-based figures for the portfolio have been calculated excluding those projects that were not rated (unable to assess or not applicable). This is unlike FY 2005, where the calculations included unrated projects. To ensure comparability, figures for the FY 2005 portfolio have been recalculated excluding unrated projects. For example, the FY 2005 APR reported that 88 percent of the projects submitted in that year had moderately satisfactory or above outcomes. This APR reports the FY 2005 figure as 92 percent.

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 terminal evaluation reports submitted in FY 2006, 54 reports (82 percent) provided sufficient information to allow assessment of sustainability of project outcomes. The key findings of this assessment are as follows:

- Of the 54 projects that were rated, the sustainability of outcomes of 35 projects (65 percent) was rated moderately likely or above. After taking into account differences in rating methodologies, this performance is similar to that of last year's cohort.
- Of the total GEF investment in rated projects ($218 million), 60 percent ($131 million) was invested in projects that were rated moderately likely or above in terms of their sustainability.
- Compared to projects from other Agencies, the sustainability ratings of World Bank projects were significantly higher.

The Evaluation Office has been rating completed projects on achievement of outcomes and sustainability of outcomes since FY 2005; this year, it carried out an additional analysis to assess 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. Of the terminal evaluation reports submitted in FY 2006, 54 reports (82 percent) provided sufficient information to allow assessment of both project outcomes and sustainability of project outcomes, as did 36 reports (88 percent) of those submitted in FY 2005. The key findings of this assessment are as follows:

- Of the rated projects from the FY 2006 cohort, 61 percent were rated moderately satisfactory or above in outcomes and moderately likely or above in sustainability. In terms of GEF investment, 56 percent was invested in projects that were rated moderately satisfactory or above in outcomes and moderately likely or above in sustainability.

- Of the rated projects from the FY 2005 cohort, 72 percent were rated moderately satisfactory or above in outcomes and moderately likely or above in sustainability. In terms of GEF investment, 86 percent was invested in projects that were rated moderately satisfactory or above in outcomes and moderately likely or above in sustainability.

The figures for FY 2005 and 2006 are not directly comparable; however, when the differences in the two years' rating methodologies are taken into account, there is little difference in the ratings of the two cohorts.

Processes

Conclusion 2: The materialization of cofinancing is on track.

There is great variation among projects in levels of cofinancing. The average materialized cofinancing has decreased slightly since FY 2005, from $4.10 per $1.00 of GEF funding to $3.80 per $1.00 of GEF funding. In general, levels of promised cofinancing are achieved. At the point of inception, projects in the FY 2006 cohort had, on average, been promised $2.10 in cofinancing per $1.00 of approved GEF grant. Of this promised cofinancing, $2.40 (114 percent) per $1.00 of approved GEF grant reportedly materialized.

The findings from the cofinancing analysis are as follows:

- Of the 66 reports submitted in FY 2006, 47 (71 percent) provided information on materialization of cofinancing. Of the 116 terminal evaluation reports submitted before FY 2006 and examined by the Evaluation Office, only 71 reports (61 percent) provided information on materialization of cofinancing (see table 1.2). Thus, despite some improvement, reporting on cofinancing continues to be an area that needs further attention.

Table 1.2

Materialization of Cofinancing as Reported in Terminal Evaluation Reports

Factor	FY of report submission					All years[a]
	2002	2003	2004	2005	2006	
Total number of terminal evaluation reports	18	15	42	41	66	**182**
Number that did not report on cofinance	7	8	12	18	19	**64**
Number that did report on cofinance	11	7	30	23	47	**118**
Approved GEF grant per project (million $)	6.2	5.5	5.9	6.4	3.5	**5.0**
Actual GEF grant per project (million $)	6.1	4.6	4.6	6.3	3.3	**4.6**
Promised cofinancing per project (million $)	29.5	8.4	43.2	9.5	7.2	**19.0**
Promised cofinancing per $1.00 of approved GEF grant	4.7	1.5	7.4	1.5	2.1	**3.8**
Reported materialized cofinancing per project (million $)	29.2	7.1	38.9	10.0	8.2	**18.3**
Reported materialized cofinancing per $1.00 of approved GEF grant	4.7	1.3	6.6	1.6	2.4	**3.6**
Materialized cofinancing per $1.00 of promised cofinancing (%)	99	85	90	106	114	**96**

a. In the APR 2005, the analysis on materialization of cofinancing was based on 70 terminal evaluation reports submitted in or before FY 2005 that had reported on cofinancing. Information collected for the Joint Evaluation of the GEF Activity Cycle and Modalities (GEF EO 2007b) allowed the Evaluation Office to incorporate data for one more project whose terminal evaluation reports were submitted in or before FY 2005. Thus, in all, 71 projects from that period have been analyzed.

- For terminal evaluations submitted in FY 2006, the average rate of reported materialized cofinancing was 114 percent of that promised. Expressed in dollar terms, for $1.00 of approved GEF grant, an average of $2.10 in cofinancing had been promised and $2.40 reportedly materialized.
- If all terminal evaluation reports submitted to the Evaluation Office so far are taken into consideration, 118 reports provide information on cofinancing. For these 118 projects, an average of 96 percent of promised cofinancing was reported to have materialized. In dollar terms, project proponents promised an average of $3.80 in cofinancing per $1.00 of approved GEF grant. The actual cofinancing reported was slightly lower, at $3.60 per $1.00 of GEF grant.

On average, the projects of the FY 2006 cohort were completed after a delay of 13 months; 17 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 and 13 months for the FY 2006 cohort. Compared to 44 percent of the projects of the FY 2005 cohort, only 17 percent of the projects in the FY 2006 cohort had delays of two years or more. Although the average delay in project completion is lower for the FY 2006 cohort, it is too early to determine if this is a trend.

The analysis shown in the FY 2005 APR suggested that outcome ratings could be correlated with project completion delays. Inclusion of the data for the FY 2006 cohort allowed the Evaluation Office to explore this issue further. It found that, although outcome ratings and project completion delays are inversely related, this relationship is very weak. Moreover, when other variables such as Implementing Agency and focal area are controlled for, it weakens further. Similarly, the relationship between project completion delays and sustainability ratings was also not confirmed.

Conclusion 3: UNEP does not adequately supervise two-thirds of its sampled projects, although improvements have been realized after the appointment of a portfolio manager. Fiduciary requirements are generally met.

Although the majority of GEF projects are rated as having moderately satisfactory supervision, the level of attention given to supervision of GEF projects is highly variable by Agency. About three-fourths of the sampled projects received supervision at the level of a minimum standard of performance, but there is substantial scope for improvement. There is some evidence that resource constraints, lack of management attention, and issues of decentralization are contributing factors.

Resource constraints were cited as a factor for many of the low-performing project supervision cases, with insufficient time spent in the field and staff limitations the main issues identified. Lack of management attention was often a contributing factor. In the World Bank, GEF projects blended with International Bank for Reconstruction and Development (IBRD) or International Development Association (IDA) operations appear to be more likely to receive management attention and well-staffed supervision teams than stand-alone operations (there has been discussion of phasing out the latter approach in some World Bank regions).

In UNDP, country office staffs directly responsible for supervision receive extensive technical backstopping from regionally based focal area specialists; this combination was the key factor in UNDP's relatively high ratings for supervision quality (see table 1.3).

Table 1.3

Percentage of Projects Rated Moderately Satisfactory or Above for Supervision

Agency			Project size	
UNDP	UNEP	World Bank	FSP	MSP
88	36	87	82	79

In UNEP-supervised projects, serious resource constraints, lack of management attention, and absence of clear guidelines for supervisory responsibility contributed to very poor results in the review (only 36 percent of cases were rated moderately satisfactory or above), though there were also instances of very good performance that indicate the potential for substantial improvement. In recent years, a GEF portfolio manager was appointed in UNEP, which has led to increased attention to and improvements in supervision. Furthermore, UNEP's fiduciary and safeguard supervision of its projects is satisfactory in general. Of 16 projects, 13 were rated moderately satisfactory or above in this regard; this finding should be considered tentative, however, since no financial management or procurement specialists were involved in the assessment.

Conclusion 4: Portfolio monitoring by the GEF Agencies tends to rate projects fully satisfactory despite important problems noted in the monitoring information on the same projects.

Project information reporting has given an unrealistically optimistic picture of portfolio health and project risks. The Evaluation Office found a marked tendency to rate projects as fully satisfactory despite problems reported relating to implementation delays, government commitment, or long-term sustainability. Little attention is being given to possible unintended effects of projects, such as the social impacts of protected area projects.

As part of the pilot assessment of supervision, the project implementation reports (PIRs) for 2005 and 2006 were examined for 55 sampled projects to identify disconnects between the number and seriousness of issues reported and project ratings assigned by IA supervision staff. Of the 55 cases, only 3 had been flagged in PIR ratings as moderately unsatisfactory or lower, with a total GEF grant value of $10 million. The desk review identified another 16 cases—for a total of 19 projects valued at $85.8 million—that could be considered at risk based on issues described in the narrative section of the PIRs, as well as reported performance of project components. The strong tendency to give optimistic performance ratings is consistent with findings of previous assessments by the World Bank's Quality Assurance Group (QAG). The 2006 PIR showed some improvement over 2005, however, with some evidence that IA managers had begun to ask staff to justify ratings that seemed inconsistent with reported problems.

The evaluation found that, at present, only the World Bank has a formal system for screening projects for potential unintended social or environmental impacts that would need to be mitigated and supervised during implementation. In the sample, one case of noncompliance with the Bank's policy was identified. UNDP and UNEP do not now have a formal system for impact screening, and no evidence was found that these aspects were taken into consideration during supervision. In some cases, UNDP and UNEP staff expressed the view that participatory design processes and the involvement of nongovernmental organizations or community groups would tend to prevent such problems, but there was no indication of attention to possible issues once implementation had begun. This presents an area of potential vulnerability (and possible reputational risk), with projects having negative impacts on certain groups, for example.

Conclusion 5: UNDP and UNEP do not sufficiently involve social and institutional specialists during supervision where this would have been appropriate.

The Evaluation Office found that social and institutional specialists were not involved in supervision to a desirable extent, especially in the sampled projects of UNDP and UNEP. The Office's report on the evaluation of the Role of Local Benefits in Global Environmental Programs recommended that in order "to strengthen generation of linkages between local and global benefits, the GEF should ensure adequate involvement of expertise on social and institutional issues at all levels of the portfolio" (GEF EO 2006b, p. 123); and GEF management holds that it is now "a regular practice at every stage of the project cycle to involve appropriate expertise and tools related to social and institutional issues by all Implementing Agencies" (GEF EO 2006b, p. 161). However, in its assessment of supervision, the Evaluation Office found that for the sampled projects only the World Bank is doing so systematically, because of its system of "do no harm" safeguard policies. This system requires all projects to be formally screened by specialists for potential safeguard issues and mitigation plans to be developed (and supervised) where issues are identified. UNDP has prepared a paper on social issues, but there is no indication that it has been made operational in project supervision. UNEP has no paper and no actual practice of involving social and institutional experts in supervision of the sampled projects.

Monitoring and Evaluation

While 78 percent of projects were rated moderately satisfactory or above on quality of monitoring, there is scope for improvement on appropriate indicators and baseline data, as well as for better quality at entry and for funding of M&E.

Of the 66 terminal evaluation reports submitted, 20 (30 percent) did not provide sufficient information on M&E to allow the Evaluation Office to assess the quality of project monitoring (see table 1.4). The key findings from the analysis of the 46 reports (70 percent) that did provide sufficient information on M&E are as follows:

- Seventy-eight percent of projects were rated as moderately satisfactory or above in their quality of project monitoring. In FY 2005, when quality of project monitoring was assessed using a different methodology, 66 percent of the rated projects were found to be moderately satisfactory or above.
- Difficult issues, such as specification of appropriate indicators and provision of baseline information, still need to be addressed in many of the projects that were rated moderately satisfactory or above.

The analysis shows linkages between quality at entry of M&E arrangements and quality of monitoring during project implementation, and provides some evidence to support the case for better funding for M&E activities. The following findings relate to the 66 projects from the FY 2006 cohort:

Table 1.4

M&E Ratings for Projects

	FY of terminal evaluation report submission		
Factor	2004	2005	2006
Total # of terminal evaluation reports	42	41	66
# that did not report on M&E	11	8	20
# that did report on M&E	29	32	46
# for which reporting was not required	2	1	0
% rated moderately satisfactory or above	55	66	78

- Forty-two projects were rated both on quality at entry of M&E arrangements and on quality of monitoring during project implementation. Of the 28 projects that were rated moderately satisfactory or above in quality at entry, the quality of monitoring during project implementation was rated moderately satisfactory or above for 27 projects (96 percent). In contrast, of the 14 projects that were rated moderately unsatisfactory or below in quality at entry, only 5 (36 percent) were rated moderately satisfactory or above in quality of monitoring during project implementation. This suggests a link between quality of M&E arrangements at entry and quality of project monitoring.
- Twenty-seven projects provided information on sufficiency of funding for M&E activities. All 20 projects that were assessed to have had adequate funding for M&E activities were also rated moderately satisfactory or above on quality of monitoring during project implementation. In contrast, of the seven projects that were assessed to have inadequate funding for M&E activities, only two were rated moderately satisfactory or above on quality of monitoring. While these numbers are not sufficient to make broad generalizations, the direction of the relationship is consistent with expectations.

Conclusion 6: A substantial proportion of terminal evaluation reports do not adequately cover issues such as sustainability, cofinancing, and M&E.

Out of the 66 terminal evaluation reports submitted in FY 2006, 20 (30 percent) did not provide sufficient information on project monitoring and 12 (18 percent) on sustainability of outcomes to allow the Evaluation Office to rate performance on these parameters. Further, 29 percent of the reports did not provide information on materialization of cofinancing. The last point is especially relevant with regard to the Council's recent decision that, in future, projects would need to report on levels of cofinancing to ensure that the principle of incrementality has been maintained.

Two factors—the maturation of the GEF portfolio and more prompt submission of terminal evaluation reports—account for the increase in number of terminal evaluations submitted. The Evaluation Office estimates that, compared to the terminal evaluations that were completed in 2004, the average time lag between terminal evaluation completion and terminal evaluation report submission for those that were completed in 2005 dropped by at least six months—a substantial improvement in performance.

Inadequate coverage of issues continues to be a problem in a significant proportion of terminal evaluation reports (see table 1.5). Some of the reports that provided performance ratings did not provide sufficient information and/or evidence to support those ratings. Thus, many of the reports do not comply with minimum requirement 3 specified in the GEF Monitoring and Evaluation Policy (GEF EO 2006a, p. 21).

Table 1.5

Submission of Terminal Evaluation Reports

	FY of report submission		
Factor	2004	2005	2006
Total number of terminal evaluation reports submitted	42	41	66
% without sufficient information on project outcomes	–	5	3
% without sufficient information on sustainability of outcomes	–	12	18
% that did not report on cofinance	29	44	29
% without sufficient information on M&E	25	20	30
% rated moderately satisfactory or above in quality	69	88	84

The outcome ratings provided by the evaluation offices of partner organizations are consistent with those provided by the GEF Evaluation Office. However, those provided in the terminal evaluation reports tend to be upwardly biased.

Since FY 2005, the Evaluation Office has been assessing the extent to which the project performance ratings provided by evaluation offices of the GEF Agencies are consistent with its own ratings. The analysis of the ratings for the FY 2006 cohort confirms last year's findings that the outcome ratings provided by the evaluation offices of the GEF Agencies are consistent with those provided by the Evaluation Office itself. While the Independent Evaluation Group (IEG) of the World Bank has been providing ratings for quite some time, this is the first year the UNEP evaluation office has provided ratings. Based on the assessment of differences between its ratings and those provided by the evaluation offices of its partner organizations, the Evaluation Office has decided that it can trust the outcome ratings provided by the IEG in the implementation completion report reviews. Beginning next year, when provided, the Evaluation Office will accept IEG outcome ratings. This is in accordance with the Office's efforts to collaborate with Agency evaluation offices and to prevent duplication of effort. The Evaluation Office is satisfied with the reviews of UNEP's evaluation unit. However, since only three reviews have been done so far, it is too early to use the UNEP evaluation unit's work as a basis.

There are major differences in the sustainability ratings provided by the Evaluation Office and the evaluation offices of the partner organizations. These differences are primarily driven by changes made by the Evaluation Office in its methodology to assess sustainability of outcomes. The Office will collaborate with its counterparts to facilitate convergence on this issue.

Both outcome and sustainability ratings provided in the terminal evaluation reports have an upward bias. For example, on a six-point scale, compared to the outcome ratings given by the Evaluation Office, those given in the terminal evaluation reports are on average higher by 0.7 points. In the analysis of ratings for the FY 2005 cohort, a similar bias was noted. Thus, candor in ratings provided in the terminal evaluation reports remains an area for improvement.

Management Action Records

The current MARs track management actions on 36 Council decisions. The Evaluation Office rated 33 percent of these decisions as having been adopted by management at high or substantial levels. For one decision, adoption was rated as negligible by both GEF management and the Evaluation Office. The Evaluation Office and management agreed on the rating on progress of adoption for 47 percent of decisions (17 of 36). For the other 53 percent, the Evaluation Office downgraded management's ratings. Many of the lower ratings given by the Office reflect the fact that the GEF Council has yet to approve proposals made to it.

Overall, the MARs demonstrate the effects of the "winds of change," in that Council decisions on older evaluations have been overtaken by recent changes, and many of the more recent decisions have led to proposals for Council consideration.

The one issue on which both GEF management and the Evaluation Office agree progress has been negligible is that of providing transparency on management information in the GEF. A Council decision in 2005 and a Council reminder in 2006 have not yet been adequately met by the Secretariat, which is fully aware of the situation. The Evaluation Office notes that making management information available in a transparent manner is not a complex issue, but instead requires sufficient human resources, energy, and dedication.

A complete version of this year's MARs are available on the GEF Evaluation Office Web site (www.gefeo.org).

1.3 Recommendations

Recommendation 1: UNDP and UNEP need to involve social and institutional expertise in project supervision.

The issue of inadequate attention to social and institutional considerations in GEF projects that had been raised by the evaluation of the Role of Local Benefits in Global Environmental Programs has been confirmed by the assessment of project supervision for UNDP and UNEP. These Agencies need to take steps to ensure that social and institutional issues are adequately supervised in GEF projects.

Recommendation 2: Special attention is required to ensure continued and improved supervision in the new GEF project cycle by ensuring adequate funding in project fees.

As the GEF moves forward with implementation of the Resource Allocation Framework and the new fee structure for Agencies, it is possible that current levels of supervision for some Agencies might be affected. The GEF Secretariat and Agencies need to take steps to ensure that quality of supervision is not negatively affected by these changes. The proposed revisions to the GEF Activity Cycle (in terms of approval procedures, processing time, and responsibilities) and to project modalities such as MSPs and FSPs should take into consideration requirements for proper supervision of GEF projects.

Recommendation 3: UNEP should develop a structural approach to supervision of its GEF portfolio.

UNEP is invited to identify ways in which it will address its relatively low performance in supervision. In recent years, improvements have been made since the appointment of a GEF portfolio manager, who has taken various actions to improve supervision and the quality of project reporting by program managers. However, the portfolio manager will need more structural support by higher levels of management to achieve more unified, consistent, and adequately funded supervision throughout UNEP's GEF portfolio.

Recommendation 4: All GEF Agencies will need to ensure that terminal evaluation reports include adequate information on sustainability of outcomes, quality of M&E systems, and cofinancing.

The terminal evaluation reports submitted to the Evaluation Office have shown little sign of improvement in reporting on sustainability of outcomes, quality of M&E systems, and cofinancing. The Agencies should take the steps necessary to meet minimum requirement 3 of the GEF Monitoring and Evaluation Policy (GEF EO 2006a), detailed in the Guidelines for Implementing and Executing Agencies to Conduct Terminal Evaluations. They should ensure that information on these issues is included in the terms of reference for terminal evaluations and that draft reports fulfill these terms of reference or adequately explain why they could not be met.

1.4 Issues for the Future

Compliance with Minimum Terminal Evaluation Reporting Requirements

Minimum requirement 3 of the GEF Monitoring and Evaluation Policy stipulates that all terminal evaluations need to assess project performance in achievement of outcomes, likelihood of sustainability of outcomes, and M&E and provide ratings for these. The policy was discussed with the GEF Agencies in the second half of 2005 and adopted in February 2006. However, most of the terminal evaluation reports submitted in FY 2006 pertain to evaluations that were completed before the policy was negotiated and implemented. This makes it difficult to assess the level of compliance by the GEF Agencies to the criteria specified in minimum requirement 3. The Evaluation Office expects that from next year on, most of the terminal evaluation reports submitted will be for evaluations conducted after the policy was adopted. Therefore, beginning with the APR for FY 2007, the Office will assess whether terminal evaluations meet minimum requirement 3.

Variation in Cofinancing Materialization

Although the reported cofinancing materialization is close to 100 percent at the portfolio level, there is high variance among projects. The Evaluation of Incremental Cost Assessment also found inconsistencies among projects in the criteria used to define cofinancing (GEF EO 2007a). To determine the reasons for high variance in materialization of cofinancing at the project level, and to ascertain the extent reported cofinancing is consistent with the manner in which it is defined by the Council, verification of actual levels of cofinancing is required. This issue will be looked at in the future.

Agency Performance Matrix

During negotiations for GEF-4, the GEF Council advocated for the development of a performance and outcomes matrix (scorecard) and required that the GEF Evaluation Office implement the following recommendation:

> The GEF Evaluation Office should report to Council through the Annual Performance Report on the performance of the GEF agencies on project-at-risk systems and the degree of independence and strength of the agency's evaluation functions, as well as on adherence to the minimum requirements for monitoring and evaluation. Furthermore, the Annual Performance Report should contain other key performance measures, to be developed into a performance and outcomes matrix in line with international methods and standards. The goal should be to set realistic and international best practice targets for ratings and achieve satisfactory ratings in all categories by 2010. Consistent with international best practice, the target for satisfactory outcome ratings should be 75% (GEF 2006, annex A, paragraph 32).

The matrix in annex E presents the Evaluation Office's response to the Council request. It covers 15 performance parameters, describes the current status of indicators and tools, and summarizes information sources and frequency of reporting in five major areas: results, processes affecting results, efficiency, quality of M&E, and learning. This matrix was discussed at the interagency meeting of February 2007, and comments and suggestions from Implementing and Executing Agencies and the GEF Secretariat have been incorporated into the present version. It should be noted that some of the measurement instruments have been developed in the context of the ongoing APR process; others are to be developed for future reports.

Each year, the Evaluation Office will circulate a draft of the performance matrix to the Agencies and ask for their comments. This feedback will be taken into account by the Office prior to finalizing the matrix for that year's APR. The Evaluation Office will inform Agencies of its reasons for any changes or decisions not to change ratings as a result of Agency comments; in cases where an Agency may disagree with such a decision, it will have the option of including a brief statement on the rating which will be included in an annex of that year's report.

Notes

1. The GEF fiscal year, like that of the World Bank, runs from July 1 to June 30. FY 2006, the primary

focus of this report, comprises the period from July 1, 2005, to June 30, 2006.

2. All dollar amounts are U.S. dollars unless otherwise indicated.

2. Scope and Methodology

2.1 Scope

Each year, the APR provides an assessment of the performance of completed GEF projects, analysis of the processes that affect accomplishment of results, and findings of the GEF Evaluation Office's oversight of project monitoring and evaluation activities across the portfolio. The APR also provides the GEF Council, other GEF institutions, and stakeholders with feedback to help improve the performance of GEF projects. Some issues are addressed by the APR annually, some biennially; others are addressed whenever there is a need to do so. The 2006 APR includes the following:

- **An overview of the extent to which GEF projects are achieving their objectives.** This overview consists of the Evaluation Office's assessment of the extent to which the completed projects for which terminal evaluation reports were submitted in FY 2006 achieved expected outcomes and the sustainability of those outcomes. Some changes have been made in the methodology used in the past to assess outcomes and their sustainability so that the assessments are more realistic. Also this year, the Evaluation Office carried out an additional analysis to assess the extent to which project outcomes that were rated moderately satisfactory or above in terms of their achievement were also rated moderately likely or above in their sustainability. The APR will continue to report annually on attainment of objectives and outcomes, and on sustainability of outcomes.
- **Report on the materialization of project cofinancing and delays in project completion.** This APR reports on the extent to which cofinancing promised at the point of project endorsement actually materialized and on delays in project completion. The Evaluation Office will continue to report on these issues on an annual basis.
- **A detailed assessment of IA quality of supervision.** This APR assesses for the first time the quality of supervision on GEF projects implemented by the IAs. The assessment appraises quality of supervision for GEF projects that were under implementation during FY 2005 and 2006. It also assesses supervision systems across the IAs and other factors that may influence supervision quality. This issue will be looked at again in the future.
- **An assessment of the quality of project monitoring.** This includes an examination of the quality of project monitoring during implementation, a determination of the sufficiency of funding for M&E activities, and an assessment of the quality at entry of M&E arrangements of completed projects. Some changes have been made in the methodology that was used last year to assess quality of project monitoring so that assessments are more realistic.

The APR reports on these M&E issues on an annual basis.

- **An assessment of the quality of terminal evaluation reports submitted by the IAs to the Evaluation Office.** This assessment, which is presented annually, provides information by focal area and Agency.
- **A presentation of findings on management action records.** This assessment, which is presented annually, reviews and follows up on the implementation status of evaluation recommendations that have been accepted by management and/or the GEF Council.

2.2 Methodology

Ensuring Reliability and Timeliness of Terminal Evaluation Reports

The project terminal evaluation reports submitted by the IAs 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.

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.

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 project ratings in the terminal evaluation report.

The reviews are also examined by a peer reviewer with substantial experience in the review of terminal evaluations. The peer reviewer provides feedback, which is incorporated by the primary reviewer in subsequent versions of the review.

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

If the terminal evaluation reports provide insufficient information to make an assessment or verify the Agency ratings on outcomes, sustainability, or quality of project M&E systems, the Evaluation Office classifies the projects as "unable to assess" and excludes them from any further analysis on the respective dimension.

This process has several limitations, with the most pervasive being that it is ultimately based on the information provided by the terminal evaluation reports. The veracity of these reports could probably be ascertained through field verification, but this is not always practicable. The Evaluation Office instead seeks to enhance reliability by incorporating in its reviews any pertinent information that it has independently gathered as part of other evaluations. Additionally, the Office looks for opportunities for targeted field verification, including setting aside time for field verification of projects during country visits carried out for other thematic evaluations. The Office will also carry out field evaluations when the findings of the terminal evaluation review or targeted field verification of a project deem independent evaluation necessary.

Another way to address the reliability concerns pertaining to terminal evaluation reports is to work with the GEF partner 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 meets most of the concerns of the Evaluation Office. The Bank's Independent Evaluation Group conducts desk reviews and verification of all implementation completion reports, which are produced by management. The IEG also carries out field verifications for 25 percent of World Bank operations. The GEF Evaluation Office has monitored IEG ratings over the last three years and has found only minor differences in the ratings it assigns versus those given by the IEG. The Office has therefore decided that, whenever provided, it will accept the IEG's validation of outcome ratings. During FY 2006, UNDP and UNEP continued their efforts to involve their central evaluation groups more directly in the evaluation of GEF projects; UNEP also worked to strengthen the independence of its evaluation group. Beginning this year, UNEP provided ratings and commentary on its completed GEF projects. The Evaluation Office is satisfied with the assessments of UNEP's evaluation unit. However, since only a few assessments have been completed so far, it is still too early to determine the overall reliability of the ratings. The Evaluation Office will continue its dialogue with the GEF partner Agencies, while reviewing their terminal evaluation reports and verifying their ratings.

Equally as important as the reliability of terminal evaluation reviews is the Evaluation Office's ability to access the reports for completed projects in a timely manner. The Office has put in place a system to track terminal evaluation report submissions and has created a database of the reports expected in any given year. Information from this database is sent to the Agencies for verification, and the Agencies are asked to submit terminal evaluation reports for completed projects or new dates for reports for extended projects. This tracking system includes all GEF projects with an original completion date since January 2001, which covers the majority of GEF projects.

The Evaluation Office has also been tracking the time between project completion and submission of terminal evaluation reports and between report completion and submission. Compared to the terminal evaluations completed in 2004, the average time lag between report completion and submission dropped by at least six months for those completed in 2005.

Limitations of the Data

The Evaluation Office used the F-test and chi square test to assess differences among groups of projects, and the findings reported here are significant at the 90 percent or higher confidence level. Regression analysis was used to assess the magnitude and direction of change associated with different variables. Nonetheless, the information obtained so far places some important limitations on the conclusions that can be derived. A large proportion of terminal evaluation reports failed to provide sufficient information on quality of project monitoring, sustainability of outcomes, and co-financing; thus, the corresponding projects are not included in the analysis for these issues. Additionally, because only a few years of data are currently available, trends cannot yet be inferred. These limitations will diminish in the coming years as the Agencies submit more terminal evaluation reports that comply with the GEF guidelines. Moreover, as the GEF project portfolio matures, the increasing number of terminal evaluation reports will permit more in-depth analysis. Larger and more reliable data sets will facilitate meaningful assessment of progress and allow the Evaluation Office to make comparisons across Agencies and focal areas.

Assessing Quality of Supervision

The Evaluation Office conducted a detailed study to assess the quality of supervision by IAs for the

GEF projects they implement. This study presents an assessment of the extent to which IA supervision systems are able to diagnose problems and address them. A stratified random sample was drawn from the GEF projects that were under implementation by the World Bank, UNDP, and UNEP during FY 2005 and 2006. Because of cost and time concerns, projects jointly implemented by IAs and/or by ExAs were not sampled; such projects comprise only 10 percent of the GEF portfolio, and their exclusion will not fundamentally alter the conclusions. Estimates of portfolio performance were made after excluding the jointly implemented projects and those implemented by the ExAs. The bias introduced due to stratification was corrected using probability weights.

MARs Assessment

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

Review of Findings

The preliminary findings of this report were presented to and discussed on various occasions with the GEF Secretariat and Agencies, and at an interagency meeting held in Washington, D.C., in February 2007. Individual reviews of project terminal evaluation reports and the results on quality of project supervision were also shared with the Agencies and the GEF Secretariat for factual verification.

3. Project Outcomes and Sustainability

This chapter discusses verified ratings on outcomes and sustainability for 66 projects (34 full-size and 32 medium-size projects) for which terminal evaluation reports were submitted in FY 2006. This is the second time the Evaluation Office has rated outcomes and sustainability.[1] Based on the experience gained during FY 2005 in rating project performance on achievement of outcomes and sustainability of outcomes, the Office felt that its approach to rating project performance needed further refinement to ensure more realistic ratings. The main limitation of the earlier methodology was that it did not account for the criticality of individual performance criteria but rather averaged the ratings for various criteria to determine an overall rating. Thus, a project could have a high rating despite having extremely low ratings for a particular critical criterion—for example, project outcomes could be rated as not sustainable due to severe financial risks but could still be rated as moderately likely if they faced negligible sociopolitical, institutional and governance, and environmental risks. The Evaluation Office therefore modified its approach for rating outcomes and sustainability of outcomes to address the issue of criticality of individual criteria. To maintain comparability with FY 2005 ratings, project outcomes and sustainability were also assessed using the previous methodology.

Eighty-four percent of the projects assessed this year were rated moderately satisfactory or above in terms of attainment of project outcomes. After controlling for the differences in rating methodologies, this performance is similar to that of the projects in the FY 2005 cohort, with both cohorts meeting the target, set during negotiations for the fourth GEF replenishment, of 75 percent of projects having satisfactory outcomes.[2]

Two-thirds (65 percent) of the projects assessed this year were rated moderately likely or above in terms of sustainability of outcomes. Sustainability ratings for World Bank projects were higher than those for projects of the other IAs. Three out of five projects (61 percent) had both moderately satisfactory and above outcome ratings and moderately likely or above sustainability ratings.

3.1 Rating Approach

The Evaluation Office rated project outcomes based on the level of achievement of project objectives and expected outcomes on a six-point scale. The criteria used to assess level of achievement included assessment of ex ante outcome **relevance**, actual **effectiveness** in achievement of outcomes, and **efficiency** in achievement of outcomes. While the same criteria were used in FY 2005 to provide the overall outcome rating, this year, relevance and effectiveness were considered to be critical criteria; the overall rating on achievement of outcomes could not be higher than the lower rating attained on either or both of the critical criteria.

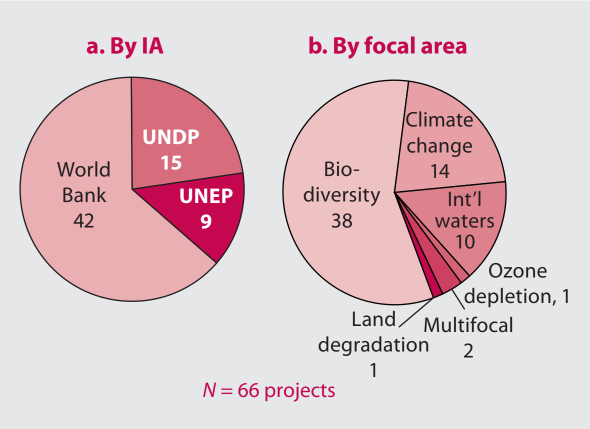
The Evaluation Office rated sustainability of outcomes based on an assessment of four key risk dimensions. Three of these dimensions—financial, sociopolitical, and institutional framework and governance—were also assessed last year; this year, the dimension of replication was replaced by environmental. Moreover, the focus of assessment shifted from ascertaining sustainability on each of these dimensions to ascertaining **risks** faced on each of these dimensions, which is believed to be more realistic to evaluate. And, unlike last year, when the overall rating was the average of ratings for individual dimensions, this year, all risk dimensions were regarded as critical. As with outcomes, overall ratings may not be higher than the lower rating on any of these dimensions. To make its rating scales comparable to those used by the evaluation offices of its partner organizations, and because assessment of risk on a six-point scale was difficult, the Evaluation Office shifted to a four-point scale in assessing sustainability of outcomes.

The Evaluation Office this year carried out an additional analysis to assess the extent to which outcomes of projects that were rated moderately satisfactory or above in achievement of outcomes were also rated moderately likely or above in sustainability.

Of the 66 terminal evaluation reports submitted, a large majority (64 percent) were for World Bank projects (see figure 3.1). In the FY 2005 cohort, by contrast, the majority of terminal evaluation reports were for UNDP projects (56 percent), with World Bank projects comprising only 29 percent of the total. This shift in IA project mix has been taken into account in assessing the differences in ratings between the two fiscal years. In terms of focal area representation, the FY 2006 cohort is similar to that for FY 2005; in both cohorts, the majority of projects were from the biodiversity focal area (see figure 3.1).

Figure 3.1

Number of Terminal Evaluation Reports by IA and Focal Area



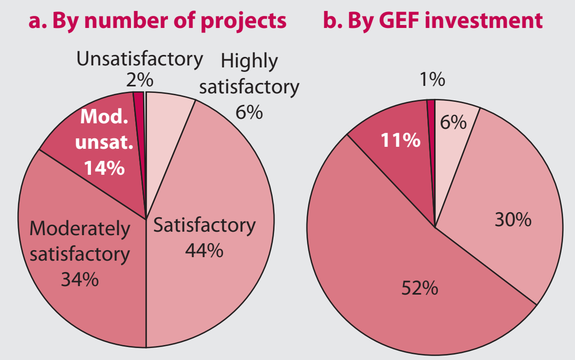
The Evaluation Office was not able to rate the level of achievement of outcomes for 2 projects and the level of sustainability of outcomes for 12 projects because insufficient information was provided in their terminal evaluation reports. Reports that did not provide sufficient information on a particular dimension were excluded from further analysis of that dimension.

3.2 Project Outcomes

Of the 64 projects whose outcomes were rated by the Evaluation Office, 84 percent were found to be moderately satisfactory or above (see figure 3.2). Similarly, of the total investment in the rated projects ($254 million), 88 percent was allocated to projects that were rated moderately satisfactory or above in outcomes (see figure 3.2). The ratings for FSPs and MSPs were similar: the outcomes of 85 percent of FSPs and 83 percent of MSPs were rated moderately satisfactory or above by the Evaluation Office. After accounting for changes in methodology, no difference was found between the outcome ratings of the cohorts for FY 2005 and 2006 (see figure 3.3).

Figure 3.2

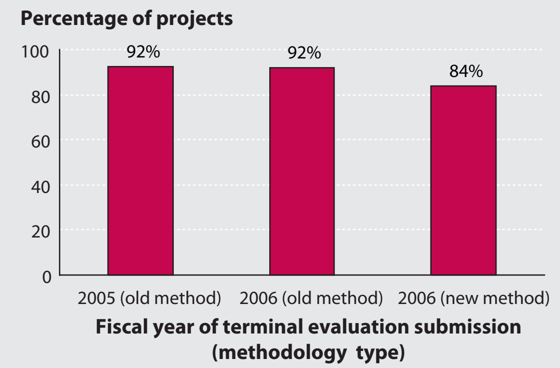
Project Outcome Ratings by Number of Projects and Project Funding



N = 64 projects; total GEF funding = $254 million

Figure 3.3

Proportion of Projects with Outcomes Rated Moderately Satisfactory or Above



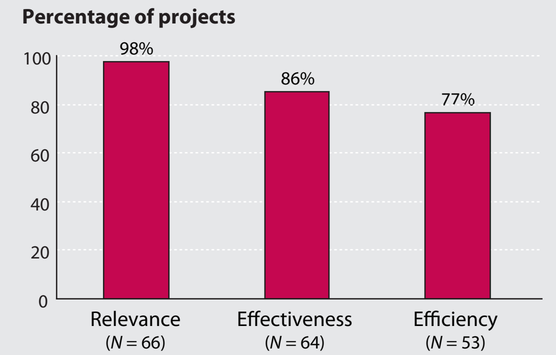
N = 39 projects in 2005; *N* = 64 projects in 2006

By outcome criterion, 98 percent of the rated projects were found to be moderately satisfactory or above in ex ante outcome relevance, 86 percent were rated moderately satisfactory or above in actual effectiveness in achievement of outcomes, and 77 percent were rated moderately satisfactory or above in efficiency of achievement of outcomes (see figure 3.4).

Outcomes of most of the World Bank projects, and of most of the biodiversity and climate change

Figure 3.4

Proportion of Projects with Outcome Dimensions Rated Moderately Satisfactory or Above



projects, were rated moderately satisfactory or above (see table 3.1). A slim majority of international waters projects—6 of 10 projects—had their outcomes rated moderately satisfactory or above. For other focal areas, the number of terminal evaluations reviewed was quite small.

The GEF investment in individual projects varied from $0.7 million to $35.0 million. Of the total investment in rated projects, $224 million (88 percent) was invested in projects whose outcomes were rated moderately satisfactory or above (see table 3.2). The GEF investment in projects rated moderately satisfactory or above varied by IA, with 91 percent of rated World Bank projects, 89 percent of UNDP projects, and 69 percent of UNEP projects being so rated by the Evaluation Office. By focal area, 89 percent of the GEF investment for biodiversity, 86 percent for climate change, and 84 percent for international waters was in projects whose outcomes were rated moderately satisfactory or above.

The Evaluation Office rated the outcomes of the following four projects as **highly satisfactory**, because they had fully achieved or exceeded the expected outcomes:

Table 3.1

Project Outcomes by IA and Focal Area

Number of projects

Rating	IA			Focal area						Total
	UNDP	UNEP	WB	BD	CC	IW	LD	MF	OD	
Highly satisfactory	0	0	4	4	0	0	0	0	0	**4**
Satisfactory	4	1	23	15	2	2	0	2	0	**28**
Moderately satisfactory	7	5	10	6	2	2	1	0	0	**22**
Subtotal	**11**	**6**	**37**	**25**	**4**	**4**	**1**	**2**	**0**	**54**
Moderately unsatisfactory	2	3	4	4	1	1	0	0	0	**9**
Unsatisfactory	1	0	0	1	1	1	0	0	0	**1**
Highly unsatisfactory	0	0	0	1	0	0	0	0	0	**0**
Subtotal	**3**	**3**	**4**	**6**	**2**	**2**	**0**	**0**	**0**	**10**
Unable to assess	1	0	1	7	8	4	0	0	1	**2**
Total	**15**	**9**	**42**	**38**	**14**	**10**	**1**	**2**	**1**	**66**

Note: BD = biodiversity; CC = climate change; IW = international waters; LD = land degradation; MF = multifocal; OD = ozone depletion.

Table 3.2

Project Outcomes by IA and Focal Area

GEF investment in million $

Rating	IA			Focal area						Total
	UNDP	UNEP	WB	BD	CC	IW	LD	MF	OD	
Highly satisfactory	0.0	0.0	14.7	14.7	0.0	0.0	0.0	0.0	0.0	**14.7**
Satisfactory	13.1	5.0	57.2	46.1	25.1	4.1	0.0	0.0	0.0	**75.3**
Moderately satisfactory	30.6	19.3	84.3	38.0	11.6	59.6	0.0	1.7	23.3	**134.2**
Subtotal	**43.6**	**24.4**	**156.1**	**98.7**	**36.7**	**63.7**	**0.0**	**1.7**	**23.3**	**224.1**
Moderately unsatisfactory	1.8	11.3	14.0	8.5	5.4	12.3	0.9	0.0	0.0	**27.0**
Unsatisfactory	3.1	0.0	0.0	3.1	0.0	0.0	0.0	0.0	0.0	**3.1**
Highly unsatisfactory	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	**0.0**
Subtotal	**4.9**	**11.3**	**14.0**	**11.6**	**5.4**	**12.3**	**0.9**	**0.0**	**0.0**	**30.1**
Unable to assess	0.8	0.0	0.8	0.8	0.8	0.0	0.0	0.0	0.0	**1.5**
Total	**49.2**	**35.6**	**170.9**	**111.1**	**42.8**	**76.0**	**0.9**	**1.7**	**23.3**	**255.8**

Note: BD = biodiversity; CC = climate change; IW = international waters; LD = land degradation; MF = multifocal; OD = ozone depletion.

- Valdivian Forest Zone: Private-Public Mechanisms for Biodiversity Conservation—an MSP in Chile
- Cape Peninsula Biodiversity Conservation Project—an FSP in South Africa
- Albarradas in Coastal Ecuador: Rescuing Ancient Knowledge on Sustainable Use of Biodiversity—an MSP in Ecuador
- Choco-Andean Corridor—an MSP in Ecuador

All of these projects were implemented by the World Bank, all were in the biodiversity focal area, all had a satisfactory monitoring and evaluation plan, all had implemented the M&E plan satisfactorily, and all had been executed in a timely manner.

The only project the Evaluation Office rated as **unsatisfactory** in terms of outcomes was the UNDP FSP Conservation of Biodiversity in the Lake Titicaca Basin. This project was closed after a 16-month extension, at which point many of its activities were yet to be completed. The project's failings were due to poor technical capacity of its hired staff and limited stakeholder participation.

The Evaluation Office rated the outcomes of nine projects as **moderately unsatisfactory**. All of these but one were so rated because of deficiencies in terms of actual effectiveness and efficiency in achieving outcomes. The exception—UNEP's international waters project on Persistent Toxic Substances, Food Security and Indigenous Peoples of the Russian North—was deficient in terms of ex ante outcome relevance. Although this project was endorsed under the contaminant-based operational program (OP10), which aims at demonstrating "ways of overcoming barriers to the adoption of best practices that limit contamination of the International Waters environment" (GEF 1996, paragraph 10.2), its objectives and expected outcomes focus on facilitating adaptation to toxic pollutants. Since projects aimed at facilitating adaptation have only recently been allowed in the GEF and then only within the climate change focal area, the project's ex ante objectives and outcomes were not relevant to GEF operational programs.

3.3 Project Outcomes Sustainability

Of the 66 terminal evaluation reports submitted in FY 2006, 12 (18 percent) did not provide sufficient information to allow assessment of sustainability of outcomes. Of the 41 submitted in FY 2005, 5 (12 percent) had not provided sufficient information. Thus, the extent to which outcomes of projects are sustainable cannot be determined for a substantial proportion of projects.

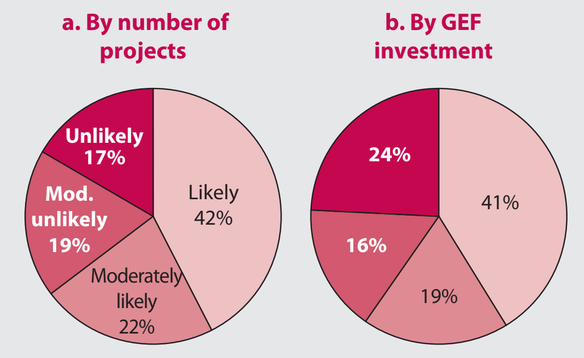
However, since insufficiency of information on sustainability of outcomes is not correlated with ratings on project outcomes and quality of project monitoring, it is likely that this shortcoming will not change the overall sustainability ratings for all completed projects reviewed in 2006. Of the 54 projects that were rated on sustainability of outcomes, 35 (65 percent) were rated moderately likely or above in terms of their sustainability (see figure 3.5).

In terms of GEF investment, 60 percent was invested in projects that were rated moderately likely or above in sustainability (see figure 3.5 and table 3.3). There was little difference between the sustainability ratings of FSPs and MSPs: 64 percent of the former and 66 percent of the latter were rated moderately likely or above.

Compared to 72 percent of the projects of the FY 2005 cohort, 84 percent of the projects of the FY 2006 cohort were rated, using the old methodology, as moderately likely or above in sustainability (see figure 3.6). The difference in the ratings of the two cohorts is not statistically significant; this

Figure 3.5

Sustainability Ratings by Number of Projects and Project Funding



N = 54 projects; total GEF funding = $218 million

Table 3.3

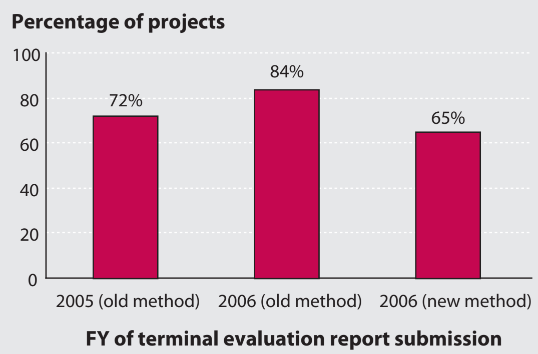
Sustainability of Project Outcomes by IA and Focal Area
GEF investment in million $

Rating	IA			Focal area						Total
	UNDP	UNEP	WB	BD	CC	IW	LD	MF	OD	
Likely	21.9	0.0	67.9	31.8	22.9	11.1	0.0	0.7	23.3	**89.8**
Moderately likely	8.2	18.2	14.4	14.4	8.2	18.2	0.0	0.0	0.0	**40.7**
Subtotal	**30.2**	**18.2**	**82.2**	**46.1**	**31.1**	**19.3**	**0.0**	**0.7**	**23.3**	**130.6**
Moderately unlikely	8.9	10.7	15.5	18.1	5.4	10.7	0.9	0.0	0.0	**35.1**
Unlikely	8.5	0.8	43.4	16.9	0.7	35.0	0.0	0.0	0.0	**52.7**
Subtotal	**17.3**	**11.5**	**58.9**	**35.0**	**6.1**	**45.7**	**0.9**	**0.0**	**0.0**	**87.7**
Unable to assess	1.8	6.0	29.8	30.0	5.6	1.0	0.0	1.0	0.0	**37.5**
Total	**49.2**	**35.6**	**170.9**	**111.1**	**42.8**	**76.0**	**0.9**	**1.7**	**23.3**	**255.8**

Note: BD = biodiversity; CC = climate change; IW = international waters; LD = land degradation; MF = multifocal; OD = ozone depletion.

Figure 3.6

Proportion of Projects with Sustainability Rated Moderately Likely or Above



N = 39 projects in 2005; *N* = 64 projects in 2006

is true for ratings on both a binary scale (likely/not likely) and the six-point scale.

Compared to projects from other Agencies, the sustainability ratings of World Bank projects were significantly higher (see table 3.4). Sustainability ratings of projects from different focal areas were comparable.

Projects that were rated moderately unlikely or lower in terms of their sustainability generally faced financial, sociopolitical, and/or institutional and governance-related risks. The outcomes of 11 percent of projects were rated as unlikely to be sustained due to sociopolitical risks (see figure 3.7). These projects were primarily those implemented in areas affected by a civil war such as the Aceh Landscape Elephant Project in Indonesia and the Energy Efficiency Market Development project in Côte d'Ivoire. The sustainability of one out of five projects (20 percent) was rated as moderately unlikely or lower due to financial risks.

Figure 3.7

Perceived Risks Underlying Projects Receiving Low Sustainability Ratings

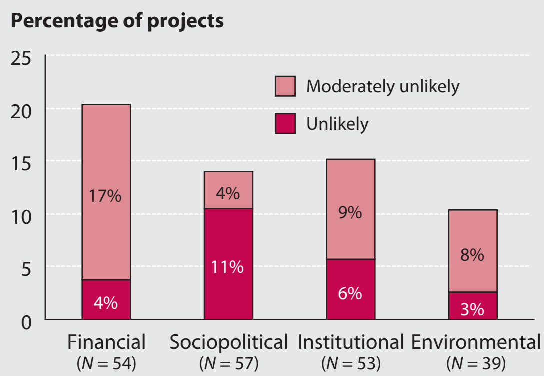


Table 3.4

Sustainability of Project Outcomes by IA and Focal Area
Number of projects

Rating	IA			Focal area						Total
	UNDP	UNEP	WB	BD	CC	IW	LD	MF	OD	
Likely	4	0	19	13	6	2	0	1	1	**23**
Moderately likely	3	4	5	5	3	4	0	0	0	**12**
Subtotal	**7**	**4**	**24**	**18**	**9**	**6**	**0**	**1**	**1**	**35**
Moderately unlikely	2	2	6	6	1	2	1	0	0	**10**
Unlikely	4	1	4	7	1	1	0	0	0	**9**
Subtotal	**6**	**3**	**10**	**13**	**2**	**3**	**1**	**0**	**0**	**19**
Unable to assess	2	2	8	7	3	1	0	1	0	**12**
Total	**15**	**9**	**42**	**38**	**14**	**10**	**1**	**2**	**1**	**66**

Note: BD = biodiversity; CC = climate change; IW = international waters; LD = land degradation; MF = multifocal; OD = ozone depletion.

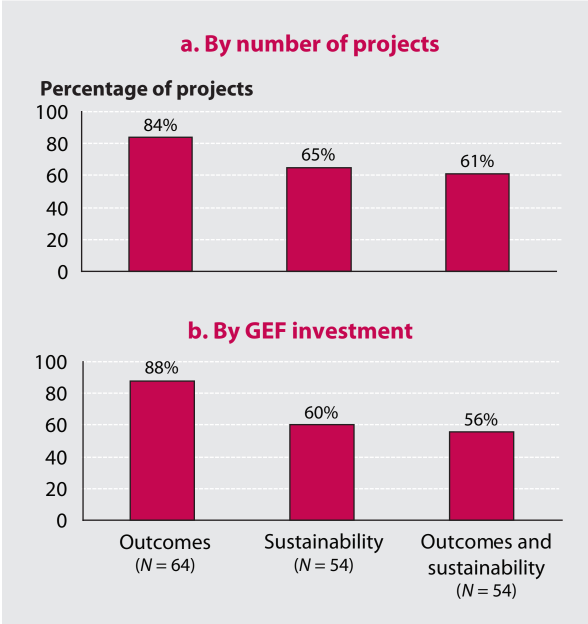
3.4 Outcomes and Outcomes Sustainability

For the first time, the Evaluation Office is presenting an assessment of the extent to which projects that are rated moderately satisfactory or above in achievement of outcomes are also rated moderately likely or above in sustainability of outcomes. Of the terminal evaluation reports submitted in FY 2006, 54 reports (82 percent) provided sufficient information to allow assessment of both project outcomes and sustainability of project outcomes. Of the rated projects from the FY 2006 cohort, 61 percent were rated moderately satisfactory or above in outcomes and moderately likely or above in sustainability (see figure 3.8). Financially, 56 percent of GEF funding for rated projects was invested in initiatives that were rated moderately satisfactory or above in outcomes and moderately likely or above in sustainability (see figure 3.8).

The Evaluation Office conducted the same analysis for the FY 2005 cohort. (Note, however, that the figures for FY 2005 are not directly comparable with those for FY 2006 due to changes in methodology.) Of the 41 terminal evaluation reports submitted for FY 2005, 36 (88 percent) provided sufficient information to allow assessment of both project outcomes and sustainability of project outcomes. Of the rated projects, 72 percent were rated moderately satisfactory or above in outcomes and moderately likely or above in sustainability. In terms of GEF investment, 86 percent was invested

Figure 3.8

Projects with Outcomes/Sustainability Rated as Moderately Satisfactory/Moderately Likely or Above



in projects that were rated moderately satisfactory or above in their outcomes and moderately likely or above in their sustainability. When the differences in the rating methodologies used in FY 2005 and 2006 are taken into account, the difference between the ratings of the two cohorts is not significant (see table 1.1).

Of the four projects that were rated highly satisfactory in achievement of outcomes, three were also rated as likely to be sustainable. The outcomes of the Choco-Andean Corridor project were rated as moderately unlikely to be sustainable because of environmental risks. Specifically, construction of an oil pipeline and charcoal production in the project area pose a threat to sustainability of outcomes of this project.

Notes

1. As used here, *sustainability* is the likelihood of continuation of project benefits after a project has been completed (GEF 2000).

2. Despite their meeting this target, the completed projects assessed during FY 2005 and 2006 do not pertain to the fourth replenishment and therefore are not subject to this requirement.

4. Factors Affecting Attainment of Project Results

The specific topics addressed in the processes segment of the APR vary from year to year; this year continues last year's discussion of materialization of cofinancing in GEF projects, albeit in less detail; it also looks at project completion delays.

Individual GEF projects vary in terms of the proportion of cofinancing that was reported to have materialized. However, at the portfolio level, the promised cofinancing tends to materialize. Performance of the FY 2006 cohort was comparable to that of past years. As promised in last year's APR, the Evaluation Office looked at the relationship between project funding and project outcomes. Last year, the Evaluation Office reported that even though prima facie there was a correlation between project outcomes and cofinancing ratios, this correlation was driven by two outliers. With the inclusion of FY 2006 data, it can be said with greater confidence that there is no relationship between cofinancing ratios and project outcomes. Further, the analysis continues to show that cofinancing ratios are not correlated with sustainability of project outcomes.

The average project completion delay was 13 months for the FY 2006 cohort, compared to 19 months for FY 2005. Only 17 percent of the projects in the FY 2006 cohort experienced delays of two years or more, compared to 44 percent for the FY 2005 cohort. It is too early to assess whether this drop indicates a trend.

The 2005 APR analysis suggested that outcome ratings could be correlated with project completion delays. Inclusion of the FY 2006 cohort data allowed the Evaluation Office to probe this issue further. It was found that, although the outcome ratings continue to be inversely related with completion delays, the relationship is quite weak. The relationship weakens further when other variables such as Agency type and focal area are controlled for. Similarly, the data for this year's cohort do not confirm a relationship between sustainability ratings and project delays.

4.1 Materialization of Cofinancing

Tracking the materialization of promised cofinancing is important because project activities are budgeted with the expectation of such cofinancing. Of the FY 2006 cohort of 66 projects, only 47 (71 percent) provided information on materialization of cofinancing. Since comparable information is not available for a sizable percentage of the completed projects, the actual cofinancing materialization for the whole portfolio could differ from the results presented here.

Total Promised Cofinancing and Actual Cofinancing

For the 47 projects from the FY 2006 cohort providing information on cofinancing materialization, $2.10 in cofinancing was promised per $1.00 of approved GEF grant, and $2.40 in cofinancing was reported to have materialized. In comparison, for the 23 projects from the FY 2005 cohort, $1.50 in cofinancing was promised per $1.00 of approved GEF grant, and $1.60 was reported to have materialized. Although both the cofinancing promised per dollar of approved GEF grant and the cofinancing that actually materialized increased for the FY 2006 cohort as compared to 2005, the difference is not significant since cofinancing ratios have a tendency to fluctuate. Figure 4.1 illustrates how the cofinancing raised per project has fluctuated for different cohorts of completed projects based on the year of submission of terminal evaluation reports.

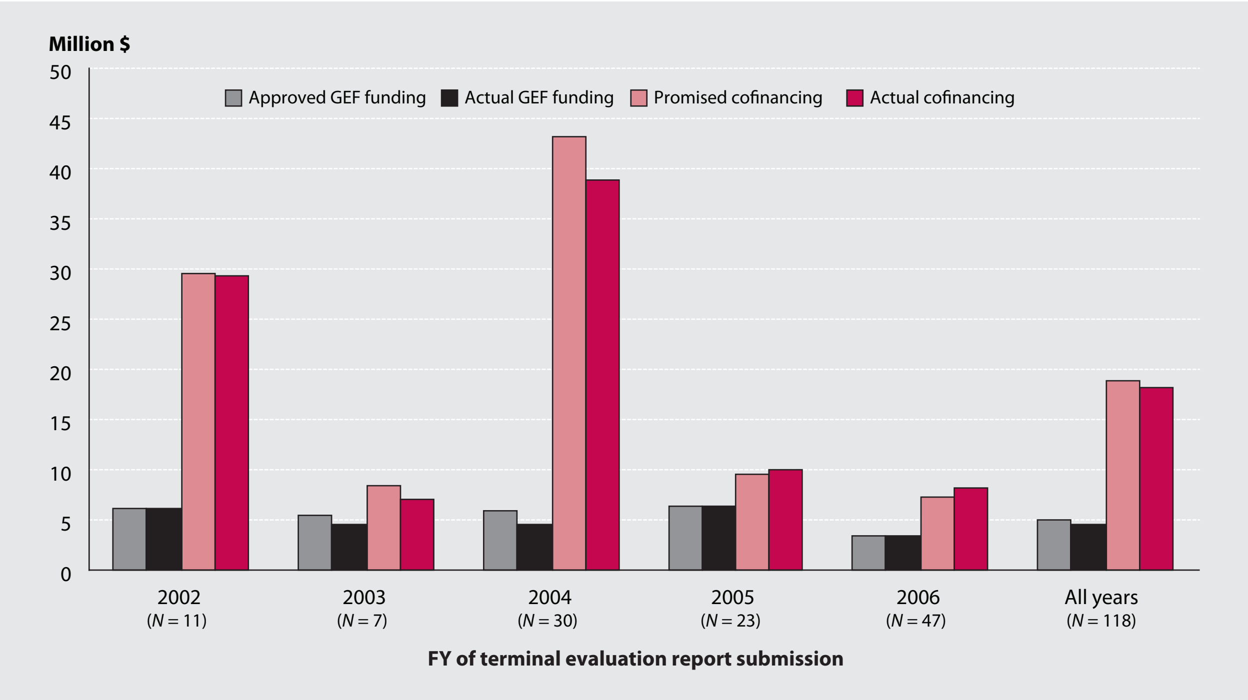
For almost two-thirds of the FY 2006 project cohort (64 percent), the reported materialized cofinancing was either less than 80 percent or more than 120 percent of that promised; this was the case for slightly over half (52 percent) of the FY 2005 project cohort. Thus, although at the portfolio level there is little difference between promised and reported materialized cofinancing amounts, there is significant variation across projects.

Trends over Time

Of the 182 terminal evaluation reports that have been submitted to the Evaluation Office since 2002, 118 provide information on cofinancing. Analysis indicates that, for the most part, promised cofinancing tended to materialize for these projects (see figure 4.1). For every $1.00 in GEF funding, $3.80 in funding from other sources was promised by the project proponents. Note, however, that these figures are skewed by the China Sichuan Gas Transmission and Distribution Rehabilitation Project, a climate change project implemented by the World Bank; for this initiative, GEF funding was $10.0 million and promised

Figure 4.1

Reported Materialization of Cofinancing: Average per Project



Note: Data for 2004 include the multimillion-dollar China Sichuan Gas Transmission and Distribution Rehabilitation Project.

cofinancing was $935.2 million, for a cofinancing ratio of 93.5. Excluding this project from the analysis reduces the cofinancing ratio to $2.20 per $1.00 of GEF grant.

The terminal evaluation reports submitted by the World Bank are more likely than those from the other Agencies to report on cofinancing: 83 percent of the terminal evaluation reports from the World Bank, compared to 52 percent from UNEP and 44 percent from UNDP, report on cofinance (see table 4.1). For the 75 World Bank projects for which cofinancing was reported, $4.30 in cofinancing was promised per $1.00 of GEF grant; 98 percent of promised cofinancing was reported to have materialized. However, much of this ratio is driven by the China Sichuan Gas Transmission and Distribution Rehabilitation Project. If this project is excluded from the analysis, the cofinancing promised per $1.00 of GEF grant approved drops to $2.30, with cofinancing materialization remaining at 98 percent. For the 28 UNDP projects and 14 UNEP projects for which cofinancing was reported, promised cofinancing per $1.00 of GEF grant was $1.00 and $1.50, respectively. Ninety-one percent of the cofinancing promised by UNDP, and 80 percent of that promised by UNEP, was reported to have materialized.

By focal area, 96 percent of the promised cofinancing materialized for 70 biodiversity projects, and 93 percent for 27 climate change projects, for which cofinancing was reported. The 13 international waters projects for which cofinancing was reported had a collective materialization of 144 percent; this was primarily driven by the Lake Ohrid Management Project, for which the reported materialized cofinancing was $80.5 million compared to a promised $21.3 million. Excluding this project from the analysis drops cofinance materialization to 95 percent.

For MSPs (46 projects), cofinancing of $1.60 was promised and $1.80 was reported to be materialized per $1.00 of GEF grant approved (110 percent materialization). For FSPs (72 projects), cofinancing of $3.90 was promised and $3.80 was reported to have materialized per $1.00 of GEF grant approved (96 percent materialization).

The increase in number of projects for which information related to cofinancing is available has allowed the Evaluation Office to assess with greater

Table 4.1

Materialization of Cofinancing as Reported in Terminal Evaluation Reports

Factor	UNDP	UNEP	World Bank	Multi-Agency	Total
Total number of terminal evaluation reports	63	27	90	2	**182**
Number that did not report on cofinancing	35	13	15	1	**64**
Number that did report on cofinancing	28	14	75	1	**118**
Approved GEF grant per project (million $)	2.6	2.9	6.1	20.2	**5.0**
Actual GEF grant per project (million $)	2.4	2.9	5.5	20.2	**4.6**
Promised cofinancing per project (million $)	2.7	4.4	26.2	134.8	**19.0**
Promised cofinancing per $1.00 of approved GEF grant	1.00	1.50	4.30	6.70	**3.80**
Reported materialized cofinancing per project (million $)	2.4	3.5	26.8	103.6	**18.3**
Reported materialized cofinancing per $1.00 of approved GEF grant	0.90	1.20	4.20	5.10	**3.60**
Materialized cofinancing per $1.00 of promised cofinancing (%)	91	80	98	77	**96**

certainty the leverage and actual materialization of cofinance for completed GEF projects, but the implications of cofinancing for project outcomes and sustainability are still not well understood. This information gap also has to do with the difficulty of determining the extent to which GEF grants are attracting new funding and/or are only attracting money that was already available through other sources that finance environmental projects. The Evaluation Office will attempt to better understand these relationships for next year's APR.

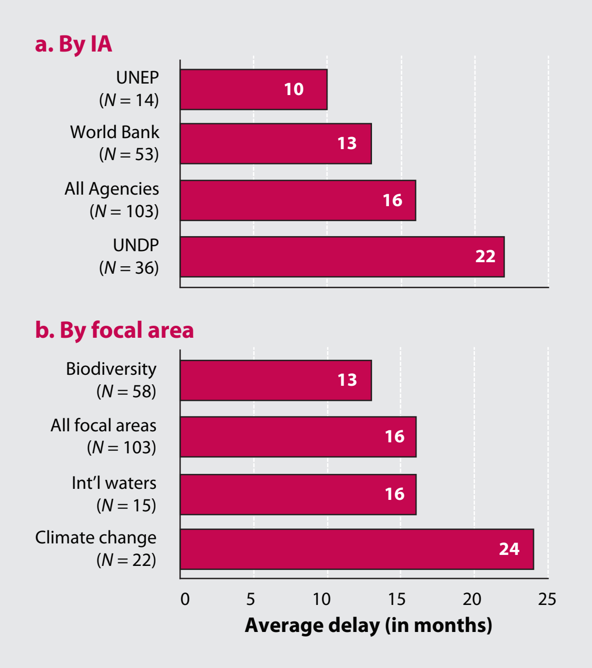
4.2 Delays in Project Completion

The Evaluation Office measured the time difference between expected closing at project start and actual closing. The average project completion delay was 13 months for the FY 2006 cohort, compared to 19 months for the FY 2005 cohort. Only 17 percent of the projects in the FY 2006 cohort, compared to 44 percent for the FY 2005 cohort, had delays of two years or more. For both cohorts, outcome ratings decreased with increases in completion delay.[1] This relationship is not confirmed when controlling for other variables, such as Implementing Agency and focal area. The data for the 2006 cohort also do not confirm a relationship between sustainability ratings and project delays.

When considering all 103 projects from the FY 2006 and 2005 cohorts for which data on completion delays are available, GEF projects have an average 16-month delay in completion, and 28 percent had a completion delay of two years or more. By Agency, UNDP's projects have significantly longer delays than those for the other Agencies, averaging 22 months from scheduled to actual completion (see figure 4.2). As discussed in chapter 5, one explanation for UNDP's longer reported completion delays could be that some of its projects, even though they were operationally closed, were not considered closed for accounting purposes. Although the average delay for UNEP projects is the lowest (10 months), the delays are not significantly different from those of other Agencies. By focal area, climate change projects experience significantly longer delays compared to those in the other focal areas, with an average delay of 24 months between planned and actual completion (see figure 4.2).

Figure 4.2

Average Delay in Project Completion by IA and Focal Area



Note

1. The relationship was polynomial (time delay squared).

5. Pilot Assessment of Project Supervision

The desk review of supervision follows the 2005 APR reviews of risk assessment by Implementing and Executing Agencies and of quality at entry of M&E by IAs. This is the first year that the Evaluation Office has conducted a supervision review, and it has been carried out on a pilot basis. The purpose of the review was not to assess the projects themselves, but to determine how well the IAs are conducting their supervision of GEF projects. Thus, an Agency may demonstrate strong supervision of a weak project and vice versa. The assessment takes into account not only the work of an individual program manager,[1] but the overall level of supervisory effort shown by the Agency (including fiduciary and management attention). **Supervision is here defined as the identification and tracking of, and response to, risks and other issues affecting project implementation and achievement of project objectives.** The focus is on learning about IA systems of supervision and benchmarking for future GEF monitoring of Agency performance. The review was conducted in response to a Council request for the Evaluation Office to look into these questions.

5.1 GEF Supervision Requirements

Several GEF documents have addressed issues related to supervision. Notably, the policy recommendations for the GEF's fourth replenishment included statements on the financial responsibilities of recipient Agencies, and the GEF Trustee was directed by the Council to prepare proposals on measures for strengthening Agency fiduciary standards. Examples of specific **standards** to be considered in line with international best practice include the following (GEF 2005b, paragraph 22):

- Independent oversight, audit and evaluation, and investigation functions
- External financial audit
- Financial management and control frameworks
- Project appraisal standards, including environmental assessments and other safeguard measures, as appropriate
- Monitoring and project-at-risk systems
- Procurement
- Financial disclosure
- Hotline and whistleblower protection, and codes of ethics

The GEF Council has established **minimum requirements** for supervision of project implementation, in parallel with Agencies' own due diligence policies and practices. There are slight variations within different Council documents as to what is expected with regard to supervision; the following is guidance from the 2006 Council-approved GEF Monitoring and Evaluation Policy, along with relevant Council guidance from previous years.

> Project fees allow Implementing and Executing Agencies to provide project cycle management services related to the GEF projects they manage. These services include portfolio development and management by regional and operational units, project identification, assistance to recipient countries in their project development and preparation, appraisal of project proposals and negotiation of GEF cofinanced operations, **supervision of GEF projects**, preparation of implementation completion reports, and reviews by the respective agency's evaluation office (GEF EO 2006a, paragraph 39; emphasis added).

Earlier GEF guidance (GEF 2004) set out the following implementation requirements:

- Mount at least one supervision mission per year, including briefing operational focal points on project progress.
- Provide technical guidance, as necessary, for project implementation.
- Pay advances to the executing agency and review financial reports.
- Prepare annual project implementation reports for submission to the GEF Evaluation Office.
- Prepare and participate in PIRs.
- Monitor and record project expenditure reports.
- Prepare periodic revisions to reflect changes in annual expense category budgets.

GEF (2002) similarly attempted to encapsulate minimal requirements of supervision:

- Supervision of project implementation (including procurement/disbursements)
- Project cost management and reporting
- Project progress review and reporting
- Preparation of appropriate interim evaluations and reviews, in accordance with Agency operational policy/practice
- Preparation of an annual PIR

Beyond these minimum requirements of the GEF, **IA/ExA internal policies** on project supervision have been evolving in response to growing attention to accountability for the use of donor funds. For example, the World Bank has devoted extensive attention to supervision requirements and procedures since the establishment of its Quality Assurance Group, which carried out a baseline study of supervision quality in FY 1997. In that year's assessment, it was found that just under two-thirds (63 percent) of all Bank projects were supervised in a satisfactory way; supervision of the remaining 37 percent was considered deficient with regard to existing Bank policies and standards (World Bank 1997). This finding underscored the need to address an internal "culture of approvals" which overemphasized the development of new projects at the expense of careful attention to the many implementation issues faced by ongoing operations. In recent years, several other development banks have launched their own internal reviews of project quality at entry and quality of supervision in response to guidance from their donors and executive boards.

The quality of World Bank supervision has since stabilized to about a 90 percent satisfactory rating, though there have been areas of lagging performance (for example, M&E, candor and realism of reporting, and so on). Reviews by the World Bank's Independent Evaluation Group have shown that those Bank projects that receive good supervision are twice as likely to have positive outcomes as those with less satisfactory supervision. These findings have been consistent for several years, and were recently validated by the QAG based on evidence from some 850 projects that had been assessed by the QAG during implementation and independently by the IEG after closing. Box 5.1, which is based on QAG experience, describes some characteristics of good supervision that should be broadly applicable to Agencies overseeing GEF projects.

Box 5.1

What Constitutes Good Supervision?

- Annual supervision strategies are agreed between task teams and managers. These strategies set clear priorities, define expected results, and form the basis for supervision resource allocation.
- During supervision, the program manager identifies problems quickly, reports them candidly, and weighs possible solutions with an eye to the project's development objectives.
- Emerging problems are addressed proactively, incorporating global best practices adapted to country circumstances.
- Project design is modified, where necessary, to reflect the changing environment and lessons of experience in close consultation with key stakeholders.
- Financial and safeguard aspects are monitored closely to ensure compliance with agreed standards, and specialized staff or consultants are used when technical issues demand expert attention.
- Project-level efforts are reinforced, where necessary, by integrating systemic implementation issues into the overall country dialogue through broader exercises such as country portfolio reviews.
- Managers provide or ensure guidance and decision making and essential mentoring and coaching to staff, especially for new or less experienced staff and during task team transitions.

Source: WB QAG 2001.

In 2005, the QAG looked retrospectively at findings from supervision assessments of 25 World Bank GEF operations from 2000 to 2005, and found that about three-fourths of these projects had been adequately supervised. About half of the GEF operations were rated as satisfactory for quality at entry, but only 38 percent had realistic assumptions for time frame and implementation capacities. Noted the QAG, "GEF operations tend to begin with significant issues of complexity and realism of project design, which are further exacerbated by any shortcomings in the areas of management attention, task team continuity, skills mix, and time in the field" (WB QAG 2005, p. 57).

5.2 Assessment Approach

The GEF Evaluation Office review of supervisory quality was conducted in three phases.

- The first phase included detailed review of 15 full-size GEF projects by the World Bank's QAG (from Quality of Supervision in FY05–06 [QSA7], the most recent of the Bank's annual quality of supervision assessments).
- The second phase comprised a desk review of supervision of 40 randomly sampled UNDP and UNEP projects to complement those already selected for the QAG review. (Six of these UNDP/UNEP projects were later dropped from the review, as discussed below.)
- The third phase entailed conference calls with IA program managers and others knowledgeable about the projects to discuss implementation issues and what had been done in supervision during FY 2005 and 2006 to help address these issues.

Only IAs were included in this review, as they account for the largest share of the current portfolio, and resource constraints prevented expanding the scope of this initial assessment to ExAs. The distribution of the sample is presented in figure 5.1; figure 5.2 shows the distribution by Agency in the GEF portfolio. Interviews were conducted with UNDP (in New York) and UNEP managers (based in Nairobi and Mexico City) to learn about internal procedures related to project supervision, and a field visit was made to Mexico and Panama to meet with UNEP and UNDP country office and regional technical staff to better understand their roles in the supervision of GEF projects. In all, 123 supervision team members, managers, or other staff were interviewed from the three IAs: 61 during the QAG review for World Bank GEF operations, and 45 and 17 from the Evaluation Office review of UNDP and UNEP projects, respectively.

Figure 5.1

Project Sample Distribution by IA and Focal Area

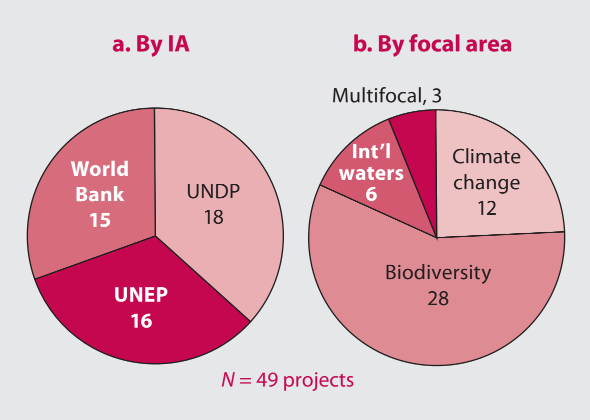
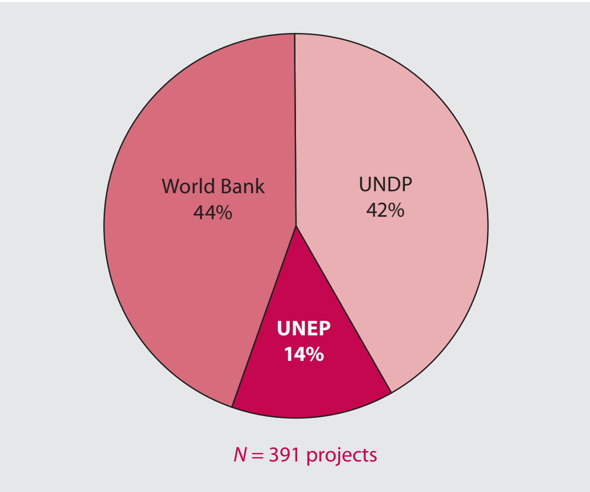


Figure 5.2

Agency Distribution in the GEF Portfolio



In carrying out the first phase of this assessment, the GEF Evaluation Office collaborated with the QAG during its conduct of QSA7. The Evaluation Office provided senior evaluation staff as panelists for the QAG assessment; in return, the QAG oversampled GEF projects (increasing the sample size from 10 to 15) in order to derive more robust findings on GEF-specific aspects of World Bank supervision. In the second phase, UNDP and UNEP policies and procedures on project supervision were reviewed, together with interviews of staff and managers, in order to better understand each Agency's internal requirements and practices as well as GEF requirements. An assessment methodology similar to that of the QAG was used to maximize comparability of findings. The sample was designed for an 85 percent confidence level with regard to IA findings, but was not expected to provide robust results by focal area or geographic region.

Several projects were dropped from the study and replaced by others when the IA involved pointed out that these projects did not entail supervision in the typical sense, but rather consisted of direct execution activities such as awareness raising, workshops, or preparation of training manuals for a particular operational program area.

The sampling process used in selecting the UNDP and UNEP projects indicated that the GEF Secretariat database is not up to date, since the sample produced six projects that ultimately had to be dropped either because they were not active during the period under review (four projects), were jointly executed (one project), or were direct execution projects (one project). All six were thus in violation of the selection criteria, reducing the sample from 40 to 34 projects. It was found that some projects in the database had been kept open for bookkeeping purposes (sometimes for a year or more); for others, the fact that the project had closed had not been recorded nor had the evaluation been forwarded to the GEF Secretariat or the Evaluation Office.

The World Bank's Business Warehouse experienced problems in accurately identifying GEF operations within the Bank's portfolio. Of 15 GEF projects in the QSA7 sample, only 10 were identified as GEF-financed in the Bank's internal project tracking database, while the other 5 were identified by Evaluation Office staff reviewing the QAG sample. In the World Bank's records, "blended" GEF projects are usually recorded as IBRD or

IDA operations, depending on the country, with the GEF component considered as cofinancing or trust funds. Previous years' supervision and quality-at-entry assessments by the QAG are likely to have missed GEF projects because of this anomaly; while Bank management is aware of the problem, no firm plans have been made to revise the system.

In all, 49 projects were sampled. To assess the performance at the GEF portfolio level, oversampling of UNEP and UNDP projects was corrected by adjusting for differential probabilities of projects being selected in the sample. The distribution and grant value of the final sample is presented in table 5.1.

Table 5.1

Distribution and Grant Value of Final Sample

Number and share	UNDP	UNEP	WB	Total
Number of projects in GEF portfolio	163	53	175	**391**
Share of GEF portfolio (%)	41.7	13.6	44.8	**100**
Number of projects in sample	18	16	15	**49**
Share of sample (%)	36.7	32.7	30.6	**100.0**
Grant value in sample (million $)	53.8	42.6	118.4	**214.8**
Grant value as % of total sample	25.0	19.8	55.1	**100.0**

UNDP and UNEP were requested to send supervision documents for the sampled projects during the period under review (FY 2005 to 2006). These included any field mission reports, midterm review or final evaluation documents, and audit reports as well as emails, management letters, or other documents pertaining to supervision issues that would help illustrate what had been done by the IA in supervising each project during the period under review. In addition, the Evaluation Office collected PIRs for FY 2005 and 2006 and original project documents. All projects reviewed, including those in the QAG assessment, are listed in annex C.

The documents collected were reviewed by a two-member panel, and UNDP and UNEP were contacted to provide any missing documents that would be needed to complete the desk review. PIRs for FY 2005 and 2006 were reviewed for internal consistency, quality of explanation of ratings, and thoroughness of risk assessment. Other documents were reviewed to determine the project time line and budget, the IA's supervisory activities, and the issues addressed by the program manager or other IA personnel during FY 2005 and 2006. Where available, midterm reviews and terminal evaluations were reviewed and any comments from the IA noted by the panel. Audits were briefly reviewed to note issues identified and the extent of follow-up on earlier audit recommendations. The panel did not attempt to examine financial management, procurement, or disbursement aspects in detail (this was done for the World Bank projects, whose QAG panels included fiduciary and safeguard specialist reviewers). Thus, an important caveat concerning this review is that the World Bank findings can be considered more robust with regard to projects' financial and due diligence aspects, and should not be compared with the more indicative findings for UNEP and UNDP projects.

The IA staff members responsible for task management were interviewed via conference call to better understand supervision during the period under review, and to discuss the general progress of and challenges facing each project. Finally, the review panel completed a supervision questionnaire for each UNDP and UNEP project, with ratings given across four major quality dimensions and numerous detailed subcategories (see annex D). The ratings were assigned using the same six-point rating system used in the FY 2006 PIR, and following the same format used by the World Bank QAG, adapted for relevance to the GEF (the World Bank projects used the standard QAG questionnaire). In this rating system, scores of 1, 2, and 3 are considered to be in the satisfactory range with the performance variations described in box 5.2; scores of 4, 5, and 6 are in the less-than-satisfactory range, also with specified variations as described in the box.

Box 5.2

Guidance Criteria for Quality Ratings

- **Highly satisfactory—**an exemplary supervision effort, demonstrating good practice in several areas, with highly proactive identification and resolution of threats to achieving the project's objective
- **Satisfactory—**satisfactory or better on all aspects; a solid supervision effort defined by sound and timely focus on implementation problems and project impact
- **Moderately satisfactory—**satisfactory on key aspects but some deficiencies and missed opportunities to strengthen prospects of project outcomes and the Agency's fiduciary role
- **Moderately unsatisfactory—**significant deficiencies in a few key aspects, which may jeopardize project outcomes and/or undermine the Agency's fiduciary role
- **Unsatisfactory—**significant deficiencies in several key aspects of supervision that jeopardize project outcomes and/or undermine the Agency's fiduciary role
- **Highly unsatisfactory—**a broad pattern of deficiencies in supervision that is likely to jeopardize project outcomes and/or undermine the Agency's fiduciary role

5.3 Supervision Quality Ratings

As part of the assessment process for supervision, the review panel began with a rapid appraisal of project quality at entry, giving attention to six aspects of the project context:

- Technical design quality
- Quality of the logframe or results framework and associated M&E arrangements
- Government commitment to the project
- Readiness for implementation at approval
- Implementation performance prior to the period under review
- Extent to which problems encountered had been identified at entry

World Bank IEG studies have shown that quality at entry is a major predictor of the likelihood of achieving project outcomes and objectives; it is also expected that a fundamental task of project supervision is to address design issues that become apparent during implementation in order to improve the prospects for success. Though the GEF has not formally established minimum standards for project design, the criteria and assessment methodology used by the QAG provide a framework that is broadly applicable to GEF projects as well, and which has been used in the past to review a substantial number of IBRD and IDA GEF operations. Table 5.2 presents the findings for design quality at entry in terms of those projects rated moderately satisfactory or better (that is, scoring 1, 2, or 3) for most or all of the design criteria.

Table 5.2

Percentage of Projects Rated Moderately Satisfactory or Better for Quality at Entry

Agency			Project size	
UNDP	UNEP	World Bank	FSP	MSP
31	29	80	54	41

The results presented in table 5.2 indicate that the World Bank performs at a high level for quality of project design, probably because of a combination of well-resourced project preparation teams together with a rigorous internal review procedure. Nevertheless, there is evidence that the Bank's GEF projects still have some design shortcomings, especially in low-capacity countries; and this year's QAG report to Bank management calls for more resources and more intensive supervision to address this (WB QAG 2007). The review of UNDP and UNEP projects revealed a broader pattern of design deficiencies; this point also emerged in interviews with program managers, many of whom indicated that their projects had had to grapple with pervasive design problems including overly ambitious or vaguely defined objectives, unrealistic time frames for implementation, poorly defined or absent monitoring indicators and other logframe deficiencies, and low levels of country capacity and ownership.

In previous reviews, the QAG had found that GEF projects lagged other World Bank operations for quality of supervision; thus, this year's results point to an encouraging improvement. UNDP appears to perform at about the same level as the World Bank, although much higher than its projects' quality-at-entry ratings (see table 5.3). That UNDP is able to achieve solid supervision despite the project design issues may be because it has offices in nearly every country, backed up by a network of focal area experts based in regional support offices. This system enables day-to-day communication with project partners by country office staff, while focal area experts can provide valuable technical advice on issues such as monitoring indicators or options for project restructuring along with a somewhat independent view of implementation progress and issues. On the whole, this multifaceted approach appears to be a very effective model, but one area of concern is whether the Resource Allocation Framework and revised fee system will lead UNDP management to look for ways to reduce staffing intensity.

Table 5.3

Percentage of Projects Rated Moderately Satisfactory or Above for Supervision

Agency			Project size	
UNDP	UNEP	World Bank	FSP	MSP
88	36	87	82	79

UNEP showed some much-needed improvements in its project supervision in 2006; for example, its PIRs were judged more candid and realistic in terms of their project progress ratings. But the evidence suggests that the Agency faces serious challenges due to a lack of clear policies and management attention to supervision, and to limited staffing and resources for field missions. UNEP also has a difficult portfolio of regional and global projects with weak institutional arrangements and poorly defined policy objectives, aggravated by a staffing profile that emphasizes scientific and technical expertise more than operational and supervision skills. Several UNEP staff members expressed concerns about the use of the GEF fee; UNEP staffing; and restrictions on field mission travel, especially for projects needing intensive supervision. It is important to point out that the assessment found cases of excellent supervision by UNEP, which illustrates that these obstacles can be overcome where program manager skills align well with the supervision needs of specific projects. Some program managers bring previous supervision experience from other Agencies, which has helped compensate for the lack of guidance and support from UNEP. On the other hand, even in these cases, the panel found evidence that individuals had had to press management to agree to additional time in the field, and that where program managers did not make this case, project problems showed a tendency to drift and accumulate. Table 5.4 provides the ratings of the sampled projects on supervision inputs, including management attention, supervision skills mix, and time in the field. This table also illustrates the particular problem faced by UNEP.

Table 5.4

Percentage of Projects Rated Moderately Satisfactory or Better for Supervision Inputs and Processes

Agency			Project size	
UNDP	UNEP	World Bank	FSP	MSP
88	43	87	82	82

Another area of concern has to do with critical review of project progress and the handling of midterm reviews. Several of the projects reviewed should have been candidates for restructuring or cancellation given that disbursements were lagging far behind or had some components that were performing very poorly, yet they were continued or extended even if their prospects for long-term sustainability seemed doubtful. The World Bank sometimes cancels nonperforming components of IBRD or IDA operations, which is considered a good business practice by both the Bank and the host country; however, this option is rarely exercised with respect to grant-financed operations such as those funded by the GEF. Several program managers shared their perception that the GEF prefers that projects approved by the GEF Council be carried through as designed, regardless of later developments.

In some cases, project cancellation should be taken as a real option, but there appears to be great reluctance to cancel even where performance has been very poor and long-term sustainability of the effort is doubtful. The panel could only speculate as to whether this was because of IA unwillingness to offend host countries or to return grant funds, or for other reasons; in any case, the Council should look to developing a clearer statement on the policy and criteria for canceling and/or reallocating funding for nonperforming grants.

Specific guidelines on handling problem projects would be helpful; these should address criteria for restructuring, when to consider cancellation, and whom to consult. A good first step in clarifying policy on cancellation is the Council decision arising from the 2006 Joint Evaluation of the GEF Activity Cycle: "At any stage of the project cycle, the beneficiary country, the Implementing or Executing Agency, or the Secretariat may recommend cancellation, termination or suspension of a project based on criteria approved by the Council" (GEF Council 2006, paragraph 24). A related issue to be resolved is the role of IA management in supporting, overseeing, and backstopping the supervision of GEF projects by program managers. Box 5.3 provides examples of good supervision practices that could be taken into account in this regard.

Within the sample, MSPs received substantially lower performance scores on supervision than did FSPs: only 65 percent of the MSPs in the sample scored moderately satisfactory or better for supervision inputs and processes. However, when weighted percentages were used to correct for the overrepresentation of some Agencies and the underrepresentation of others within the sample so as to reflect actual proportions in the GEF portfolio, it was found that the supervision ratings accorded MSPs were no different than those for FSPs. What can be concluded is that the poor supervision by UNEP and its disproportionate representation in the sample creates a downward biasing on results when the effects of each Agency's performance are not considered and controlled before extrapolating to the portfolio.

5.4 Project Reporting, Progress Ratings, and Risk Assessment

The PIRs for FY 2005 and 2006 generally present a positive outlook and give the impression that the

Box 5.3

Examples of Good Practice

Four projects were rated highly satisfactory for supervision, one each from the World Bank and UNEP, and two from UNDP: Energy Efficiency Project (Bulgaria, World Bank); An Integrated Ecosystem Approach to Conserve Biodiversity and Minimize Habitat Fragmentation in Three Selected Model Areas in the Russian Arctic (UNEP); Conservation, Restoration and Wise Use of Calcareous Fens (Slovak Republic, UNDP); and Conservation of Globally Significant Forest Ecosystems in Suriname's Guayana Shield (UNDP).

Three of these projects were FSPs, and one was an MSP; two were in biodiversity, one was in climate change, and one was multifocal. Three are quite recent projects, approved in 2004 or 2005; the other dates from 2000. Although the Bulgaria project was also rated highly satisfactory for quality at entry, the other three exhibited design problems, and two had substantial design shortcomings that needed to be addressed during supervision.

Characteristics of excellent supervision as noted by reviewers included the following:

- The task team has been extremely proactive, comprises staff knowledgeable about the requirements of the project and conversant with mistakes made in similar projects, and has had strong support from management.
- The program manager set up a monthly action list, and warning letters were sent when implementation performance was unsatisfactory. Management was notified of the problem, and major efforts were launched to turn around the project.
- Management in-country spends considerable time in supervising this project. The logframe and indicators were revised to better focus on the achievement of the global objective. Reporting reflects actual conditions and clearly communicates all relevant risks, progress toward achieving the objective, and implementation issues; and explains thinking and actions for follow-up.

GEF portfolio is healthy and relatively nonrisky, in spite of some implementation delays and similar issues. Most projects reviewed are rated fully satisfactory, with moderate or low risk overall. At-risk status is taken as a rating of moderately unsatisfactory or lower for objectives (called development objective by some IAs) or outcomes. Of 55 projects reviewed for PIR ratings (including those excluded from the supervision assessment for reasons described earlier), only 3 were rated moderately unsatisfactory or lower for objectives or outcomes in the FY 2006 PIR. These three projects—two FSPs and one MSP—consisted of one each conducted by the World Bank, UNEP, and UNDP, with a combined GEF grant value of about $10 million.

In contrast, the Evaluation Office identified 19 projects, with a total GEF grant value of $85 million, that appear to be at risk based on other evidence presented in the PIRs (narrative description of project issues, description of progress of key components, and so on). The evaluation also found inconsistencies in the reporting of risks in PIRs: in some cases, the risk sections had been left blank or the reported risk rating was inconsistent with the issues identified (for example, one project cited six critical risk areas but gave the overall risk as moderate). There was little difference noticed among IAs in this regard: all three demonstrated a significant tendency toward reporting overall project ratings that appear to be inconsistent with other evidence presented in the PIRs (see table 5.5). Concrete examples of discrepancies included projects rated fully satisfactory even though

- one or more components were described in the PIR as facing serious implementation delays;
- the project was described as struggling with design flaws, overambitious objectives, and an unrealistic time frame for implementation;

Table 5.5

Projects at Risk: APR versus PIR Analysis

Status	UNDP	UNEP	WB	Total
Sample size	24	16	15	**55**
At risk: APR	7	6	6	**19**
Grant value (million $)	23	16.8	46[a]	**85.8**
At risk: PIR	1	1	1	**3**
Grant value (million $)	3.5	0.9	5.6	**10.0**

a. Does not include IBRD/IDA components.

- chronic shortages of counterpart funding hindered implementation;
- serious deficiencies were reported in project management units or M&E systems, or on field sites;
- government policy changes and/or lack of commitment were undermining project objectives.

Figures 5.3 through 5.6 compare the at-risk status (for both development objective and implementation performance) of the APR sample as reflected in PIR ratings with the ratings of the Evaluation Office.

Table 5.6 presents ratings on the candor and realism of supervision reporting by the IAs. These findings are broadly consistent with QAG reports on candor and realism of internal project ratings, which points to a systemic problem of incentives and management oversight for project supervision in general. The GEF is understood to be in the business of risk, but more realistic appraisal of risk is needed to improve risk management and risk mitigation, and to channel additional supervision support to the riskiest operations.

The review also found that better guidelines are needed for the reporting process, beginning with some basic questions: Who is the end user? What are the incentives for candor? Proper communication of risks and sensitive issues is important, but the evaluation found some evidence of self-censorship in reporting. There were cases in which IA program managers withheld information from reports in order to reduce the chances of "interference" from the GEF Secretariat or the GEF Evaluation Office. There is also a need for clarity of roles—including the respective functions of IA management, the GEF Secretariat, and the GEF Evaluation Office—in the reporting process.

Table 5.6

Percentage of Projects Rated Moderately Satisfactory or Above for Candor and Realism of Supervision Reporting

Agency			Project size	
UNDP	UNEP	World Bank	FSP	MSP
75	29	80	77	60

Project supervision is apparently not relying sufficiently on social specialists. This echoes a finding of the evaluation of the Role of Local Benefits in Global Environmental Programs, which found that "relatively few [GEF] projects have matched their commitments to stakeholder involvement with a nuanced understanding of local social issues in a proactive manner or systematically drawn on social expertise in project design and implementation" (GEF EO 2006b, p. 123). That evaluation recommended that in order "to strengthen generation of linkages between local and global benefits, the GEF should ensure adequate involvement of expertise on social and institutional issues at all levels of the portfolio." In their management response, the GEF Secretariat and IAs maintained that while such might not have been the case during the GEF's early days, "today, it is a regular practice at every stage of the project cycle to involve appropriate expertise and tools related to social and institutional issues by all Implementing Agencies" (GEF EO 2006b, p. 161).

However, in its assessment of supervision, the Evaluation Office found that, for the sampled

Figure 5.3

Portfolio at Risk for Implementation Performance, PIR Ratings

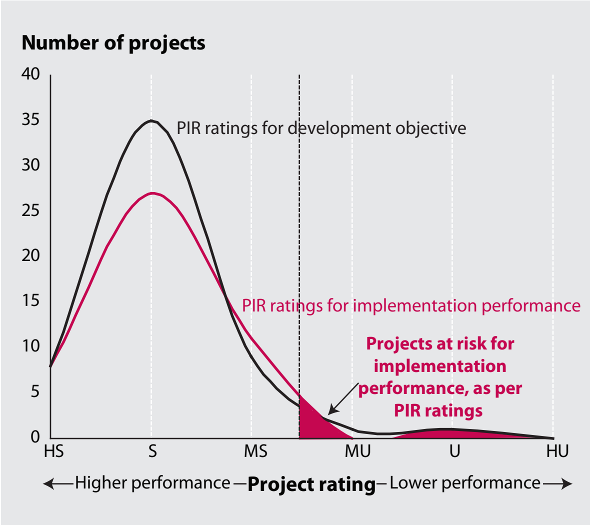


Figure 5.4

Portfolio at Risk for Development Objective, PIR Ratings

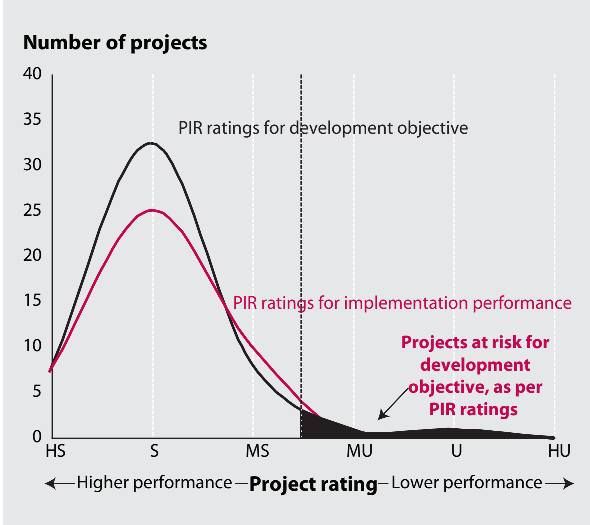


Figure 5.5

Portfolio at Risk for Implementation Performance, APR Ratings

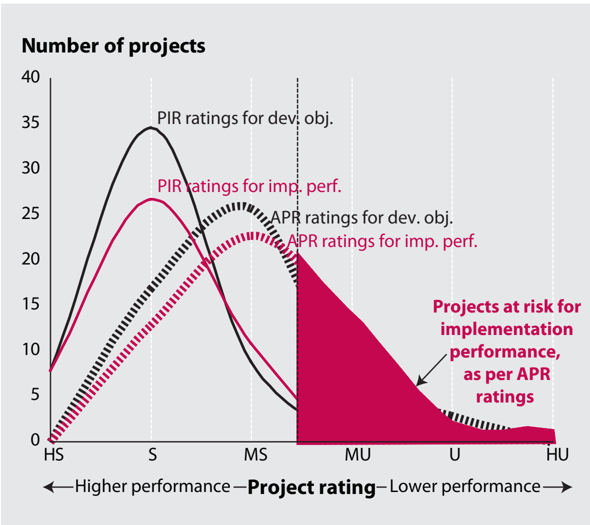
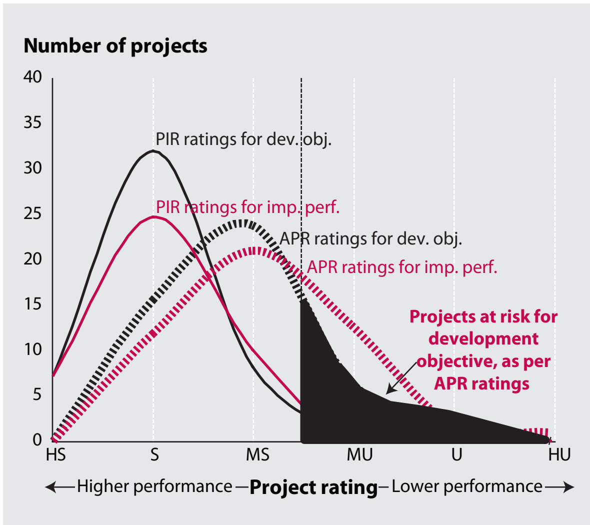


Figure 5.6

Portfolio at Risk for Development Objective, APR Ratings



projects, only the World Bank is involving such experts systematically. The Bank's system of "do no harm" safeguard policies requires all projects to be formally screened by specialists for potential safeguard issues and for mitigation plans to be developed (and supervised) where issues are identified. UNDP has prepared a paper on social issues, but there is no indication that it has been made operational in project supervision. UNEP has no paper and no actual practice of involving social and institutional experts in supervision of the sampled projects.

5.5 Portfolio Risk

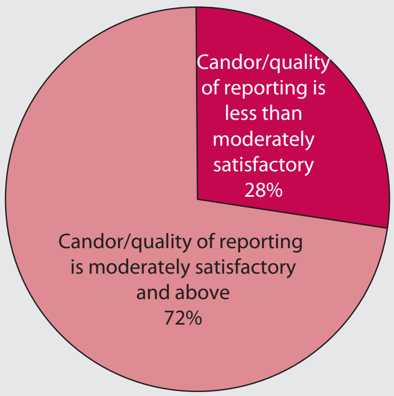
The narrative in some PIRs reviewed for this evaluation indicates that IAs give less weight to external factors (such as government administration changes), even though these may sometimes represent a significant risk to the project and especially to its long-term sustainability. Clearer guidance on the treatment of exogenous risk factors would be helpful to program managers. IA supervision staff expressed the perspective that reports should not "penalize" a project for the impacts of circumstances outside its control, which is understandable if the report rating is used to assess project staff performance. However, from the perspective of GEF management, all risks—whether internal or external—must be taken into account to supervise the GEF portfolio properly. Thus, guidelines on managing the accumulation of multiple risks are needed. Many GEF operations take place in countries with very limited technical and management capacity, not to mention complex climates of civil unrest and post-conflict transitions. At present, as long as no individual risk is rated as critical, there seems to be no objective basis for flagging a project as being in a higher risk category, even where a substantial number of moderate risks have been identified.

In several cases reviewed for this assessment, one or more of the risks identified during project design had manifested during implementation, with consequences for project performance—yet the overall risk rating had not been revisited. Management information systems such as UNDP's ATLAS and the World Bank's Structural Adjustment Program can incorporate and track a large number of discrete risk factors and automatically assign projects into a particular category (for example, "potential problem project") as soon as a predefined number of flags have been reached. But these systems, just like GEF managers, need relevant information in order to manage supervision responses. Even the World Bank's quite sophisticated portfolio monitoring system has been found to underrepresent portfolio risk because of staff reluctance or inability to report on projects from the perspective that includes all the information that is relevant and crucial to upper management.

This disconnect in information sharing is observed in the sample and is assessed by dimension 4, candor and quality of project performance reporting, of the supervision questionnaire. Projects reviewed in this assessment performed worst on this dimension, with 63 percent of the sample performing moderately satisfactory or better on it and 42 percent performing satisfactory or better. At the portfolio level, taking into account Agency proportional representation, this means that 72 percent of projects could be expected to perform moderately satisfactory or better (see figure 5.7), and 44 percent could be expected to perform satisfactory or better, in terms of candor of reporting.

Figure 5.7

Portfolio-Wide Ranking on Candor and Realism of Supervision Reporting



5.6 Supervision Fees

The evaluation looked at the use of supervision fees provided by the GEF to the Agencies. Table 5.7 provides a summary of fees paid for the projects in the FY 2005 and 2006 sample over the life of the project; it should be noted that about one-fifth of the sample entered the portfolio before the fee system was introduced. The amount of the fee ranged from $143,000 to just over $1.5 million. With one exception, the standard fee size for an MSP was $146,000. (The exception was an "emergency" UNDP MSP in the Galapagos Islands, which received $98,000.) The average fee size for an FSP was $571,000.

Table 5.7

Agency Fees for Projects in the APR Sample
Million $

Funding and fees	UNDP	UNEP	WB	Total
Total GEF funding	65.5	41.8	121.3	**228.6**
MSPs and FSPs	63.1	40.0	116.7	**219.8**
PDF	1.4	1.8	4.6	**7.8**
Cofinancing	140.4	37.8	341.2	**519.4**
IA fee	4.2	3.2	8.2	**15.6**
IA % of sample	26.9	20.4	52.7	**100.0**

Note: PDF = project development facility. Data do not include corporate fees. Some of the sampled projects entered the GEF portfolio before the institution of Agency fees.

The UNEP and UNDP program managers interviewed were not able to provide project-level supervision costs, and the panel was told that those Agencies do not yet have a system in place for charging staff time or travel costs against specific projects, although such a system is being planned. The World Bank has tracked staff supervision costs for many years; table 5.8 provides details on budgeted and actual supervision expenditures for the 15 World Bank projects in the sample. The costs shown in the table exclude preparation costs for projects that became effective during the period under review, which explains part of the large variation seen in year-to-year costs (in such cases, most of the costs would fall under project preparation rather than supervision).

Several points emerged from the review of the Bank's supervision costs. The World Bank assigns supervision coefficients for its projects based on a number of factors, such as past experience with the sector and country, and project size and duration. Additional resources may be added to provide for any specialized staff inputs (for example, for projects triggering safeguard policies such as resettlement, dam safety, indigenous peoples, or potential for environmental impacts). Most of these coefficients are established at the regional management level, and there are variations in practice from one region to another. One factor noticed during QSA7 was that some regions have established guidelines limiting task team expenditures on variable costs (travel and consultants); in several cases, this was cited by program managers as a factor limiting their ability to provide more intensive supervision of complex or challenging projects. This means that even where unexpended supervision funds may be available, they cannot be applied toward additional travel or consultant costs if this would exceed the variable cost share of the budget.

The Agency fee received from the GEF also covers a number of central unit costs which are taken off the top: predetermined portions are set aside for the evaluation unit, the legal department, the trust fund department, the loan department, and the treasury. There are also regional management overheads to be taken into account as well as adjustments for inflation. For FY 2006, the average supervision budget coefficient for a GEF-financed World Bank FSP was set at $78,000; this increased to $80,000 for FY 2007. The supervision budget for an MSP in FY 2006 was $16,000, which was raised to $18,000 in FY 2007.

Table 5.8

World Bank Supervision Costs for GEF Operations for 15 GEF Projects in QSA7 Assessment

Project name	Grant amount (million $)	Agency fee (thousand $)	Supervision budget (thousand $)			Supervision actual (thousand $)			Budget variance (%)
			FY 2005	FY 2006	Aver-age	FY 2005	FY 2006	Aver-age	
Bulgaria – Energy Efficiency	10.0	841.0	94	133	113.5	7	105	56	
Peru – Participatory Management of Protected Areas[a]	14.8	1515.0	66	51	58.5	72	73	72.5	124
Regional – Integrated Ecosystem Management in Indigenous Communities	9.0	1016.3	330	100	215	167	166	166.5	77
Burundi – Agricultural Rehabilitation and Sustainable Land Management[a]	5.0	437.0	0	60	30	2	117	59.5	198
Costa Rica – Ecomarkets[a]	8.0	878.0	53	48	50.5	61	61	61	121
Uruguay – Energy Efficiency	6.9	715.0	60	60	60	58	56	57	95
Tajikistan – Community Agriculture and Watershed Management	4.5	680.0	26	95	60.5	72	96	84	139
Tunisia – Gulf of Gabes Marine and Coastal Resources Protection	6.1	666.0	40	60	50	87	69	78	156
Vietnam – Demand-Side Management and Energy Efficiency Program	5.5	572.0	37	55	46	70	85	77.5	168
Guinea-Bissau – Coastal and Biodiversity Management	4.8	502.0	20	50	35	11	45	28	80
Bosnia-Herzegovina – Water Quality Protection	8.5	391.5	60	60	60	69	57	63	105
Pakistan – Protected Areas Management[b]	10.8		45	60	52.5	23	69	46	88
Serbia – Reduction of Enterprise Nutrient Discharges Project[b]	9.0		50	60	55	7	91	49	89
Mozambique – Coastal and Marine Biodiversity Management[b]	3.7		12	60	36	16	71	43.5	121
Morocco – Protected Areas Management[b]	10.1		0	50	25	69	72	70.5	282
Total	**116.7**	**8,213.8**	**893**	**1,002**	**947.5**	**791**	**1,233**	**1,012**	**107**
Project average	7.8	746.7	59.5	66.8	63.2	52.7	82.2	67.5	126

a. Blended project; amount given is GEF grant only.

b. No GEF fee.

World Bank management is concerned about the processing complexity of GEF operations, which are the only trust-funded operations within the Bank that are processed with the same requirements as those for large investment operations; Montreal Protocol operations, for example, are handled under simpler procedures. In addition, the Bank has concerns about implementation delays, which increase administrative costs. For this reason, regional management units have an incentive to withhold supervision expenditures as a hedge against likely project extensions. This sometimes creates a dilemma for task teams, which need to balance management pressure to control costs against the supervision needs of complex and innovative projects, especially in countries with weak technical and management capacity. The QSA7 review found that the adequacy of supervision resources for GEF operations was somewhat below the Bank average for FY 2005 and 2006.

5.7 Other Observations

Several other findings emerged during this pilot assessment of project supervision:

- One intent of **midterm reviews** is to enable program managers to be alerted to the necessity of restructuring or reorienting—or even canceling—projects in a timely manner. However, the evaluation found numerous examples of projects that had postponed their midterm review because of implementation delays. (In one case, the midterm review was held six months before project closing.) In the past, a GEF guideline held that a review should take place at the end of the second year of implementation. This guidance should be reinstated, even in cases of implementation delay, as it provides an opportunity for adaptive management and project rework. Several of the projects reviewed had waited too long to address fundamental problems, leaving program managers little time (extensions notwithstanding) to make up the lost ground. Guidelines for project revision should be streamlined to avoid the perception that restructuring imposes too heavy a bureaucratic burden on a project.
- **Fiduciary standards** are sometimes unclear. In one regional project reviewed, different country offices applied different thresholds for audit requirements; audit requirements for MSPs also vary across Agencies. The GEF Trustee is preparing a paper on fiduciary standards which may help resolve these questions.
- There is also a lack of clarity about **requirements and procedures for project changes**—what needs approval from the GEF Secretariat and what can be approved by the IA. In some projects reviewed, it appeared that rather modest changes took an inordinate amount of time on the part of project managers and partners, distracting them from more substantive supervision tasks. On the other hand, it is also evident that, in many countries, project partners are sometimes poorly prepared to meet the recordkeeping and reporting requirements of internationally financed operations, thereby placing an additional responsibility on supervision (and preparation) teams.

Note

1. The IA staff member responsible for supervision is variously termed a task manager, task team leader, or program manager. This report uses "program manager" throughout to refer to this staff member.

6. Quality of Project Monitoring

This chapter addresses the quality of GEF project monitoring. Section 6.2 continues the analysis begun in chapter 5 of quality of monitoring in completed projects during implementation. Section 6.3 explores the linkages of quality of project monitoring during implementation with quality of M&E plan at entry and with adequacy of funding for M&E.

Of the 66 terminal evaluation reports reviewed for FY 2006, only 46 provided sufficient information to allow the Evaluation Office to rate quality of project monitoring. Although a greater proportion of terminal evaluation reports are providing sufficient information on M&E than was the case in earlier years, the proportion that does not do so remains substantial (30 percent). This gap affects the extent to which the findings for the rated projects could be considered representative of the entire cohort.

Of the rated projects, 78 percent were rated as moderately satisfactory or above in quality of project monitoring. This positive rating for the vast majority of projects does not mean that difficult issues such as specification of appropriate indicators and provision of baseline information have always been addressed. More work is required in these areas.

The analysis found linkages between quality at entry of M&E arrangements and quality of monitoring during project implementation, and some evidence to support the case for better funding for M&E activities.

6.1 Rating Approach

The analysis used 107 terminal evaluation reports—41 from FY 2005 and 66 from FY 2006—submitted by the IAs to the Evaluation Office. Of these reports, 78 provided sufficient information to rate quality of monitoring during project implementation.

The Evaluation Office rates quality of project monitoring based on how well a project's M&E system was implemented. To do so, it 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.

The Evaluation Office provided an overall rating for quality of project monitoring during project implementation on a six-point scale. The methodology used for assessment of quality of monitoring for this year's APR is different from that used previously. In the 2005 APR, the Evaluation Office considered both quality of implementation of M&E system and usage of M&E for project management. For this year, these two criteria were combined to create an overall rating on quality of monitoring during project implementation. To determine the difference in findings resulting from this new methodology, the FY 2006 cohort was also assessed using the old methodology. Analysis showed that there was little difference between the M&E ratings for projects using the two methodologies. Therefore, in this report, the ratings for the FY 2005 cohort have been compared directly with those of the FY 2006 cohort, even though the methodologies used for the ratings are not identical.

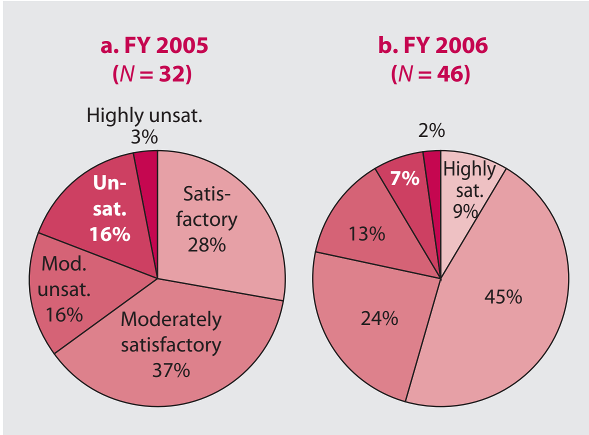
6.2 Overall Findings

As shown in figure 6.1, the proportion of projects rated moderately satisfactory or above on quality of monitoring was 66 percent for the FY 2005 cohort and 78 percent for 2006. Note that the proportion of terminal evaluation reports providing insufficient information was 22 percent in FY 2006. Thus, M&E information is unknown for a substantial proportion of GEF projects.

Among the 46 rated projects of the FY 2006 cohort, 84 percent of World Bank projects and all five UNEP projects had moderately satisfactory or above ratings (see table 6.1). By focal area, 81 percent of biodiversity projects were rated moderately satisfactory or above. For other focal areas, either the proportion of terminal evaluation reports that did not provide sufficient information is substantial (climate change and international waters), or the number of projects is too small (land degradation, multifocal, and ozone depletion) to allow for meaningful findings. By project size, 62 percent of full-size projects were rated as having moderately satisfactory or better project monitoring during implementation, as were 92 percent of medium-size projects—a surprising finding. A plausible explanation could be that MSPs are smaller and comparatively less complex, and therefore quality of monitoring could be relatively better. To ascertain whether this actually is the case, the Evaluation Office will continue to study this issue.

Figure 6.1

Quality of Monitoring during Project Implementation for FY 2005 and 2006



6.3 Factors Affecting Quality of Monitoring

In FY 2005, the Evaluation Office assessed quality at entry of M&E arrangements for the projects endorsed that year. Beginning with this year's APR, the Office is assessing quality at entry of the M&E arrangements for completed projects for which terminal evaluation reports have been submitted. It has also started assessing adequacy of funding for M&E activities and will assess the linkages of these variables with quality of monitoring during project implementation. Quality at entry of M&E arrangements was assessed on a six-point scale; adequacy

Table 6.1

Quality of M&E during Project Implementation by IA and Focal Area
Number of projects

Rating	IA			Focal area						Total
	UNDP	UNEP	WB	BD	CC	IW	LD	MF	OD	
Highly satisfactory	0	0	4	4	0	0	0	0	0	**4**
Satisfactory	5	4	12	15	2	2	0	2	0	**21**
Moderately satisfactory	0	1	10	6	2	2	1	0	0	**11**
Subtotal	**5**	**5**	**26**	**25**	**4**	**4**	**1**	**2**	**0**	**36**
Moderately unsatisfactory	2	0	4	4	1	1	0	0	0	**6**
Unsatisfactory	2	0	1	1	1	1	0	0	0	**3**
Highly unsatisfactory	1	0	0	1	0	0	0	0	0	**1**
Subtotal	**5**	**0**	**5**	**6**	**2**	**2**	**0**	**0**	**0**	**10**
Unable to assess	5	4	11	7	8	4	0	0	1	**20**
Total	**15**	**9**	**42**	**38**	**14**	**10**	**1**	**2**	**1**	**66**

Note: BD = biodiversity; CC = climate change; IW = international waters; LD = land degradation; MF = multifocal; OD = ozone depletion.

of funding was rated on a binary scale (adequate or inadequate).[1] Among the projects that were rated, the quality at entry of the M&E plan was rated moderately satisfactory or above for 59 percent.

Forty-two projects were rated both on the quality at entry of their M&E arrangements and on their quality of monitoring during project implementation. Of the 28 projects that were rated moderately satisfactory or above for the former measure, 27 were rated moderately satisfactory or above for the second measure as well; this is a correlation of 96 percent. Moreover, of the 14 projects that were rated moderately unsatisfactory or below in quality at entry of M&E arrangements, only 5 (36 percent) were rated moderately satisfactory or better in quality of monitoring during project implementation. Thus, better quality at entry of M&E arrangements seems to be associated with better quality of monitoring during project implementation.

Adequacy of Funding for M&E Activities

Twenty-eight projects were rated on adequacy of funding for their M&E activities. Of these, 27 projects had also been rated on quality of monitoring during project implementation. All 20 projects that were rated to have had adequate funding for M&E activities were also rated moderately satisfactory or above on their quality of monitoring during project implementation. In contrast, of the seven projects that were rated to have inadequate funding for M&E activities, only two (29 percent) were rated moderately satisfactory or above on quality of monitoring during project implementation. While these numbers are not sufficient to make broad generalizations, the direction of the relationship is consistent with the expectations of the Evaluation Office: that adequate funding for M&E is an important condition for satisfactory implementation of a project M&E system.

Quality at Entry of M&E Arrangements

Of the 66 projects in the FY 2006 cohort, quality at entry of M&E arrangements was rated for 49 projects. The remainder could not be rated because neither terminal evaluation reports nor project documents provided sufficient information.

Note

1. For some of these projects—especially those approved during the GEF pilot phase—project appraisal documents submitted for Chief Executive Officer endorsement were not available. For others, even though their project appraisal document was available, the annexes that were supposed to include the M&E plan were not.

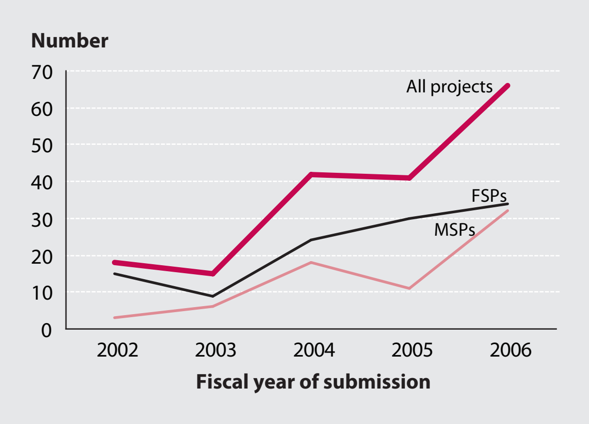
7. Quality of Terminal Evaluation Reports

Terminal evaluation reports that provide an accurate assessment of project accomplishments and shortcomings are not only essential in forming building blocks for the APR's assessment of project outcomes and sustainability but are also useful as a learning tool. To date, 182 terminal evaluation reports have been submitted to the Evaluation Office, and the number of reports submitted each year has been increasing (see figure 7.1). Two factors underlie this increase—the maturation of the GEF portfolio and more prompt submission of terminal evaluation reports.

This is the third year the GEF Evaluation Office has rated the quality of project terminal evaluation reports; 149 have been rated thus far. Since the methodology for assessing the quality of these has remained more or less stable over the period, the Evaluation Office is able to assess preliminary trends in changes in their quality.[1]

Figure 7.1

Number of Terminal Evaluation Reports Submitted for Review



An assessment of the terminal evaluation reports submitted in FY 2006 shows that there are many areas in which performance continues to lag. For example, a substantial proportion of the reports do not provide sufficient information to facilitate rating of sustainability of outcomes (18 percent) or quality of monitoring (30 percent). Similarly, 29 percent did not report on cofinancing. One reason for this lack of information could be an absence of clear guidelines at the time when most of these terminal evaluation reports were completed. For example, even though the GEF Monitoring and Evaluation Policy was negotiated with the IAs in the second half of 2005, most of the terminal evaluation reports that were submitted in FY 2006 pertained to terminal evaluations completed before July 2005. Based on minimum requirement 3 of the GEF Monitoring and Evaluation Policy, and after consultations with the IAs, the Evaluation Office will issue guidelines on conducting terminal evaluations. Beginning with the next APR, the Evaluation Office will also track the extent to which the terminal evaluation reports submitted by the IAs are in compliance with minimum requirement 3.

Since FY 2005, the Evaluation Office has been assessing the extent to which the project performance ratings for completed projects provided by the evaluation offices of partner organizations are consistent with its own ratings.[2] Based on its assessment, the Evaluation Office has decided that the **outcome ratings** provided by the World Bank's IEG on implementation completion report reviews can be trusted as being in accord with its own ratings. Therefore, beginning next year, the Office will accept IEG outcome ratings. This provision is in accordance with the Evaluation Office's effort to collaborate with its counterparts within the GEF partner Agencies and to prevent duplication of effort. The Office appreciates the efforts of UNEP's evaluation unit in providing commentary and ratings on completed projects. Because only three reviews have been done so far, it is too early for the Evaluation Office to work with and assess this material.

There is a substantial difference between the **sustainability ratings** provided by the Evaluation Office and those from the evaluation offices of the partner organizations. This difference is primarily driven by changes made by the Evaluation Office in its methodology to assess sustainability of outcomes. The Office will collaborate with its counterparts within the partner organizations to facilitate convergence on this issue.

Although the outcome ratings provided by partner Agency evaluation offices are similar to those provided by the GEF Evaluation Office, those provided in the terminal evaluation reports differ significantly. Compared to ratings provided by the Evaluation Office, **terminal evaluation report ratings** tend to assess a project's outcome achievements more optimistically. Many of the terminal evaluation reports that did not provide sufficient information on a specific dimension to allow the Evaluation Office to assess performance had rated the respective project on performance in that dimension. Since evidence shows that the ratings provided in the terminal evaluation reports are comparatively optimistic, they were not accepted by the Evaluation Office. More unbiased ratings in terminal evaluation reports are required before the Office could consider them to be reliable.

7.1 Rating Approach

The assessment approach adopted for the terminal evaluation reports submitted in FY 2006 was the same as that adopted for FY 2005; those submitted in FY 2004 had been assessed using a slightly different but comparable methodology. The reports submitted in FY 2005 and 2006 were assessed using 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 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?

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

7.2 Findings

Of the 66 terminal evaluation reports submitted for review in FY 2006, the quality of 54 (82 percent) was rated moderately satisfactory or above (see table 7.1). This figure is comparable to the 88 percent so rated in the FY 2005 cohort but higher than the 69 percent for the FY 2004 cohort.

The terminal evaluation reports of 78 percent of medium-size projects and 88 percent of full-size projects were rated moderately satisfactory or above. This difference is not statistically significant in the context of the 2006 cohort. However, when all 149 terminal evaluation reports rated so far are considered on a six-point scale, there is a significant difference in the ratings for MSPs and FSPs. This difference largely reflects—as was confirmed during discussions with the IAs—lesser attention being given by the IAs to quality control for the MSP terminal evaluation reports.

Beginning this year, the Evaluation Office compared ratings for quality of terminal evaluation reports based on the year of report completion rather than the year of report submission, because the high variation in time lag between completion and submission of individual reports makes it difficult to observe underlying trends in changes in report quality.[3] However, using year of completion as a basis for comparison imposes its own limitation, since all the terminal evaluation reports for some cohorts, especially the most recent one, have not been submitted; thus, the ratings for as-yet-unsubmitted reports are not known. For example, although the 149 terminal evaluation reports reviewed so far include those for evaluations completed from 2001 to 2006, most were completed in FY 2004 and 2005.

Figure 7.2 presents the proportion of terminal evaluation reports with quality rated as moderately satisfactory or above on the basis of year of terminal evaluation report completion. No clear trend is visible. As the number of cohorts and the number of terminal evaluation reports that have been reviewed within each cohort increase, the Evaluation Office will be able to assess trends.

Given the small number of terminal evaluation reports reviewed each year, comparing perfor-

Table 7.1

Quality of Terminal Evaluation Reports by IA and Focal Area
Number of reports

	IA			Focal area						
Rating	UNDP	UNEP	WB	BD	CC	IW	LD	MF	OD	Total
Highly satisfactory	0	0	1	1	0	0	0	0	0	1
Satisfactory	4	2	19	15	5	3	0	1	1	25
Moderately satisfactory	9	4	15	16	7	4	0	1	0	28
Subtotal	**13**	**6**	**35**	**32**	**12**	**7**	**0**	**2**	**1**	**54**
Moderately unsatisfactory	0	3	4	4	0	2	1	0	0	7
Unsatisfactory	1	0	3	1	2	1	0	0	0	4
Highly unsatisfactory	1	0	0	1	0	0	0	0	0	1
Subtotal	**2**	**3**	**7**	**6**	**2**	**3**	**1**	**0**	**0**	**12**
Unable to assess	0	0	0	0	0	0	0	0	0	0
Total	**15**	**9**	**42**	**38**	**14**	**10**	**1**	**2**	**1**	**66**

Note: BD = biodiversity; CC = climate change; IW = international waters; LD = land degradation; MF = multifocal; OD = ozone depletion.

Figure 7.2

Proportion of Terminal Evaluation Reports with Quality Rated Moderately Satisfactory or Above

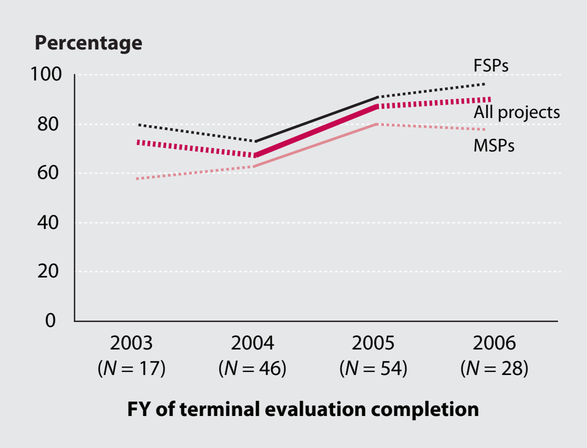
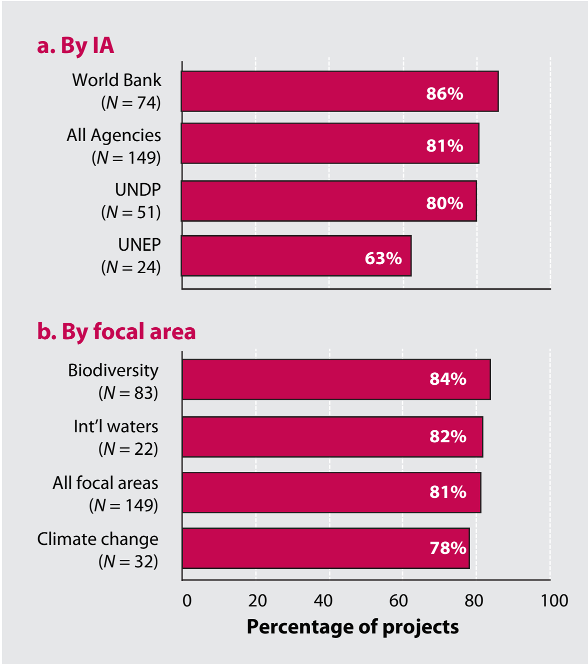


Figure 7.3

Proportion of Terminal Evaluation Reports with Quality Rated Moderately Satisfactory or Above, by IA and Focal Area



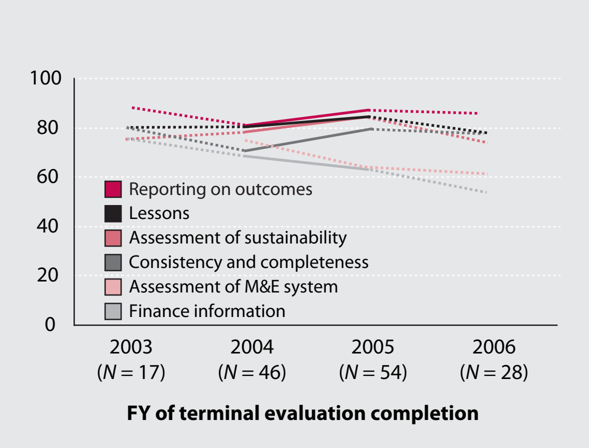
mance by IA and focal area in report quality ratings is difficult within each cohort. However, if terminal evaluation reports of all the cohorts are considered together, some comparisons are possible; these findings should in no way be considered definitive, given the relatively small sample size. When all 149 terminal evaluation reports that have been reviewed for quality are considered, a greater proportion of those by the World Bank (86 percent) was rated moderately satisfactory or above in quality than for the other Agencies (see figure 7.3). Also, compared to the other IAs, a smaller proportion of the terminal evaluation reports submitted by UNEP (63 percent) was rated moderately satisfactory or above in quality. These differences are significant even after accounting for differences in project mix in terms of focal area and scope (country versus regional/global projects). There is little difference by focal area in terms of the proportion of terminal evaluation reports rated moderately satisfactory or better in quality (see figure 7.3).

A detailed assessment of the ratings of terminal evaluation reports shows that many quality dimensions are addressed better than others. For example, 90 percent of the reports were rated moderately satisfactory or above in providing adequate information on relevant outcomes and achievement of objectives. Similarly, 84 percent of the reports were rated moderately satisfactory or above in providing an insightful description of the lessons that could be learned from the project. In comparison, only 59 percent were similarly rated for providing adequate financial information and only 60 percent for adequate information on assessment of M&E systems.

Figure 7.4 depicts the trends in proportion of projects rated moderately satisfactory or above for various quality dimensions based on year of terminal evaluation completion. The lines denote the period for which most reports of the cohorts have been reviewed. No clear trend is evident for each dimension. However, consistently more reports provide adequate information on outcomes and

Figure 7.4

Proportion of Terminal Evaluation Reports with Quality Rated Moderately Satisfactory or Above on Each Quality Dimension



on insightful description of lessons learned than for project financial aspects and assessment of M&E systems. Thus, the quality of terminal evaluation reports needs to be improved in terms of financial information and assessment of M&E systems. Draft guidelines on conducting terminal evaluations issued by the Evaluation Office to the IAs address these concerns. It is hoped that, in future, terminal evaluation reports will provide adequate information on these dimensions.

7.3 IA versus Evaluation Office Ratings

For many projects whose terminal evaluation reports have been reviewed by the Evaluation Office, ratings on outcomes, sustainability, and M&E systems are also provided by the terminal evaluation reports and independent evaluation offices of the respective Agency. Although the methodologies followed by the IA evaluation offices may not be identical to those followed by the Evaluation Office, the criteria on which performance is rated are often the same. The Evaluation Office compares its ratings with those provided by Agency evaluation offices and terminal evaluation reports using both an expanded scale (six- or four-point scale) and a binary scale to determine the extent to which ratings are consistent.

Among the IAs, the evaluation office of the World Bank has been providing ratings on outcomes, sustainability, M&E, and quality of terminal evaluation reports for some of the World Bank GEF projects. The UNEP independent evaluation office started providing such ratings this past year.

Of the 66 terminal evaluation reports submitted in FY 2006, outcome ratings have been provided by the evaluation offices of the respective Agencies for 20 projects—17 World Bank projects and 3 UNEP projects. Compared to the outcome ratings given by the Evaluation Office, those given by the Agency evaluation offices are, on average, 0.3 points higher on a six-point scale. Although this difference is statistically significant, the magnitude of difference is not substantial. Similarly, binary scale ratings accorded by the Evaluation Office on the one hand and the Agency evaluation offices on the other were identical for all but 1 of the 20 projects. In the past, the outcome ratings provided by the IA evaluation offices, especially those provided by the Independent Evaluation Group of the World Bank, have been consistent with GEF Evaluation Office ratings. In future, when provided, the Evaluation Office will accept the IEG outcome ratings. For UNEP, it is too early to assess the extent to which the ratings of its evaluation unit are consistent with those provided by the Evaluation Office.

For the FY 2006 cohort, there was a substantial difference between the sustainability ratings provided by the Evaluation Office and by the evaluation offices of partner organizations. This difference primarily reflects changes made by the Evaluation Office in its methodology for assessing sustainability of outcomes this year. The Evaluation Office will collaborate with its sister offices to facilitate convergence on this issue.

In the 66 terminal evaluation reports submitted this year, outcome ratings have been provided for 30 projects—19 from the World Bank, 5 from UNDP, and 6 from UNEP. Compared to the outcome ratings given by the Evaluation Office, those given in the terminal evaluation reports were, on average, higher by 0.7 points on a six-point scale. This difference is both significant and substantial in magnitude. The difference remains significant even when ratings are compared using a binary scale. Thus, compared to ratings provided by the Evaluation Office, the ratings provided in Agency terminal evaluation reports tend to assess a project's outcome achievements more optimistically. Because terminal evaluation reports are expected to present a fair and objective assessment of project performance, the ratings they provide need to be candid and reflective of and consistent with report contents; this is an area targeted for future improvement.

Notes

1. The terminal evaluation reports submitted in FY 2004 were not rated on the quality of their assessment of project M&E systems. However, analysis showed that inclusion/exclusion of criteria does not result in substantive differences in the overall ratings of the quality of terminal evaluation reports. The overall quality rating of one report would have changed from moderately satisfactory to moderately unsatisfactory; the rating for another report would have changed in the reverse direction.

2. The World Bank's Independent Evaluation Group has been providing ratings for quite some time; this year, UNEP's evaluation unit also provided ratings.

3. If year of report submission is used as a basis for assessment, changes in report quality anticipated as a result of Evaluation Office and Agency improvement measures will be less discernible, since reports completed in a year could be submitted over several years.

8. Management Action Records

The MARs keep track of the level of adoption of Council decisions on the basis of evaluation findings and recommendations. The two purposes of the MARs are (1) to provide the Council with a record of its decisions on the follow-up to evaluation reports, proposed management actions, and the actual status of these actions; and (2) to increase the accountability of GEF management regarding Council decisions on M&E issues. The GEF Council approved the format and procedures for the GEF MARs at its November 2005 meeting and requested that the GEF Evaluation Office prepare updated MARs to be presented to the Council for review and follow-up on an annual basis.

The rating categories for the progress of adoption of Council decisions were agreed upon in a consultative process between the Evaluation Office and the GEF Secretariat and 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 plans and actions for adoption are in a very preliminary stage
- **N/A—**not applicable

In this year's exercise, the Evaluation Office found it necessary to introduce a new rating:

- **Not possible to verify yet—**verification will have to wait until more data are available or proposals have been further developed

The first MAR was presented to the Council in June 2006, but the preparatory process was flawed. Consequently, it was impossible for the Evaluation Office to verify the ratings in time for the Council meeting. This year's MARs are the first to present ratings of GEF management and the verification of these ratings by the Evaluation Office. They track management actions on Council decisions based on eight Office reports:[1]

- "Annual Performance Report 2004" (GEF/ME/C.25/1, May 2005)
- "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)
- "GEF Country Portfolio Evaluation: Costa Rica" (GEF/ME/C.28/5, May 2006)
- "Evaluation of Incremental Cost Assessment" (GEF/ME/C.30/2, November 2006)
- "Evaluation of the Experience of Executing Agencies under Expanded Opportunities in the GEF" (GEF/ME/C.30/4, 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)

Five older evaluation reports have become overtaken by the recent changes in the GEF and can be considered no longer relevant. Their MARs have been archived.[2]

The current MARs track management actions on 36 Council decisions. The Evaluation Office rated 33 percent of these decisions as having been adopted by management at high or substantial levels. For 3 percent of the decisions, adoption was rated as negligible.

The Evaluation Office and management agreed on the rating of progress of adoption for 47 percent of decisions (17 of 36). On the other 53 percent, the Evaluation Office downgraded management's ratings. As shown in table 8.1, most disagreements between management and GEF Evaluation Office ratings are in the higher levels of adoption (high and substantial). Many of the lower ratings given by the Office reflect the fact that proposals to the Council have yet to be approved by it. If and when the Council does approve these proposals, substantial adoption may have occurred. For next year's MARs, this issue may be discussed further and may lead to additional guidelines.

There are several Council decisions on which progress in adoption by management is significant. These include the adoption of the terminal evaluation review processes by the UNDP and UNEP evaluation units, and the GEF Secretariat proposal to provide a "level playing field" for Implementing and Executing Agencies.

The only decision whose adoption was assessed as negligible by both GEF management and the Evaluation Office relates to the Council's June 2005 decision requesting increased transparency in the GEF project approval process through an improved management information system. Work on this system is still in its early stages and has not yet led to any visible improvement in the information that is available on where proposals are in the approval process. The Evaluation Office reiterates its viewpoint that making information available in a transparent way is not difficult, but does require discipline in gathering information as well as diligent staff effort to present this information on the Web site in an accessible format. Software does not need to be written nor a full system put into place in order to manage information.

Table 8.1

Ratings of GEF Progress toward Adopting Council Decisions

	GEF Evaluation Office ratings					
Management ratings	**High**	**Substantial**	**Medium**	**Negligible**	**Not possible to verify yet**	**Sum of management ratings**
High	5	2	6			**13**
Substantial		5	8		2	**15**
Medium			6		1	**7**
Negligible				1		**1**
Not possible to verify yet						
Sum of GEF EO ratings	**5**	**7**	**20**	**1**	**3**	**36**

Note: Highlighted fields show agreement between the ratings of management and the GEF Evaluation Office; fields to the right of the diagonal represent higher ratings by management than by the Evaluation Office (except in the case of not possible to verify yet).

Another issue that the Evaluation Office believes needs to be better adopted is the Council decision to integrate local benefits in a more systematic way in all stages of the GEF Activity Cycle. Management rates the adoption of this decision as high, but the Evaluation Office assesses the adoption rate for this decision as medium. The Office assessment is based on this year's APR findings, which indicate that the integration of social issues into supervision of GEF projects, when appropriate, has been insufficient in UNDP and UNEP. Furthermore, the GEF Secretariat still lacks expertise in this area—there has been no social scientist post in the Secretariat since 2003.

The Evaluation Office also notes that the quality of terminal evaluations is still in need of improvement. Recommendations from both the 2004 and 2005 APRs call for improved quality of terminal evaluations, but even though GEF Agencies have developed and tested terminal evaluation review processes, they still need to ensure that terminal evaluation reports include adequate information on sustainability of outcomes, quality of M&E systems, and reporting on cofinancing.

The many changes in the MARs from last year's version to this year's are an indication of a dynamic GEF, going through a process of change. For five older evaluation reports, the Council decisions have been overtaken by new developments, thus making tracking of the adoption of these old Council decisions no longer relevant. This does not mean that the lessons from past experience of these older reports have been overlooked or are now forgotten—rather, they have been integrated into more recent efforts by the new GEF Chief Executive Officer (CEO) to renew the strategies of the GEF and fulfill the GEF-4 replenishment agreement. Furthermore, the difference in ratings of level of adoption between the Evaluation Office and GEF management generally does not reflect a disagreement about the direction of the adoption. In most cases, the Office believes that management is on track—just not as far along as management considers itself to be. This is typically because management is expecting the Council to agree with its proposals, whereas the Evaluation Office tends to await Council decisions on these proposals.

The lack of progress in providing transparency on management information in the GEF is a cause for concern. A Council decision in 2005 and a Council reminder in 2006 have not yet been adequately met by the Secretariat, which is fully aware of the situation and has rated its own performance in this regard as negligible. The Evaluation Office again calls attention to the fact that making management information available in a transparent manner is not a complex issue, but instead requires sufficient human resources, energy, and dedication.

Notes

1. Documents are listed by their Council designation so as to clearly indicate the date when the respective Council decisions were made.

2. This concerns the following reports:

   - "GEF's Engagement with the Private Sector" (GEF/ME/C.23/Inf.4, May 2004)
   - "Program Study on Biodiversity" (GEF/ME/C/C.24/Inf.1, November 2004)
   - "Program Study on Climate Change" (GEF/ME/C.24/Inf.2, November 2004)
   - "Program Study on International Waters" (GEF/ME/C.24/Inf.3, November 2004)
   - "Review of the GEF Operational Program 12: Integrated Ecosystem Management" (GEF/ME/C.25/5, May 2005)

   The MAR on GEF's Engagement with the Private Sector was archived because the new approach to the private sector that is being proposed by the

GEF Secretariat makes previous recommendations on this issue obsolete. The other reports have been overtaken by the new focal area strategies the GEF Secretariat is presenting to the Council. Recommendations from the program studies were also incorporated into the GEF replenishment agreement and the recommendations of the GEF Third Overall Performance Study.

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 effectiveness.[1] The ratings on the outcomes of the project will be based on performance on the following criteria:[2]

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[3])?

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) as follows:

- **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.** Unable to assess outcomes on this dimension.

The calculation of the overall outcomes score of projects will consider all three criteria, of which relevance and effectiveness are critical. The overall score will be determined satisfying two different conditions: (1) the overall score on project outcomes will have a constraint that it cannot be higher than the lower score among the critical criteria of relevance and effectiveness; and (2) the overall score cannot be higher than the average score of all three criteria calculated using the following formula:

$$\text{Outcomes} = (a + b + c) \div 3$$

In case the average score calculated using this formula is lower than or equal to the lower score among the critical criteria of relevance and effectiveness, then the average score will be the overall score. In case the average score is higher than the lower score among the critical criteria of relevance and effectiveness, then the lower score among the critical criteria will be the overall rating. If the rating for the critical dimensions is unable to assess, then the overall rating for outcomes will also be unable to assess.

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 framework and governance, and environmental.

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

- **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?

- **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 it as 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?
- **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.
- **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 a 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 Catalytic Role

The terminal evaluation guidelines have provided no specific guidance regarding assessment of the catalytic role and replication in GEF projects.[4] The upcoming evaluation on catalytic role will provide a framework to carry out this assessment in future. Meanwhile, terminal evaluation reviews will seek to identify four different levels of catalytic results. These results need not be linear or cumulative; thus, it is possible that one is present while others might not be present. The levels of catalytic results recorded will be as follows:

a. **Production of a public good.** This would be the lowest level of catalytic result. It might include the production of knowledge, a new technology, or a new approach. At the lowest level of catalysis, the catalytic effect as such is left to the market. No actions are undertaken by the project to propagate or promote the public good that has been created.

b. **Demonstration.** Under this category, the project that produces a public good takes steps to promote propagation. This could include setting up a demonstration site, working to get the message out, and providing assistance to others interested in repeating the experience.

c. **Replication.** Experience is repeated either inside or outside the project, in other GEF projects, or by other agencies or actors within the country or internationally.

d. **Scaling up.** The project activities lead to changes at the system level. New or proposed approaches become widely accepted, or they become the law of the land. This normally would involve some kind of policy decision.

Problems might occur when the term "mainstreaming" is used when referring to scaling up and replication regarding GEF operations. Further terminology in this regard will be developed by the upcoming evaluation of the catalytic role.

Catalytic results will not be rated by terminal evaluation reviews.

A.5 Criteria for Assessment of Quality of Project M&E Systems

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

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

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

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

  - Was sufficient funding provided for M&E in the budget included in the project document?
  - Was sufficient and timely funding provided for M&E during project implementation?
  - 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 M&E System = a

A.6 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 in each criterion 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 report will be rated as follows:

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

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

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

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

A.7 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?

Notes

1. *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).
2. *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.
3. The GEF Secretariat and IAs are currently seeking to better align the focal area program indicators and tracking tools with focal area strategic priorities and project objectives. This will enable the aggregation of outcomes and impacts for each focal area to annually measure progress toward targets in the program indicators and strategic priorities.
4. *Replication* refers to the repeatability of the project under similar contexts based on lessons learned and experience gained. Actions to foster replication include dissemination of results, seminars, training workshops, and field visits to project sites (GEF 2000).

Annex B. Terminal Evaluation Tables

Table B.1

Outcomes and Sustainability Ratings in FY 2006
Number of projects

Sustainability rating	Outcome ratings: Highly satisfactory	Moderately satisfactory	Satisfactory	Moderately unsatis-factory	Unsatisfac-tory	Highly unsatisfac-tory	Unable to assess	Total
Likely	3	15	5	0	0	0	0	**23**
Moderately likely	0	5	5	2	0	0	0	**12**
Moderately unlikely	1	2	4	3	0	0	0	**10**
Unlikely	0	0	5	3	1	0	0	**9**
Unable to assess	0	6	3	1	0	0	2	**12**
Total	**4**	**28**	**22**	**9**	**1**	**0**	**2**	**66**

Table B.2

Outcomes and Sustainability Ratings in FY 2006
GEF investment in million $

Sustainability rating	Outcome ratings: Highly satisfactory	Moderately satisfactory	Satisfactory	Moderately unsatis-factory	Unsatisfac-tory	Highly unsatisfac-tory	Unable to assess	Total
Likely	13.8	35.2	40.9	0.0	0.0	0.0	0.0	**89.8**
Moderately likely	0.0	18.7	14.7	7.3	0.0	0.0	0.0	**40.7**
Moderately unlikely	1.0	1.5	22.4	10.2	0.0	0.0	0.0	**35.1**
Unlikely	0.0	0.0	41.1	8.5	3.1	0.0	0.0	**52.7**
Unable to assess	0.0	19.9	15.1	1.0	0.0	0.0	1.5	**37.5**
Total	**14.7**	**75.3**	**134.2**	**27.0**	**3.1**	**0.0**	**1.5**	**255.8**

Table B.3

Materialization of Cofinancing by Focal Area

Factor	Biodiversity	Climate change	International waters
Number of terminal evaluation reports reviewed	99	43	26
Number of reports providing actual cost information	70	27	13
Promised $ of cofinance per GEF $ approved	1.7	9.1	1.8
Promised cofinancing that materialized (%)	96	93	144

Table B.4

Terminal Evaluations Reviewed by the Evaluation Office in FY 2006

GEF ID	Name	Project size	Region	IA	Focal area	GEF amount approved (million $)
5	Regional – Capacity Building for the Adoption and Application of Energy Codes for Buildings	MSP	Asia	UNDP	CC	0.99
17	South Africa – Conservation of Globally Significant Biodiversity in Agricultural Landscapes through Conservation Farming	MSP	Africa	WB	BD	0.75
18	Kenya – Lewa Wildlife Conservancy	MSP	Africa	WB	BD	0.73
20	South Africa – Conservation Planning in Thicket Biome	MSP	Africa	WB	BD	0.74
21	Micronesia – Community Conservation and Compatible Enterprise development in Pohnpei	MSP	Asia	UNDP	BD	–
26	Indonesia – Aceh Elephant Landscape Project	MSP	Asia	WB	BD	0.72
55	Regional – West Africa Pilot Community-Based Natural Resource Management Project	FSP	Africa	WB	BD	7.00
58	Brazil – National Biodiversity Project	FSP	LAC	WB	BD	10.00
60	Costa Rica – Tejona Wind Power	FSP	LAC	WB/ IDB	CC	3.30
88	Regional – Lake Victoria Environmental Management Project	FSP	Africa	WB	IW	35.00
101	Uganda – Institutional Capacity Building for Protected Areas Management and Sustainable Use	FSP	Africa	WB	BD	2.00
103	Costa Rica – Biodiversity Resources Development	FSP	LAC	WB	BD	7.00
107	Ukraine – ODS Phase-out	FSP	ECA	WB	OD	23.34
113	Regional – Lake Ohrid Management	FSP	ECA	WB	IW	4.10
117	Nicaragua – Atlantic Biological Corridor	FSP	LAC	WB	BD	7.10
118	Senegal – Sustainable and Participatory Energy Management	FSP	Africa	WB	CC	4.70
121	Honduras – Biodiversity in Priority Areas Project	FSP	LAC	WB	BD	7.00
123	Latvia – Solid Waste Management and Landfill Gas Recovery	FSP	ECA	WB	CC	5.12
133	Panama – PAMBC Atlantic Mesoamerican Biological Corridor Project	FSP	LAC	WB	BD	8.40
134	South Africa – Cape Peninsula Biodiversity Conservation Project	FSP	Africa	WB	BD	12.30
202	Regional (Bolivia, Peru) – Conservation of Biodiversity in the Lake Titicaca Basin	FSP	LAC	UNDP	BD	3.11
218	Central African Republic – A Highly Decentralized Approach to Biodiversity Protection and Use: The Bangassou Dense Forest	FSP	Africa	UNDP	BD	2.50

GEF ID	Name	Project size	Region	IA	Focal area	GEF amount approved (million $)
250	Mongolia – Biodiversity Conservation and Sustainable Livelihood Options in the Grasslands of Eastern Mongolia	FSP	Asia	UNDP	BD	5.11
370	India – Development of High-Rate Biomethanation Processes as Means of Reducing Greenhouse Gas Emissions	FSP	Asia	UNDP	CC	5.50
400	Regional – Conservation and Sustainable Use of Dryland Agro-Biodiversity of the Fertile Crescent	FSP	Asia	UNDP	BD	8.12
408	Benin – National Parks Conservation and Management Project	FSP	Africa	WB	BD	6.76
424	Laos – Off-grid Electrification Pilot Demonstration, A Component of the Laos Southern Provinces Rural Electrification	MSP	Asia	WB	CC	0.74
460	Regional – Preparation of a Strategic Action Programme for the Dnieper River Basin and Development of SAP Implementation Mechanisms	FSP	ECA	UNDP	IW	7.00
461	Regional – Determination of the Priority Actions for the Further Elaboration and Implementation of the Strategic Action Programme for the Mediterranean Sea	FSP	ECA/ Africa	UNEP	IW	6.10
466	El Salvador – Promotion of Biodiversity Conservation within Coffee Landscapes	MSP	LAC	WB	BD	0.75
540	Thailand – Building Chiller Replacement Program	FSP	Asia	WB	CC	2.50
570	Cote d'Ivoire – Energy Efficiency Market Development	MSP	Africa	WB	CC	0.73
583	Brazil – Implementation of Integrated Watershed Management Practices for the Pantanal and Upper Paraguay River Basin	FSP	LAC	UNEP	IW	6.62
584	Global International Waters Assessment	FSP	Global	UNEP	IW	6.79
586	Brazil – Integrated Management of Land Based Activities in the São Francisco Basin	FSP	LAC	UNEP	IW	4.77
625	Colombia – Sustainable Use of Biodiversity in the Western Slope of the Serrania del Baudo	MSP	LAC	WB	BD	0.73
637	Macedonia – Development of Mini-Hydropower Plants	FSP	ECA	WB	CC	1.50
654	Poland – Zakopane/Podhale Geothermal District Heating and Environment Project	FSP	ECA	WB	CC	5.40
656	Samoa – Marine Biodiversity Protection and Management	MSP	Asia	WB	BD	0.90
659	South Africa – Sustainable Protected Area Development in Namaqualand	MSP	Africa	WB	BD	0.75
681	Panama – Effective Protection with Community Participation of the New Protected Area of San Lorenzo	MSP	LAC	WB	BD	0.73
775	Ecuador – Choco-Andean Corridor	MSP	LAC	WB	BD	0.98
791	Regional – Formulation of a Strategic Action Programme for the Integrated Management of the San Juan River Basin and Its Coastal Zone	FSP	LAC	UNEP	IW	3.92
794	Regional – Catalyzing Conservation Action in Latin America: Identifying Priority Sites and Best Management Alternatives in Five Globally Significant Ecoregions	MSP	LAC	UNEP	BD	0.75
798	Philippines – Sustainable Management of Mt. Isarog's Territories	MSP	Asia	UNDP	BD	0.75
799	Philippines – Conservation of the Tubbataha Reef National Marine Park and World Heritage Site	MSP	Asia	UNDP	BD	0.75
800	Seychelles – Marine Ecosystem Management	MSP	Africa	WB	BD	0.75

GEF ID	Name	Project size	Region	IA	Focal area	GEF amount approved (million $)
807	Russia – Persistent Toxic Substances, Food Security and Indigenous Peoples of the Russian North	MSP	ECA	UNEP	IW	0.73
844	Chile – Valdivian Forest Zone: Private-Public Mechanisms for Biodiversity Conservation	MSP	LAC	WB	BD	0.73
845	Indonesia- The Greater Berbak-Sembilang Integrated Coastal Wetlands Conservation Project	MSP	Asia	WB	BD	0.73
846	Ecuador – Albarradas in Coastal Ecuador: Rescuing Ancient Knowledge on Sustainable Use of Biodiversity	MSP	LAC	WB	BD	0.73
880	China – Targeted Research Related to Climate Change	FSP	Asia	UNDP	CC	1.72
932	Russia – Sustainable Conservation of BD in Four Protected Areas of Russia's Kamchatka	FSP	ECA	UNDP	BD	2.10
940	Malaysia – Biomass-based Power Generation and Cogeneration in Palm Oil Industry – Phase 1	FSP	Asia	UNDP	CC	4.03
941	China – Demonstration for Fuel-Cell Bus Commercialization	FSP	Asia	UNDP	CC	5.82
979	Costa Rica – Biodiversity Conservation in Cacao Agro-forestry	MSP	LAC	WB	BD	0.73
1020	Colombia – Conservation and Sustainable Development of the Mataven Forest	MSP	LAC	WB	BD	0.73
1224	Global (Brazil, Cote d'Ivoire, India, Indonesia, Kenya, Mexico, Uganda) – Conservation and Sustainable Management of Below Ground Biodiversity, Phase I	FSP	Global	UNEP	BD	5.02
1280	China – Passive Solar Rural Health Clinics	MSP	Asia	WB	CC	0.78
1303	Russia – Strengthening Protected Areas Network for Sikhote-Alin Mountain Forest Ecosystems Conservation in Khabarovsky Kray	MSP	ECA	WB	BD	0.75
1378	Global – Assessment of Soil Organic Carbon Stocks and Change at National Scale	MSP	Global	UNEP	MF	0.98
1394	Regional (Burkina Faso, Cameroon, Egypt, Ethiopia, Ghana, Kenya, Niger, Nigeria, Senegal, South Africa, Zambia, Zimbabwe) – Climate, Water and Agriculture: Impacts on and Adaptation of Agro-Ecological Systems in Africa	MSP	Africa	WB	MF	0.70
1397	Mexico – Private Land Mechanisms for Biodiversity Conservation	MSP	LAC	WB	BD	0.73
1424	Indonesia – Forests and Media Project (INFORM)	MSP	Asia	WB	BD	0.94
1444	Regional (Estonia and Russia) – Development and Implementation of the Lake Peipsi/Chudskoe Basin Management Plan	MSP	ECA	UNDP	IW	1.00
2469	Regional – Supporting Capacity Building for the Elaboration of National Reports and Country Profiles by African Parties to the UNCCD	MSP	Africa	WB	LD	0.90

Note: BD = biodiversity; CC = climate change; ECA = Europe and Central Asia; IW = international waters; LAC = Latin America and the Caribbean; LD = land degradation.

Annex C. Projects in Sample

Agency	Country/ region	Project	GEF ID no.	Approval date	Type	Grant (mil. $).	Focal area
UNDP	Pacific Islands	Pacific Islands Renewable Energy Programme	1058	Oct. 02	MSP	0.7	CC
UNDP	Slovak Republic	Conservation, Restoration and Wise Use of Calcareous Fens	1681	Jun. 04	MSP	0.977	BD
UNDP	Ecuador	Galapagos Oil Spill : Environmental Rehabilitation and Conservation	1409	Jun. 01	MSP	0.53	BD
UNDP	Brazil	Demonstrations of Integrated Ecosystem and Watershed Management in the Caatinga, Phase I	609	Feb. 04	FSP	4.1	MF
UNDP	Chile	Conservation and Sustainable Use of Chiloé Globally Significant Biodiversity	1021	May 01	MSP	1	BD
UNDP	Kazakhstan	Integrated Conservation of Priority Globally Significant Migratory Bird Wetland Habitat	838	Aug. 03	FSP	8.71	BD
UNDP	Slovenia	Removing Barriers to the Increased Use of Biomass as an Energy Source	658	Dec. 01	FSP	4.3	CC
UNDP	Regional	Southern Africa Biodiversity Support Programme	260	Aug. 00	FSP	4.5	BD
UNDP	Bhutan	Linking and Enhancing Protected Areas in the Temperate Broadleaf Forest Ecoregion of Bhutan	1852	Apr. 03	MSP	0.792	BD
UNDP	Uzbekistan	Establishment of the Nuratau-Kyzylkum Biosphere Reserve as a Model for Biodiversity Conservation	855	Jan. 01	MSP	0.725	BD
UNDP	Cape Verde	Integrated Participatory Ecosystem Management In and Around Protected Areas, Phase I	1124	Oct. 03	FSP	3.585	BD
UNDP	South Africa	Pilot Production and Commercial Dissemination of Solar Cookers	1311	June 02	MSP	0.8	CC
UNDP	Suriname	Conservation of Globally Significant Forest Ecosystems in Suriname's Guayana Shield	661	Aug. 00	FSP	9.24	BD
UNDP	Mongolia	Commercialization of Super Insulated Building Technology	22	Oct. 01	MSP	0.725	CC
UNDP	Algeria	Conservation and Sustainable Use of Globally Significant Biodiversity in the Tassili and Ahaggar National Parks	1114	Jan. 04	FSP	3.54	BD
UNDP	Syria	Supply-Side Efficiency and Energy Conservation and Planning	264	Nov. 98	FSP	4.07	CC
UNDP	Nepal	Landscape-scale Conservation of Endangered Tiger and Rhinoceros Populations in and Around Chitwan National Park	906	Apr. 01	MSP	0.75	BD

Agency	Country/ region	Project	GEF ID no.	Approval date	Type	Grant (mil. $).	Focal area
UNDP	Mexico	Action Plan for Removing Barriers to the Full-scale Implementation of Wind Power	1284	Jun. 03	FSP	4.736	CC
UNDP	*China*	*Targeted Research Related to Climate Change*	*880*	*June 02*	*FSP*	*1.724*	*CC*
UNDP	*India*	*Enabling Activity for the Preparation of India's Initial Communication to the UNFCCC*	*11*	*Sept.00*	*FSP*	*2.0*	*CC*
UNDP	*Global*	*Removal of Barriers to the Introduction of Cleaner Artisanal Gold Mining and Extraction Technologies*	*1223*	*May 02*	*FSP*	*6.8*	*IW*
UNDP	*Philippines*	*Conservation of the Tubbataha Reefs National Marine Park and World Heritage Site*	*799*	*Aug. 00*	*MSP*	*0.725*	*BD*
UNDP	*Latvia*	*Economic and Cost-effective Use of Wood Waste for Municipal Heating Systems*	*914*	*Mar. 01*	*MSP*	*0.75*	*CC*
UNDP	*Tunisia*	*Barrier Removal to Encourage and Secure Market Transformation and Labeling of Refrigerators*	*576*	*Apr. 99*	*MSP*	*0.71*	*CC*
UNEP	Regional	Global Environmental Citizenship	464	Mar. 98	FSP	3.212	MF
UNEP	Regional	Sustainable Conservation of Globally Important Caribbean Bird Habitats: Strengthening a Regional Network for a Shared Resource	1604	Sept. 03	MSP	0.974	BD
UNEP	Kenya	Support to the Implementation of the National Biosafety Framework	1371	Sept. 01	MSP	0.51	BD
UNEP	Namibia	Support to the Implementation of the National Biosafety Framework	1372	Jul. 02	MSP	0.672	BD
UNEP	Regional	Biodiversity Conservation and Integration of Traditional Knowledge on Medicinal Plants in National Primary Health Care Policy in Central America and Caribbean	1410	Oct. 01	MSP	0.725	BD
UNEP	Global	Development of a Strategic Market Intervention Approach for Grid-Connected Solar Energy Technologies	1599	Apr. 04	MSP	0.975	CC
UNEP	Regional	Development and Implementation of Mechanisms to Disseminate Lessons Learned and Best Practices in Integrated Transboundary Water Resources Management in Latin America and the Caribbean	1426	Mar. 03	MSP	0.972	IW
UNEP	Regional	Development of a Wetland Site and Flyway Network for Conservation of the Siberian Crane and Other Migratory Waterbirds in Asia	1097	Mar. 03	FSP	10.35	BD
UNEP	Regional	Conservation of Gramineae and Associated Arthropods for Sustainable Agricultural Development in Africa	1344	Sept. 01	MSP	0.947	BD
UNEP	Global	Solar and Wind Energy Resource Assessment	1281	Jun. 01	FSP	6.512	CC
UNEP	Regional	Determination of Priority Actions for the Further Elaboration and Implementation of the Strategic Action Programme for the Mediterranean Sea	461	Dec. 00	FSP	4.61	IW
UNEP	Regional	An Integrated Ecosystem Management Approach to Conserve Biodiversity and Minimize Habitat Fragmentation in Three Selected Model Areas in the Russian Arctic	1163	May 04	FSP	3.375	MF

Agency	Country/ region	Project	GEF ID no.	Approval date	Type	Grant (mil. $).	Focal area
UNEP	Global	Promoting Ecosystem-based Approaches to Fisheries Conservation and LMEs	2474	May 04	MSP	0.995	IW
UNEP	Regional	Managing Hydrogeological Risk in the Iullemeden Aquifer System	2041	Dec. 03	MSP	0.958	IW
UNEP	Kenya	Joint Geophysical Imaging Methodology for Geothermal Reservoir Assessment	1780	Dec. 02	FSP	0.979	CC
UNEP	Global	In-situ Conservation of Crop Wild Relatives through Enhanced Information Management and Field Application	410	Mar. 04	FSP	5.827	BD
WB	Serbia	Reduction of Enterprise Nutrient Discharges Project	2141	May 05	FSP	9.02	IW
WB	Costa Rica	Ecomarkets	671	Jun. 00	FSP	8	BD
WB	Pakistan	Protected Areas Management	87	Apr. 01	FSP	10.08	BD
WB	Uruguay	Energy Efficiency	1179	May 04	FSP	6.875	CC
WB	Regional	Integrated Ecosystem Management in Indigenous Communities	1092	Dec. 04	FSP	4	BD
WB	Peru	Participatory Management of Protected Areas	1101	Mar. 03	FSP	14.8	BD
WB	Tajikistan	Community Agriculture and Watershed Management	1872	Jun. 04	FSP	10.8	BD
WB	Bosnia-Herzegovina	Water Quality Protection	2143	Jun. 05	FSP	8.9	IW
WB	Tunisia	Gulf of Gabes Marine and Coastal Resources Protection	1174	Mar. 05	FSP	6.31	BD
WB	Vietnam	Demand-Side Management and Energy Efficiency Program	1083	Jun. 03	FSP	5.5	CC
WB	Burundi	Agricultural Rehabilitation and Sustainable Land Management	2357	Jul. 04	FSP	5	LD
WB	Guinea-Bissau	Coastal and Biodiversity Management	1221	Nov. 04	FSP	3	BD
WB	Mozambique	Coastal and Marine Biodiversity Management	648	Jun. 00	FSP	5.6	BD
WB	Bulgaria	Energy Efficiency	2117	Mar. 05	FSP	10.0	CC
WB	Morocco	Protected Areas Management	409	Jan. 00	FSP	10.5	BD

Note: Projects in italics were closed and not reviewed for this assessment. BD = biodiversity; CC = climate change; IW = international waters; LD = land degradation; MF = multifocal.

Annex D. Supervision Questionnaire

Project Supervision Assessment

Agency:

Country: **Project Name:**

Id No.: **Focal Area/OP:**

GEF Amount (million $):

Date Approved:

Person(s) Interviewed:

Panel Member(s):

Assessment Date:

Overall Rating: ___

Comment:

Summary Ratings Sheet

Assessment Rating

1 = Highly Satisfactory
2 = Satisfactory
3 = Moderately Satisfactory
4 = Moderately Unsatisfactory
5 = Unsatisfactory
6 = Highly Unsatisfactory
NA = Not Applicable

1. Focus on Results
2. Supervision of Fiduciary/Due Diligence Aspects
3. Adequacy of Supervision Inputs and Processes
4. Candor and Quality of Project Performance Ratings

Overall Assessment

Comment: The overall assessment is not an average of the assessments of the constituent elements of supervision and do not include rating applied to the quality at entry. The panel used its judgment in weighing the relative importance of each supervision element given the country and the project context. In making the assessment, the panel considered the importance of each category, and within each category, the various questions, to supervision quality.

Part A. Context

(This section assesses project quality at entry. These ratings were not aggregated with those in part B.)

a. Assessment of project design at entry

b. Assessment of quality of logframe or results framework (including intermediate performance indicators) as well as readiness of implementation arrangements for M&E

c. Government commitment to the project

d. Readiness of implementation at approval

e. Overall performance prior to period under review

f. Extent to which the problems encountered during implementation have been identified at entry

Part B. Supervision Ratings

1. Focus on Results

1.1 Identification and Assessment of Problems

- a. Timely identification and assessment of implementation problems?
- b. Timely identification and assessment of possible threats to global environmental objective?

1.2 Focus on Sustainability (extent to which supervision paid attention to this aspect)

- a. Borrower and stakeholder ownership
- b. Technical assistance, training, and capacity building

1.3 Actions Taken and Follow-Up

- a. Appropriateness of advice and proposed solutions
- b. Appropriateness and speed of IA follow-up action
- c. Impact and effectiveness of IA actions
- d. Quality and timeliness of midterm review (if any)
- e. Quality and timeliness of restructuring plan (if any)

1.4 Performance Monitoring

- a. Extent to which the task team made use of the global environmental objective and intermediate outcome indicators to assess the project's implementation, as reported in the PIR
- b. Extent to which the indicators (both quantitative and qualitative) have been used to identify and address potential obstacles to the achievement of the global environmental objective (attention to long-term objective)

2. Supervision of Fiduciary/Due Diligence Aspects

2.1 Procurement

2.2 Financial Management

2.3 Legal Aspects e.g., Legal compliance; clarity/timeliness of advice

2.4 Environmental Aspects e.g., Environmental impact

2.5 Social Aspects e.g., Social impact

3. Adequacy of Supervision Inputs and Processes

3.1 Staffing

a. Staff continuity

b. Supervision skill mix

3.2 Supervision Activities

a. Quality of mission preparation and effectiveness of time spent in the field

b. If there was joint supervision with cofinancers, quality of coordination and other supervision activities

3.3 Quality of Supervision Documentation

3.4 Effectiveness of Relationships

a. With the government

b. With donors and other stakeholders

3.5 Management Inputs

a. Adequacy and speed of management attention and actions

b. Adequacy of supervision budget

c. Effectiveness of budget use

4. Candor and Quality of Project Performance Reporting

4.1 Extent to Which Ratings Reflect Actual Conditions

4.2 Adequate Explanation of Ratings, and of Any Change in Ratings

4.3 Accuracy of Ratings of Project Components and Risk Assessment

4.4 Quality and Timeliness of Data (including intermediate outcome indicators) to Support the Key Performance Indicators

Annex E. Agency Performance and Outcomes Matrix

Parameter	Agency	Current status of development of indicators and tools used to measure performance	Information source and frequency of reporting
Results			
1. Project outcomes	IAs, ExAs	Performance is measured on the basis of desk reviews that verify ratings of project terminal evaluations. Since the 2005 APR, the Evaluation Office has been tracking differences between its ratings and those given by Agencies' independent evaluation offices. As differences diminish, the Office will verify ratings on a sample of projects and will accept and report to the GEF Council on the basis of Agencies' independent evaluation office ratings.	Reporting on desk-verified ratings will continue on an annual basis. Every four years, as part of the focal area program studies, and on an opportunistic basis (for example, as part of country portfolio reviews), projects will be selected for field verification. The information provided by the APR will also be incorporated in a broader analysis of results that will be carried out by the focal area program studies.
2. Risks to the sustainability of outcomes	IAs, ExAs	Performance will be measured on the basis of desk reviews that verify the risk to project outcome ratings from project terminal evaluations. Since the 2005 APR, the Evaluation Office has been tracking differences between its ratings and those given by Agencies' independent evaluation offices. As differences diminish, the Office will verify ratings on a sample of projects and will accept and report to the GEF Council on the basis of Agencies' independent evaluation office ratings.	
Processes affecting results			
3. Quality of supervision and adaptive management	IAs, ExAs	The Evaluation Office is conducting the first quality of supervision review for GEF IAs as part of the 2006 APR. For World Bank projects, this review has been carried out in collaboration with the Bank's Quality of Supervision Assessment 7. Through this initial review, the Office will develop criteria and tools to assess project supervision in other GEF Agencies, considering various aspects of supervision to include consideration of GEF criteria, recommendations from previous evaluations, and other relevant IA and ExA contractual commitments (use of GEF fees, for example). The quality of supervision review will address adaptive management of GEF Agencies by examining the extent and forms by which Agencies identify, track, and respond to risk and problems affecting project implementation.	Reporting will be on a biennial basis. The Office will continue to collaborate with World Bank quality of supervision assessments. Relevant findings from other evaluations carried out by the Office such as the Local Benefits Study, country program evaluations, and focal area program studies will be used in this reporting as well.

Parameter	Agency	Current status of development of indicators and tools used to measure performance	Information source and frequency of reporting
Efficiency			
4. Project preparation elapsed time	GEF Secretariat, IAs, ExAs	The indicator is the average number of months required to prepare projects. These data are generated by a combination of information provided by Agencies and the GEF Secretariat database. Agencies will be accountable for the elapsed time of preparation tasks for which they are responsible. The Evaluation Office will explore ways to define standards for elapsed time and to improve the reliability of data for future reporting once the Council has approved proposed changes to the Activity Cycle.	This parameter is normally reported annually in the project performance report, now the responsibility of the GEF Secretariat. The APR will verify this information on an annual basis. The Joint Evaluation of the Activity Cycle reported on this issue as well.
5. Implementation completion delays	IAs, ExAs	The indicator is the average number of months of delays in scheduled and actual project closing. This information has been gathered since 2005 by the Evaluation Office as part of the Joint Evaluation of the Activity Cycle and the APR process.	Reporting will continue on an annual basis.
6. Materialization of cofinancing		Average rate of promised and realized cofinancing. This information has been gathered since 2005 by the Evaluation Office as part of the APR process from project terminal evaluations submitted to the Office by IAs and ExAs.	Reporting will continue on an annual basis. Also, as part of the focal area program studies, the Evaluation Office will assess the reliability of cofinancing figures reported in terminal evaluations by IAs and ExAs.
Quality of M&E			
7. Independence of Agency central evaluation units	IAs, ExAs, Evaluation Office	The Office will draw on the ratings, self-reporting, and peer reviews carried out in the context of the multilateral development banks' Evaluation Cooperation Group and the United Nations Evaluation Group. The charter and mandate of the various evaluation units will also provide evidence of their degree of independence.	Reporting will be on an annual basis, although the data may not change for all Agencies each year. The year of validity of the rating will be incorporated in the reporting.
8. Independence of terminal evaluations or independent review of terminal evaluations	IAs	The Office will assess the independence of the terminal evaluation process, including formulation of terms of reference, selection of evaluators, and review of evaluations.	Reporting will be on an annual basis and will include actions that IAs and ExAs have undertaken to further strengthen the independence of their evaluation processes.
9. Realism of risk assessment (robustness of project-at-risk systems).	IAs, ExAs, GEF Secretariat	The 2005 APR developed an inventory of IA and ExA project-at-risk systems that includes 13 performance parameters. On the basis of these parameters, the Office will develop a tool to assess and rate Agency systems to identify and manage risk for the 2007 APR. The Office will also develop a tool to assess the treatment of risk by the GEF Secretariat during the project approval process.	The Office will report on project-at-risk systems starting in 2007. The frequency of this reporting (biennial or once during a four-year replenishment period) will be decided later on the basis of cost calculations.

Parameter	Agency	Current status of development of indicators and tools used to measure performance	Information source and frequency of reporting
10. Robustness of program result indicators and tracking tools	GEF Secretariat, task forces	An assessment tool will be developed by the Evaluation Office on the basis of the SMART criteria for indicators. Special attention will be given to GEF-specific priorities such as scientific soundness of the system (particularly when there is a heavy reliance on the use of proxy indicators), validity of aggregation, and extent to which GEF indicator systems are properly integrated with IA and ExA monitoring systems.	Focal area program evaluations, conducted every four years, will include an assessment of focal area indicators and tracking tools.
11. Quality assurance of project M&E arrangements at entry	GEF Secretariat, IAs, ExAs	The 2005 APR developed a rating tool using 13 parameters of quality for the M&E at entry.	Reporting will be on a biennial basis.
12. Quality of project M&E during implementation	IAs, ExAs	Average verified ratings obtained during terminal evaluation reviews. The Evaluation Office has been tracking differences between its ratings and those given by Agencies' independent evaluation offices. As differences diminish, the Office will verify ratings on a sample of projects and will accept and report to the GEF Council on the basis of Agencies' independent evaluation office ratings.	Reporting will be on an annual basis. Every four years, as part of the focal area program studies, a sample of projects will be selected for field verification.
13. Quality of project terminal evaluation			
Learning			
14. Management action record.	GEF Secretariat, IAs, ExAs	The MAR tracks and rates actions taken to address Council decisions on recommendations in Evaluation Office reports. The indicator for this parameter will be the average rating reported in the MAR.	Reporting will be on an annual basis.
15. Quality of lessons in terminal evaluations	IAs, ExAs	Average ratings of the Office's terminal evaluation reviews of lessons presented in project evaluation reports.	

Note: SMART = specific, measurable, achievable, realistic, and timely.

Annex F. Management Response

F.1 Introduction

We welcome the *GEF Annual Performance Report 2006*, prepared by the GEF Evaluation Office. The 2006 APR provides an assessment of project outcomes and sustainability, delays in project completion, materialization of cofinancing, and quality of the monitoring and evaluation arrangements at the point of Chief Executive Officer endorsement. This year's APR also looks for the first time at the quality of supervision of GEF projects by the three Implementing Agencies.

According to the report, the assessment of project outcomes, project sustainability, and delays in project completion relies on an analysis of 66 projects for which terminal evaluation reports were submitted by the IAs to the Evaluation Office in FY 2006. For assessment of the materialization of cofinancing, all the 182 terminal evaluation reports submitted since January 2001 were considered. Of these, 118 (65 percent) terminal evaluation reports provided information on actual materialization of cofinancing. The assessment of quality of M&E arrangements at the point of CEO endorsement is based on the 74 full-size projects that were CEO endorsed in FY 2005.

The report acknowledges that the findings presented have several limitations due to the fact that they are based on terminal evaluation reviews, which are in turn based on information provided by the IA terminal evaluation reports. Nevertheless, the discussion of the issues assessed in the APR 2006 provides a series of useful insights that can contribute to portfolio management at the GEF.

F.2 Results

We welcome the report's conclusion that completed GEF projects remain on target to achieve the 75 percent satisfactory outcomes as agreed upon in the GEF-4 replenishment agreement. According to the report, out of the completed GEF projects submitted in FY 2006, 84 percent were rated moderately satisfactory or above in achievement of outcomes; moreover, of the total investment in the rated projects ($254 million), 88 percent ($224 million) was allocated to projects that were rated moderately satisfactory or above.

In terms of sustainability, 65 percent of projects were rated moderately likely or above in sustainability of outcomes. Of the total investment in the rated projects ($218 million), 60 percent ($131 million) was invested in projects that were rated moderately likely or above in sustainability. It would be helpful if the Evaluation Office could provide recommendations for improving on the 65 percent satisfactory ratings for sustainability of outcomes through an examination of whether there is potential for improvement through GEF program strategies, project design, financing plans, or supervision.

We also welcome the additional analysis undertaken by the Evaluation Office to assess the extent to which projects that were rated moderately satisfactory or above in achievement of outcomes were also rated as moderately likely or above in sustainability of outcomes. Of the rated projects from the FY 2006 cohort, 61 percent were rated moderately satisfactory or above in outcomes and moderately likely or above in sustainability.

F.3 Processes

We are pleased that the materialization of cofinancing is on track. For terminal evaluation reports submitted in FY 2006, the average rate of reported materialization was 114 percent of the promised cofinancing.

For this report, the Evaluation Office carried out a pilot assessment of project supervision by IAs. We are concerned with the conclusion that "UNEP does not adequately supervise two-thirds of its sampled projects, although improvements have been realized after the appointment of a portfolio manager." However, it should be noted that the assessment also found that "Fiduciary requirements are generally met." In relation to this assessment, UNEP has noted the following:

> During the past year UNEP has taken a number of steps to both take stock of issues related to project supervision and also to improve its systems for GEF project implementation oversight. UNEP therefore welcomes the GEF Pilot Assessment as it provides an additional independent view of current project supervision practices and a baseline against which the organization will be able to measure progress towards the improvement of quality of project supervision in the future.

We are concerned with the conclusion that portfolio monitoring by the GEF Agencies tends to rate projects fully satisfactory despite important problems noted in the monitoring information on the same projects. According to the report, there is a marked tendency to rate projects fully satisfactory despite problems reported in areas such as implementation delays, government commitment, or long-term sustainability.

This assessment was carried out as part of the pilot assessment of supervisions. For the 55 project implementation reports reviewed from 2005 and 2006, only 3 had been flagged in PIR ratings as moderately unsatisfactory or lower (grant value of $10.0 million). The desk review identified another 19 projects (grant value of $85.8 million), which could be considered at risk based on issues described in the narrative section of the PIRs, as well as reported performance of project components. It is, however, encouraging that the 2006 PIRs showed some improvement over 2005 with evidence that IA managers had begun to ask staff to justify ratings that seemed inconsistent with reported problems.

F.4 Monitoring and Evaluation

We are pleased with the finding that 78 percent of projects were rated moderately satisfactory or above on quality of monitoring. We agree that there is scope for improvement on appropriate indicators and baseline data, as well as for better quality at entry for funding of M&E. We believe that the development of strategic programming for GEF-4 focal area strategies and their associated indicators (GEF 2007) will allow the Agencies to make further improvements in this area.

We are concerned with the Evaluation Office's conclusion that a substantial proportion of terminal evaluation reports do not adequately cover issues such as sustainability, cofinancing, and M&E. Out of 66 terminal evaluation reports submitted in FY 2006, 30 percent did not provide sufficient information on project monitoring, 18 percent did not provide sufficient information on sustainability outcomes, and 29 percent did not provide information on the materialization of cofinancing.

In considering the three issues that have been raised, it would have been helpful if the APR had stated clearly the standards or criteria that were used to determine the adequacy of information for rating sustainability.

F.5 Other Findings

The World Bank would like to provide specific responses to a section of the APR document as follows:

a. It is important to provide an accurate description of the status of "blended" IBRD/IDA-GEF operations [chapter 5]. GEF projects are not recorded as IBRD/IDA operations; rather, they are considered supplements of "parent" IBRD/IDA projects. All GEF projects, whether blended or not, are part of a distinct GEF product line and are coded accordingly with separate project identification numbers. The systems issues identified in the report are complex; they relate primarily to concerns by the Bank's Quality Assurance Group about double-counting the "parent" IBRD/IDA operation in its portfolio assessments.

b. The World Bank has been proactive in restructuring or canceling "nonperforming projects" unlike the characterization in the APR [chapter 5]. There is a specific indicator of portfolio performance for this purpose, called the proactivity index, which is monitored for all Bank operations, including GEF grants. Furthermore, over the past decade, the Bank has dropped or canceled more than $300 million worth of GEF grants—not including dozens of projects with partial cancellations. Refer to World Bank (2006).

F.6 Recommendations

Recommendation 1: UNDP and UNEP need to involve social and institutional expertise in project supervision.

UNDP notes this recommendation and the findings on which it is based. While UNDP already involves social and institutional expertise in project supervision, UNDP will examine how it can further strengthen this.

UNEP acknowledges that social and institutional expertise is important for adequate supervision of specific projects. As such, UNEP will include guidance on this issue in the section on project supervision standards of its GEF Operations Manual.

Conclusion 4 of the APR 2006 indicates that at present only the World Bank has a formal system for screening projects for potential unintended social or environmental impacts that would need to be mitigated and supervised during implementation. UNEP would like to clarify that potential negative social or environmental effects are routinely considered in the formal project design review process (that is, in the project quality-at-entry checklist); however, because UNEP's projects do not generally involve development interventions on the ground, the application of complex social and environmental safeguards such as those required for World Bank projects is not warranted. UNEP management is nonetheless willing to consider additional cost-effective measures to identify and mitigate these risks.

Recommendation 2: Special attention is required to ensure continued and improved supervision in the new GEF project cycle by ensuring adequate funding in project fees.

The results-based management framework that is presented for Council discussion will provide the platform for the Secretariat to develop tools to monitor the portfolio, in coordination with the GEF Agencies. While developing these tools, care will be taken to ensure that they do not duplicate, but rather build on, the monitoring activities undertaken by the Agencies.

The Agencies will ensure that the appropriate level of resources received from fees are directed toward supervision of projects.[1]

Recommendation 3: UNEP should develop a structural approach to supervision of its GEF portfolio.

UNEP notes that the above recommendation is in line with its management approach to strengthen overall project supervision. UNEP is therefore pleased to report on the following changes it has put in place over the last year to improve the overall supervision of its GEF portfolio:

a. UNEP has undertaken an intensive exercise of improving human resource capacity for the supervision of its GEF project portfolio. This has included the strengthening of its human resource capacity in persistent organic pollutants, climate change, and international waters by recruiting senior professionals in FY 2006, which would show improvement in project supervision in the next fiscal year. Management has also taken the decision to further strengthen UNEP's international waters and biodiversity teams and will soon conclude new recruitments to further develop the capacity for each of the teams where additional human resource needs have been identified.

b. UNEP has also been developing standard processes to instill a more structured approach to project implementation supervision as follows. UNEP developed and applied an enhanced GEF PIR process (piloted in GEF FY 2006). The organization improved the design and has started implementation of several tools to meet the requirements of the GEF M&E Policy set in place during the previous year. A project-at-risk system was developed and piloted from GEF FY 2005 for the implementation review of all its GEF projects. This will provide a tool for alerting management of problem projects so that corrective actions can be undertaken rapidly. UNEP has carried out an analysis of the staff time required for overseeing different types of GEF projects. This analysis is being used to more evenly balance the workload to ensure adequate human resources are assigned to project supervision.

c. UNEP is also institutionalizing a process of annual project quality of supervision reviews which will be conducted by UNEP's portfolio manager in consultation with UNEP's Evaluation and Oversight Unit. Minimum standards for the oversight of GEF projects are being put in place which these annual reviews will focus on. Starting with the June 2007 work program, UNEP's GEF projects, once endorsed by the CEO, will include project supervision plans to be put in place before UNEP proceeds with implementation. These plans will establish project supervision tasks and their costings. A GEF Operations Manual is under development with an expected completion date of August 2007. The manual includes a section devoted to project supervision and will include the standards for both financial/fiduciary and substantive aspects of project supervision and a description of the associated processes. It will also include standards to take into consideration, and social and institutional issues where these issues may affect project implementation quality. The manual will be shared with the GEF Evaluation Office for an ex ante review of UNEP's proposed GEF project management and supervision standards. This will be used to facilitate training of staff responsible for providing GEF project implementation supervision to ensure a more uniform approach.

Although UNEP is improving its project supervision practices as noted above, UNEP makes the following observations concerning the findings of the Evaluation Office pilot assessment:

a. The pilot assessment found that UNEP has satisfactory standards concerning supervision of fiduciary/safeguard aspects. This finding is that 13 projects out of 16 in the sample were rated moderately satisfactory or above (82 percent of projects scored 3 or above on the 6-point scale, where 6 is the lowest rating).

b. There are a number of factual errors in the pilot assessment which affect the ratings of individual projects. These could have been corrected if Agencies had been given the opportunity to review the details of the APR findings in draft form. For example, one project was rated 6 for realism of project performance rating, while that project had been rated by UNEP's PIR as being a project at risk (substantial risk rating) and had also been rated in the same PIR as moderately unsatisfactory for progress toward development objectives, demonstrating the realism in UNEP's project performance rating. Another example is that several projects that had just begun implementation by June 2004 were given a poor score for overall implementation performance prior to the period under review. In a few cases, the Evaluation Office consultants made mistakes in what they recorded during the interviews with UNEP task managers. For example, in a UNEP project reviewed, the GEF evaluation states that there was no training provided to the project staff, yet this was done and is included in the documentation provided to the evaluators.[2]

c. Concerning candor in Agency PIR ratings, it should be noted that two UNEP projects in the pilot assessment sample now have terminal evaluations which were submitted to the Evaluation Office. The independent evaluation ratings are broadly consistent with the PIR rating provided by the task managers, while the Evaluation Office assessment concludes that the task managers had significantly overrated performance in the PIR.

d. For future exercises, it is suggested that a clear explanation of the factors considered in the supervision assessment are given and comments to support the ratings are disclosed. This would help the Agencies understand the basis for the rating and help them to identify ways in which project supervision can be improved.

Recommendation 4: All GEF Agencies will need to ensure that terminal evaluation reports include adequate information on sustainability of outcomes, quality of M&E systems, and cofinancing.

Evaluation reports prepared for GEF-financed projects are expected to meet minimum requirement 3 of the GEF Monitoring and Evaluation Policy. In line with these requirements, Agencies will ensure that terminal evaluation reports include information on sustainability of outcomes, quality of M&E systems, and assessment of cofinancing realized. However, we would appreciate more guidance from the Evaluation Office to identify concrete steps to be taken to improve the quality and level of information presented in terminal evaluation reports.

We welcome the Agency Performance and Outcomes Matrix (scorecard) presented by the Evaluation Office and look forward to receiving the matrix in next year's APR.

Notes

1. It is important to note that Agency fees cover the full life cycle of the projects—preparation, supervision, and evaluation phases. In addition, while the fees are indeed provided on a project basis, the Agencies manage the application of resources received from the fee on a portfolio basis, redirecting resources toward projects in their respective portfolios that require more due diligence efforts, on the basis of factors such as country and sector context, project complexity, and other institutional operational practices and budget guidelines.

2. **Response from the GEF Evaluation Office:** The GEF Evaluation Office reviewed all nine projects for which UNEP had identified factual errors in the pilot assessment. UNEP identified a total of 30 errors, of which in five instances, the review found UNEP's identification to be correct. These five factual errors have been rectified in this report. None of these errors was sufficiently substantive to merit a change in the ratings of the respective projects.

References

The GEF Council documents cited here (indicated with the designation "GEF/C.xx") are available on the GEF Web site, www.thegef.org, under Documents/ Council Documents. GEF Evaluation Office documents can be found on the GEF Evaluation Office Web site, www.gefeo.org, under Publications.

GEF (Global Environment Facility). 1996. "Operational Program Number 10 Contaminant-Based Operational Program." http://thegef.org/Operational_Policies/Operational_Programs/OP_10_English.pdf.

———. 2000. "GEF Project Cycle." GEF/C.16/Inf.7.

———. 2002. "An Interim Report on a Revision of the Fee Structure." GEF/C.20/5.

———. 2004. "Proposal for Revising the Fee System." GEF/C.23/8.

———. 2005b. "Policy Recommendations for the Fourth Replenishment of the GEF Trust Fund." GEF/R.4/30.

———. 2006. "Summary of Negotiations on the Fourth Replenishment of the GEF Trust Fund." GEF/C.29/3.

———. 2007. "Focal Area Strategies and Strategic Programming for GEF-4." GEF/C.31/10.

GEF Council (Global Environment Facility Council). 2006. "Joint Summary of the Chairs GEF Council Meeting December 5–8, 2006." GEF/C.30.

GEF EO (Global Environment Facility Evaluation Office). 2006a. *The GEF Monitoring and Evaluation Policy.* Evaluation Document No. 1. Washington, DC.

———. 2006b. *The Role of Local Benefits in Global Environmental Programs.* Evaluation Report No. 30. Washington, DC.

———. 2007a. *Evaluation of Incremental Cost Assessment.* Evaluation Report No. 34. Washington, DC.

———. 2007b. *Joint Evaluation of the GEF Activity Cycle and Modalities.* Evaluation Report No. 33. Washington, DC.

OECD DAC (Organisation for Economic Co-operation and Development Development Assistance Committee). 2002. *Glossary of Key Terms in Evaluation and Results Based Management.* www.oecd.org/dataoecd/29/21/2754804.pdf.

World Bank. 1997. *Quality of Supervision Report.* Washington, DC.

———. 2006. "Summary of IBRD Project and IA Fee Cancellations." GEF/C.28/Inf.14.

WB QAG (World Bank Quality Assurance Group). 2001. *Supervision Quality in FY00.* Washington, DC.

———. 2005. *Annual Performance Report on Portfolio Performance.* Washington, DC.

———. 2007. *Quality of Supervision in FY05–06 (QSA7): A QAG Assessment.* Washington, DC.

GEF Evaluation Office Publications

Number	Title	Year
Evaluation Reports		
39	Joint Evaluation of the GEF Small Grants Programme	2008
38	GEF Annual Performance Report 2006	2008
37	GEF Country Portfolio Evaluation: Samoa (1992–2007)	2008
36	GEF Country Portfolio Evaluation: The Philippines (1992–2007)	2008
35	Evaluation of the Experience of Executing Agencies under Expanded Opportunities in the GEF	2007
34	Evaluation of Incremental Cost Assessment	2007
33	Joint Evaluation of the GEF Activity Cycle and Modalities	2007
32	GEF Country Portfolio Evaluation: Costa Rica (1992–2005)	2007
31	Annual Performance Report 2005	2006
30	The Role of Local Benefits in Global Environmental Programs	2006
29	Annual Performance Report 2004	2005
28	Evaluation of GEF Support for Biosafety	2006
	Third Overall Performance Study	2005
	GEF Integrated Ecosystem Management Program Study	2005
	Biodiversity Program Study	2004
	Climate Change Program Study	2004
	International Waters Program Study	2004
Evaluation Documents		
ED-1	The GEF Evaluation and Monitoring Policy	2006
ED-2	GEF Evaluation Office Ethical Guidelines	2008

Evaluation Office
Global Environment Facility

Global Environment Facility
Evaluation Office
1818 H Street, NW
Washington, DC 20433
USA

www.gefeo.org