

# ANNUAL PERFORMANCE REPORT 2005



GLOBAL ENVIRONMENT FACILITY  
EVALUATION OFFICE



**Global Environment Facility  
Evaluation Office**

# **GEF Annual Performance Report 2005**

**August 2006**

Evaluation Report No. 31

© 2006 Global Environment Facility Evaluation Office  
1818 H Street, NW  
Washington, DC 20433  
Internet: [www.thegef.org](http://www.thegef.org)  
Email: [gefevaluation@thegef.org](mailto:gefevaluation@thegef.org)

All rights reserved.

The findings, interpretations, and conclusions expressed herein are those of the authors and do not necessarily reflect the views of the GEF Council or the governments they represent.

The GEF Evaluation Office does not guarantee the accuracy of the data included in this work. The boundaries, colors, denominations, and other information shown on any map in this work do not imply any judgment on the part of the GEF concerning the legal status of any territory or the endorsement or acceptance of such boundaries.

### **Rights and Permissions**

The material in this work is copyrighted. Copying and/or transmitting portions or all of this work without permission may be a violation of applicable law. The GEF encourages dissemination of its work and will normally grant permission promptly.

ISBN: 1-933992-04-2

### **Global Environment Facility**

Director of the GEF Evaluation Office: Robert D. van den Berg

Task Manager: Aaron Zazueta, Senior Evaluation Officer

Evaluation Team:

Neeraj Negi, Consultant

Frederick Swartzendruber, Consultant

Antonio del Monaco, Evaluation Specialist

Editing and layout: Nita Congress

Printing: Graphic Communications

Cover photo: © Haroldo Castro/Conservation International

Evaluation Report No. 31

A FREE PUBLICATION

# Contents

<b>Acronyms and Abbreviations .....</b>	<b>vi</b>
<b>Foreword.....</b>	<b>vii</b>

## CONTEXT

<b>1. Background, Main Conclusions, and Recommendations .....</b>	<b>1</b>
1.1 Background.....	1
1.2 Main Conclusions.....	2
1.3 Recommendations.....	6
1.4 Issues for the Future.....	6
<b>2. Scope and Methodology .....</b>	<b>8</b>
2.1 Scope.....	8
2.2 Methodology.....	9

## RESULTS

<b>3. Project Outcomes and Sustainability .....</b>	<b>11</b>
3.1 Ratings Approach .....	11
3.2 Project Outcomes.....	11
3.3 Sustainability of Project Outcomes.....	14

## PROCESSES

<b>4. Factors Affecting Attainment of Project Results .....</b>	<b>17</b>
4.1 Materialization of Cofinancing .....	17
4.2 Relationship between Project Funding and Project Outcomes and Sustainability.....	19
4.3 Delays in Project Completion.....	20

## MONITORING AND EVALUATION

<b>5. Assessment of Project-at-Risk Systems of GEF Partner Agencies.....</b>	<b>21</b>
5.1 Assessment Approach .....	21
5.2 GEF Requirements for Project Preparation and Implementation .....	22
5.3 Agency Reporting Systems.....	23
5.4 Ensuring Comparability.....	26

<b>6. Quality of Project Monitoring.....</b>	<b>28</b>
6.1 Quality of Monitoring during Project Implementation .....	28
6.2 Systems to Ensure Quality of M&E Arrangements at Entry .....	31
<b>7. Quality of Terminal Evaluation Reports.....</b>	<b>38</b>
7.1 Assessment Approach .....	38
7.2 Findings.....	39
7.3 IA Versus GEF Evaluation Office Ratings.....	41
<b>8. Management Action Records .....</b>	<b>44</b>

## ANNEXES

A. Terminal Evaluation Report Review Guidelines.....	46
B. Characteristics of Agency Project-at-Risk Monitoring Systems .....	50
C. Inventory of Project-at-Risk Monitoring Systems .....	51
D. Methodology and Parameters for Reviewing M&E Arrangements at Entry.....	53
E. Portfolio Performance on M&E Arrangements at Entry, by Parameter .....	56
F. Ratings for Terminal Evaluation Reports Reviewed, by Project .....	60
G. Quality of Terminal Evaluation Reports, by IA and Assessment Criteria.....	62
H. Management Response .....	64

## BOXES

3.1: The Rating Scale.....	12
5.1: GEF Partner Agencies .....	22
5.2: World Bank Risk Rating Categories.....	24
5.3: Enhancing the Independence of Internal Risk Monitoring.....	25
5.4: Independence through External Risk Monitoring.....	25
6.1: IA Initiatives to Improve the Quality of Monitoring .....	30
7.1: Changes at UNEP and UNDP to Improve the Quality of Project Terminal Evaluation Reports .....	40
7.2: Binary Rating Scale.....	42

## FIGURES

3.1: Number of Terminal Evaluation Reports.....	11
3.2: Project Outcome Ratings.....	12
3.3: Ratings for Sustainability of Outcomes .....	14
3.4: Sustainability of Project Outcomes, by Assessment Criteria.....	16
4.1: Materialization of Cofinancing Trend for Completed Projects .....	19
4.2: Relationship between Outcomes and Leveraged Funds .....	19
4.3: Effect of Delay on Project Outcomes and Sustainability.....	20
6.1: Quality of Project M&E Systems .....	29
6.2: Number of Projects .....	32
7.1: Quality of Terminal Evaluation Reports .....	39

7.2: Percentage of Terminal Evaluation Reports Rated Moderately Satisfactory or Above, by Assessment Criteria.....	40
7.3: Percentage of Terminal Evaluation Reports Not Providing Sufficient Information on Quality of Project M&E System .....	41
7.4: IA and GEF Evaluation Office Ratings on Outcomes and Sustainability.....	41
8.1: Management Action Record.....	44

## **TABLES**

3.1: Project Outcomes by IA (number of projects).....	12
3.2: Project Outcomes by Focal Area (number of projects).....	12
3.3: Project Outcomes by IA (GEF investment in million \$).....	13
3.4: Project Outcomes by Focal Area (GEF investment in million \$).....	13
3.5: Outcomes Sustainability by IA (number of projects) .....	15
3.6: Outcomes Sustainability by Focal Area (number of projects).....	15
3.7: Outcomes Sustainability by IA (GEF investment in million \$) .....	15
3.8: Outcomes Sustainability by Focal Area (GEF investment in million \$) .....	15
4.1: Actual Compared to Promised Cofinancing.....	18
4.2: Cofinancing Ratios and Actual Cofinancing by Region, IA, and Focal Area .....	18
5.1: Agency Risk-Monitoring Inventory .....	27
6.1: Quality of Project Monitoring by IA.....	30
6.2: Quality of Project Monitoring by Focal Area .....	30
6.3: Projects with Overall Compliance by IA .....	33
6.4: Projects with Overall Compliance by Focal Area .....	34
7.1: Ratings on Quality of Terminal Evaluation Reports, by IA .....	39
7.2: Average Difference between IA and GEF Evaluation Office Ratings on a Binary Scale.....	43
7.3: Difference between IA and GEF Evaluation Office Ratings on a Six-Point Scale .....	43

## Acronyms and Abbreviations

---

<b>APR</b>	Annual Performance Report	<b>MAR</b>	Management Action Record
<b>CEO</b>	Chief Executive Officer	<b>MIS</b>	Management Information System
<b>EA</b>	Executing Agency	<b>OD</b>	Ozone Depletion
<b>EBRD</b>	European Bank for Reconstruction and Development	<b>PIR</b>	Project Implementation Review
<b>FY</b>	Fiscal Year	<b>PMIS</b>	Project Management Information System
<b>GEF</b>	Global Environment Facility	<b>POP</b>	Persistent Organic Pollutant
<b>IA</b>	Implementing Agency	<b>STAP</b>	Scientific and Technical Advisory Panel
<b>IEG</b>	Independent Evaluation Group	<b>UNDP</b>	United Nations Development Programme
<b>M&amp;E</b>	Monitoring and Evaluation	<b>UNEP</b>	United Nations Environment Programme





## Foreword

---

This is the second Annual Performance Report (APR) of the Global Environment Facility (GEF) Evaluation Office on the results of the GEF, as reported on through the monitoring and evaluation (M&E) systems of the GEF partners. It provides an independent assessment of these results and also focuses on performance issues. The second report offers substantially more information than the first one: More end-of-project evaluation reports become available as the GEF matures. Additionally, M&E in the GEF has become more reliable and is reporting on achievements in increasingly promising ways. Given the encouraging developments in ensuring better indicators throughout project portfolios, we hope to build these annual performance reports into a solid overview of GEF results over time.

The continuing emphasis on accountability for results—and on learning from past experiences—has led many international organizations to report on results in annual reports. With this second APR, the GEF has firmly established its commitment in this regard. In addition, the GEF Secretariat is now fully engaged in portfolio monitoring and reporting on progress to the GEF Council.

The Council reviewed the *Annual Performance Report 2005* and the management response (included as annex H) at its 28th meeting in June 2006, and made four decisions. It requested that the GEF Secretariat redraft project review guide-

lines and standards to ensure compliance with the new M&E minimum requirements, and consider ways to enhance the contribution of Scientific and Technical Advisory Panel reviews during the process. Second, the GEF Secretariat should support focal area task forces with corporate resources to continue the development of indicators and tracking tools to measure the results of GEF operations in the various focal areas. Third, the Evaluation Office should continue its oversight of project-at-risk systems and interact with the Implementing and Executing Agencies and report to the GEF Council on progress. Lastly, the Implementing and Executing Agencies should share, in accordance with their operational policies and procedures, mid-term and terminal evaluations with the GEF focal points in a timely manner.

The Council also requested that the GEF partner agencies continue to follow up on the recommendations made in last year's APR regarding the need to improve terminal evaluation reports. Discussions and decisions show that the APR has become a valuable source of information and basis for decisions in a short time.

The task manager for the APR was Aaron Zazueta, Senior Evaluation Officer. He was supported by Evaluation Office staff and independent consultants. He developed the framework for the report and the scope for the different chapters, oversaw the review of project terminal evaluation reports,

and ensured the consistency of the APR by closely supervising the drafting of its various chapters. Chapter 5 was drafted by Frederick Swartzendruber, consultant. Antonio del Monaco, Evaluation Specialist, coordinated the review of the terminal evaluation reports, which were reviewed by various consultants and colleagues in the GEF Evaluation Office. Neeraj Kumar Negi, consultant, and Frederick Swartzendruber carried out the review of systems to ensure quality of M&E arrangements at entry.

Various intermediary products were discussed with the GEF Secretariat and the 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. The final responsibility for this report remains firmly with this office.

A handwritten signature in black ink, appearing to read 'Rob D. van den Berg', with a stylized flourish extending to the right.

Rob D. van den Berg  
Director, Evaluation Office

## CONTEXT

# 1. Background, Main Conclusions, and Recommendations

## 1.1 Background

This is the second 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 focus more on assessing the results of GEF activities and overseeing monitoring and evaluation (M&E) operations. The higher quality of terminal evaluation reports submitted by the IAs in fiscal year (FY) 2005 also allowed for the inclusion in this APR of an assessment of the extent to which projects are achieving their objectives. Furthermore, this report presents a detailed account of some aspects of project results, of processes that may affect project results, and of M&E arrangements across the GEF partnership.

For the first time, the APR includes an assessment of project outcomes, project sustainability, delays in project completion, materialization of cofinancing, and quality of M&E arrangements at the point of GEF Chief Executive Officer (CEO) endorsement. For the assessment of project outcomes, project sustainability, and delays in project completion, 41 projects were considered for which terminal evaluation reports were submitted by the IAs to the Evaluation Office in FY 2005. Altogether, the GEF had invested \$260 million in

these 41 projects.<sup>1</sup> For assessment of materialization of cofinancing, all 116 terminal evaluation reports submitted after January 2001 were considered. Of these, 70 (60 percent) provided information on actual materialization of cofinancing. The GEF has invested a total of \$380 million in these 70 projects and has been able to leverage an additional \$1.77 billion in cofinancing. For assessment of quality of M&E arrangements at point of CEO endorsement, the 74 full-sized projects that were CEO-endorsed in FY 2005 were considered. The GEF has approved an overall investment of \$535 million in these projects.

As in the 2004 APR, this APR includes an assessment of quality of project monitoring and of quality of terminal evaluation reports. For this assessment, 83 terminal evaluations were considered, of which 41 were submitted in FY 2005 and 42 in FY 2004. This allowed comparisons of performance during these two years. The GEF invested about \$460 million in these 83 projects.

The findings presented have several limitations. Most are based on the terminal evaluation reviews, which are in turn based on the information provided by IA terminal evaluation reports. This introduces uncertainty into the verification process, which the Evaluation Office seeks to mitigate by incorporating in its reviews any per-

<sup>1</sup>Note that all dollar figures cited in this report are US\$.

tinient information it has independently gathered through other evaluations. The Evaluation Office is also seeking to improve the independence of terminal evaluation reports by more fully involving the central evaluation units of partner agencies in the process. Another limitation is that the assessment of partners' project-at-risk systems is based on agency self-reports.

For many issues on which performance is being reported in this APR, information is presently available only for FY 2005. Comparisons across years will be possible in future APRs. For assessment of quality of terminal evaluation reports, data are available for FY 2004 and 2005. Although this allows for comparison of performance in these two 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. These limitations will be mitigated in future with accumulation and availability of data for more cohorts.

Project outcomes and their sustainability is one of the topics addressed this year for the first time by the APR. A high proportion of the projects, both in terms of their number and the financial resources allocated to them, were rated as moderately satisfactory or higher.<sup>2</sup> This is a very positive finding. Nonetheless, a more authoritative assessment of the results of GEF operations will be possible as information on more projects becomes available for analysis in the coming years. It should also be mentioned that, despite the positive ratings for outcomes and sustainability, deficient project and program monitoring across the GEF system ham-

---

<sup>2</sup>The Evaluation Office uses a six-point rating scale in assessing project outcomes, sustainability, and quality of terminal evaluation reports. This scale classifies performance on a specific dimension into six gradations: highly satisfactory or highly likely, satisfactory or likely, moderately satisfactory or moderately likely, moderately unsatisfactory or moderately unlikely, unsatisfactory or unlikely, or highly unsatisfactory or highly unlikely.

pers efforts to aggregate results. Only by putting in place robust M&E systems at the project and program levels will the GEF be able to demonstrate the extent of its contributions toward addressing critical global environmental problems.

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. MARs are discussed in chapter 8 of this report as an M&E tool enabling GEF learning from evaluation reports. MARs are published and updated on the GEF Evaluation Office website, which can be accessed through [www.thegef.org](http://www.thegef.org).

## 1.2 Main Conclusions

### Results

***Conclusion 1: Most of the completed GEF projects assessed this year have acceptable performance in terms of outcomes and sustainability.***

**Attainment of project outcomes.** The Evaluation Office rated project outcomes based on the level of achievement of project objectives and expected outcomes. The key findings of this assessment are:

- Of the 41 GEF projects reviewed in FY 2005, 88 percent were rated moderately satisfactory or above in their outcomes.
- In terms of the effectiveness of use of GEF funds, 95 percent of the \$260 million allocated to the projects reviewed in FY 2005 went to projects that achieved moderately satisfactory or better outcomes.

**Sustainability of project outcomes.** The Evaluation Office rated sustainability based on four key criteria: financial resources, sociopolitical issues, institutional framework and governance, and replication. The key findings are:

- Seventy-six percent of the projects were rated moderately likely or above in terms of their sustainability. Of the 23 United Nations Development Programme (UNDP) projects that were assessed, 7 (30 percent) were in the moderately unlikely category—just below the level where project performance could be considered acceptable. This presents an opportunity for improvement.
- In terms of GEF funding, 80 percent of the allocated funds were for projects with a sustainability rating of moderately likely or better.
- Among the criteria used to determine sustainability, projects tend to be weaker in their financial viability.

## Processes

***Conclusion 2: The projects examined have realized almost all cofinancing promised at project inception, except for global projects and those in Africa.***

The key findings from the cofinancing analysis are:

- Most of the projects achieved the cofinancing promised at inception. On average, projects were promised \$4.4 per GEF dollar and achieved \$4.1 per GEF dollar.
- The projects with higher promised cofinancing as a percentage of GEF funds tended to meet the expected cofinancing better than projects with lower promised cofinancing as a percentage of GEF funds.
- The Latin America and the Caribbean region has the highest level of actual cofinancing, with 141 percent of promised cofinancing actually materializing. The lowest levels of actual cofinancing as a percentage of that promised are found among global projects and projects in Africa, for which 66 percent and 76 percent of promised cofinancing materialized, respectively.

***Conclusion 3: Excessive delay in project completion is associated with lower performance in terms of outcomes and sustainability.***

The analysis of the 41 projects reviewed by the Evaluation Office in FY 2005 shows that outcome and sustainability ratings tend to be lower for projects with completion delays greater than 24 months. However, this association does not imply causality because excessive delay in project completion is more likely to be a result of other factors than an underlying cause affecting outcomes and sustainability. The Evaluation Office will further analyze the underlying causes in other evaluations such as the Joint Evaluation of the GEF Activity Cycle and Modalities, as well as future APRs, to ascertain the extent of and specific forms in which project delay affects project outcomes and sustainability.

## Monitoring and Evaluation

***Conclusion 4: The quality of monitoring shows signs of improvement, but much work remains to be done.***

In this report, the Evaluation Office continues the analysis of quality of monitoring initiated in 2004. The assessment shows:

- Compared to FY 2004, there was an improvement in the quality of project monitoring systems in FY 2005. The number of projects with moderately satisfactory or better ratings increased from 39 percent in 2004 to 52 percent in 2005.
- The actions taken by the IAs to address weaknesses in project monitoring systems have led to improvements. However, 24 percent of the projects had monitoring systems rated as moderately unsatisfactory or worse, and 20 percent of the terminal evaluation reports submitted to the Evaluation Office did not provide sufficient information to rate project M&E; thus, there is considerable room for improvement.



**Conclusion 5: A substantial proportion of projects did not meet the 2003 minimum M&E requirements at entry and would not have met the minimum M&E requirements of the new M&E policy.**

An assessment of project compliance with the minimum M&E requirements at CEO endorsement, in which the M&E arrangements of all the 74 full-sized projects that were CEO-endorsed in 2005 were assessed, shows:

- Fifty-eight percent of the projects comply with the 2003 minimum requirements for M&E arrangements at the point of CEO endorsement.
- By IA, the UNDP projects have better ratings than those of the World Bank on some compliance parameters; across focal areas, climate change projects have better ratings than do biodiversity projects on some parameters. Differences in ratings among agencies are due to the level of management attention accorded to M&E issues. The differences in the ratings among focal areas are due to the level of technical difficulties encountered in monitoring.

**Conclusion 6: Due to gaps in the present project review process, M&E concerns are not being adequately addressed.**

The major gaps and weaknesses in the review process are as follows:

- At present, there is insufficient guidance for GEF Secretariat reviewers to address M&E issues adequately and consistently.
- Standards applied by GEF Secretariat reviewers vary.
- The 2003 minimum requirements for M&E were interpreted in a variety of ways, especially with regard to the identification of baseline data.
- Although focal area task forces are developing project-level indicators and tracking tools, these

are not yet sufficiently developed to adequately measure project-level results.

- Despite significant progress by focal area task forces in developing indicators and tracking tools, technical difficulties still must be overcome to adequately measure and aggregate results.

**Conclusion 7: The present project-at-risk systems at the GEF partner agencies vary greatly and may have to address such issues as insufficient frequency of observations, robustness and candor of assessments, overlap and redundancy, and independent validation of risk.**

Assessment of the project-at-risk systems of the GEF partner agencies addressed only the issue of system design as reported by the respective agency to the Evaluation Office. This assessment did not examine the actual internal reports to determine the degree of compliance with formal procedures. The key findings of, and issues identified by, this assessment are:

- Many partner agencies monitor projects-at-risk using a “warning flag” system which tracks self-rated project performance through a corporate management information system (MIS). These ratings are aggregated and rolled up for portfolio-level reporting. The project-at-risk assessment systems of development bank partner agencies generally have most of the desirable characteristics of such systems (for example, independent project-level risk validation), whereas the systems of other partner agencies may lack many of these characteristics.
- Insufficient frequency of observations undermines the reporting power inherent in an MIS—computer reports can be generated any time, yet the underlying data are often updated only once a year.
- It may be difficult to ensure the robustness and candor of self-assessment.

- Managers and staff worry about the proliferation of monitoring and reporting systems, overlap or redundancy, and staff reporting burdens.
- Most agencies lack formal arrangements for independent validation of self-reported project-at-risk assessments. Only the European Bank for Reconstruction and Development (EBRD) has a formal process of project-level risk validation independent of the business unit. In the World Bank, the Quality Assurance Group performs a similar function but at a more aggregate level.

***Conclusion 8: Overall quality of terminal evaluation reports is improving, but there are still some areas where major work is necessary.***

The Evaluation Office began rating the quality of project terminal evaluation reports in 2004, which allows for comparison with 2005 reports.

- Compared to FY 2004, there has been a marked improvement in the overall quality of terminal evaluation reports in FY 2005—especially of those submitted by UNDP and the World Bank.
- A detailed assessment of the quality of terminal evaluation reports using the Evaluation Office criteria shows that IAs are addressing most of the key issues the Evaluation Office identified last year.
- There is little difference in the outcome and sustainability ratings given by the Evaluation Office and the various IAs when a binary scale is used.<sup>3</sup> On the six-point scale, there is no difference between the ratings accorded by the Evaluation Office and the World Bank's Independent Evaluation Group (IEG); the United Nations

<sup>3</sup>The Evaluation Office's six-point rating scale can be converted to a binary scale by collapsing it into two modalities—acceptable (moderately satisfactory or moderately likely and above ratings) and unacceptable (moderately unsatisfactory or moderately unlikely and below ratings).

Environment Programme (UNEP) tends to rate its projects a point higher than does the Evaluation Office. Since many of the terminal evaluation reports submitted by UNDP did not provide outcome and sustainability ratings, robust inferences cannot be drawn about the overall reliability of its terminal evaluation ratings.

- The terminal evaluation reports continue to be weak in assessing quality of monitoring; this is especially true of terminal evaluation reports in the climate change focal area. They also frequently fail to report on actual costs, including total costs and a breakdown by activity of GEF financing and cofinancing by other sources. Thus, despite improvement in the overall quality of the terminal evaluation reports submitted by the IAs, there are gaps in the information provided.

## **Management Action Records**

The first exercise in tracking the rate of adoption of GEF Council decisions on evaluation reports through the MARs has been a mixed experience, which will need to be improved when the MARs are next presented to Council in June 2007. Differences of interpretation on how adoption should be rated caused delays on the GEF management (GEF Secretariat/IAs) side, which meant that the Evaluation Office received the MARs too late to verify management's ratings. Based on other evaluations and insights gained through a consultative process, the Evaluation Office has indicated in the MARs how it perceives the rate of adoption so far. The Evaluation Office is confident that, based on the experience gained through the process, it will be possible to present verified ratings to the Council in June 2007. The MARs have been published as an information document for the Council (GEF/ME/C.28/Inf.2).

Verification was possible on one rating. In the MAR of the 2004 APR, management assesses as

“medium” the rate of adoption of the June 2005 Council decision that the transparency of the GEF approval process needs to be improved. A medium rate means that there has been “some adoption in operational and policy work, but not to a significant degree in key areas.” This assessment is based on the work that has been done to upgrade the GEF’s MIS. Based on evidence the Evaluation Office gathered in field visits for the pilot Costa Rica Country Portfolio Evaluation and the Joint Evaluation of the GEF Activity Cycle and Modalities, it has been able to verify this assessment and has downgraded the rate of adoption to “negligible.” This downgrade is based largely on the fact that information on project status in the approval process is still not available in a systematic way: For project proponents at the country level, nothing has changed since before the Council’s 2005 decision. The Costa Rica Country Portfolio Evaluation proposes that the Council reinforce its decision, and the MAR on the 2004 APR underscores the need for this reinforcement.

### 1.3 Recommendations

**Recommendation 1: The GEF Secretariat should redraft project review guidelines and standards to ensure compliance with the new M&E minimum requirements. Further consideration should also be given to ways to enhance the contribution of STAP reviews during the process.**

Lack of guidance has been identified as a problem that has led reviewers to apply their own perspective rather than abiding by a standard for minimum M&E requirements. Consideration should also be given to a more clearly defined role for the Scientific and Technical Advisory Panel (STAP) roster reviewers in the assessment of scientific and technical aspects of project indicators.

The GEF Secretariat should modify the Proposal Agreement Review template by adding a separate

section for candor and realism of risk assessment. This will ensure that any risk-related issue flagged at any stage of the review is followed up on during later stages of project processing.

**Recommendation 2: The GEF Secretariat should support focal area task forces with corporate resources to develop indicators and tracking tools to measure the results of GEF operations in the various focal areas.**

In recent years, focal area task forces have undertaken various actions to develop tools to measure the environmental results of GEF operations. Such efforts need to be intensified to further the development of indicators and tracking tools. This will require corporate investments to address the technical challenges specific to each focal area, build consensus on indicators, define ways to roll up results at the portfolio level, and find ways to address issues related to attribution of outcomes to GEF initiatives.

The ongoing work of the IAs to improve the quality of terminal evaluations should continue. IA terminal evaluation reports still have major information gaps. They are weak in terms of assessing project monitoring systems and in reporting actual project costs, including total costs and a breakdown by activity of GEF funding and cofunding. UNDP needs to fully engage its central evaluation group in the process, and UNEP needs to further enhance the independence of its central evaluation group to improve the quality of its terminal evaluations and address differences in ratings. Progress in this area will be tracked through the MAR of the previous APR and in assessments of future APRs.

### 1.4 Issues for the Future

The findings on the quality of M&E arrangements at entry confirm the importance of the minimum requirements for M&E of the *GEF Monitoring and*



*Evaluation Policy* (Evaluation Document 2006, No. 1). The new policy asks projects to provide adequate baseline information on indicators at the point of work program inclusion. The Evaluation Office will issue guidelines on the minimum requirements and how they will be evaluated in the coming APRs. The Evaluation Office will carry out another assessment of M&E quality assurance systems in the coming years to follow up on progress in implementation of the 2006 GEF M&E policy. It will also look at M&E during project implementation to ensure compliance with GEF M&E requirements both at entry and during project execution.

The review of the project-at-risk monitoring systems of GEF partner agencies shows that there is a need to enhance accounting and validation in existing agency frameworks. This is particularly the case for those frameworks that depend almost exclusively on self-assessment by management—a practice that is most prevalent among the partner agencies that are not development banks. Issues such as institutional culture and incentive structure to manage project risks also need to be assessed. In future assessments of the project-at-

risk systems of the partner agencies, the Evaluation Office will review actual internal reports to determine degree of compliance with the formal procedures of the system design.

The present analysis of the links between the promised level of cofinancing and project outcomes/sustainability is inconclusive. Although the analysis of the full set of projects does show an inverse relationship between levels of cofinancing and outcome/sustainability ratings, the relationship does not hold when the outliers are dropped from the analysis. However, there might be a point beyond which a higher level of promised cofinancing could be associated with a higher risk of a project losing sight of GEF objectives. As the number of projects with terminal evaluations increases, it will be possible to draw more robust inferences.

While causality is not implied, project delay might be a proxy indicator for the risk involved in projects. The Evaluation Office will look to further assess the association between implementation delays and outcomes and sustainability and will seek to identify the factors underlying this association.

# CONTEXT

## 2. Scope and Methodology

---

### 2.1 Scope

The APR provides an annual presentation of the performance of the GEF's completed projects, the processes that affect the accomplishment of results, and the 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 of the issues are addressed by the APR annually, some biennially; others are addressed whenever there is a need to do so. The 2005 APR includes:

- **An overview of the extent to which the GEF projects are achieving their objectives (chapter 3).** 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 2005 achieved expected outcomes and the sustainability of those outcomes. The APR will report annually on attainment of objectives and outcomes.
- **An analysis of the materialization of project cofinancing by region and IA (chapter 4).** This is accompanied by an analysis of the links between project cofinancing and project outcomes and sustainability. The Evaluation Office will report on these issues on an annual basis.
- **An analysis of the correlation between project implementation delays and project outcomes and sustainability (chapter 4).** The Evaluation Office will report on these issues on an annual basis.
- **An inventory of the present risk monitoring practices of the GEF IAs and EAs (chapter 5).** This inventory addresses the approaches used by the GEF partner agencies to track risk. It also identifies the strengths and weaknesses of agencies' current project-at-risk monitoring systems. The APR will report on these systems biennially.
- **An assessment of the quality of project monitoring (chapter 6).** This includes an examination of the quality of M&E at project completion and an assessment of the quality assurance systems of project M&E arrangements at CEO endorsement. The APR will report annually on the former; reporting on the latter will be performed biennially.
- **An assessment of the quality of terminal evaluation reports submitted by the IAs to the Evaluation Office (chapter 7).** This assessment is presented annually and provides information by focal area and IA. As part of the FY 2005 assessment of terminal evaluation report quality, this year's APR also assesses reporting on project M&E systems during project implementation.

## 2.2 Methodology

The project terminal evaluation reports submitted by the IAs and EAs 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 terminal evaluation 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, evaluate their internal consistency, and 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 and proposed rating modifications are examined by a senior evaluation officer in the respective focal area who either confirms or rejects the initial reviewer's conclusions and ratings. When projects are downgraded below moderately satisfactory (for outcomes) or below moderately likely (for sustainability), a second senior evaluation officer in the Evaluation Office also examines the review to ensure that the new ratings are justified. If the terminal evaluation reports provide insufficient information to make an assessment or verify the IA 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 mitigate uncertainty by incorporating in its reviews any pertinent information that it has independently gathered as part of other evaluations. Additionally, the Evaluation Office takes several approaches to targeted field verification, including setting aside time for field verification of projects during country visits carried out for other thematic evaluations. The Evaluation 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 Evaluation Office has monitored IEG ratings over the last two years and has found only minor differences in the ratings it assigns versus those given by the IEG. Therefore, the Evaluation Office will rely on the IEG's validation of terminal evaluation reports and, where necessary, complement these with a relatively minor effort to address GEF-specific information needs. During FY 2005, UNDP and UNEP took steps to involve their central evaluation groups more directly in the evaluation of GEF projects; additionally, UNEP worked to strengthen the independence of its central evaluation group. The Evaluation Office will continue its dialogue with the GEF partner agencies, while reviewing their

terminal evaluation reports and verifying their ratings.

An important issue related to the reliability of terminal evaluation reviews as a major source of information to the APR is whether the Evaluation Office is able to access the terminal evaluation reports of completed projects in a timely manner. To ensure this, the Evaluation Office has put in place a system to track terminal evaluation report submissions and has created a database of the terminal evaluation reports expected in any given year. Information from this database is sent to the IAs for verification, and the agencies are asked to submit terminal evaluation reports for completed projects or new dates for terminal evaluation reports for extended projects. This tracking system includes all GEF projects with an original completion date since January 2001; this covers the majority of GEF projects.

Beginning in FY 2005, the Evaluation Office is also tracking the time between project completion and submission of terminal evaluation reports and between terminal evaluation report completion and submission. The analysis of the 2005 data shows that, on average, terminal evaluation reports were received by the Evaluation Office 7.8 months after their completion and 10.5 months after project completion. This average is well within the 12-month limit set by the new *GEF Monitoring and Evaluation Policy* (Evaluation Document 2006, No. 1). However, 11 terminal evaluation reports (27 percent of a total of 41) were submitted a year after project completion: 1 of 6 from UNEP, 3 of 12 from the World Bank, and 7 of 23 from UNDP.

Special reviews were carried out on the quality assurance systems for M&E arrangements at CEO endorsement and on the status of project-at-risk systems. Annexes D and B provide a description of the methodology and the data collection instrument used, respectively, in these assessments.

The Evaluation Office used the F-Test and chi square test to assess differences among groups of projects, and the findings reported here as significant are at the 90 percent or higher confidence level. The regression analysis was used to assess the magnitude and direction of change associated with different variables. Nonetheless, the information obtained so far place some important limitations on the conclusions that can be derived. In some cases—such as the assessment of outcomes and sustainability, factors affecting sustainability, and the assessment of the implications of lag time during implementation—the numbers are still relatively small, hampering the ability to draw firm conclusions. In the case of the assessment of project monitoring at completion, a large proportion of terminal evaluation reports failed to provide sufficient information; thus, a significant proportion of those projects are not included in the analysis. Additionally, trends cannot yet be inferred, because only two years of data are currently available. These limitations will diminish in the coming years as IAs and EAs 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 enable meaningful assessment of progress and allow the Evaluation Office to make comparisons among agencies and focal areas.

The preliminary findings of this report were presented to and discussed on various occasions with the GEF Secretariat and IAs/EAs, including at focal area task force meetings held in November and December 2005 and the Interagency Meeting held in Washington, D.C., in January 2006. Individual reviews of project terminal evaluation reports and the results on quality of project M&E arrangements at entry were also shared with the IAs and the GEF Secretariat for factual verification.

## RESULTS

### 3. Project Outcomes and Sustainability

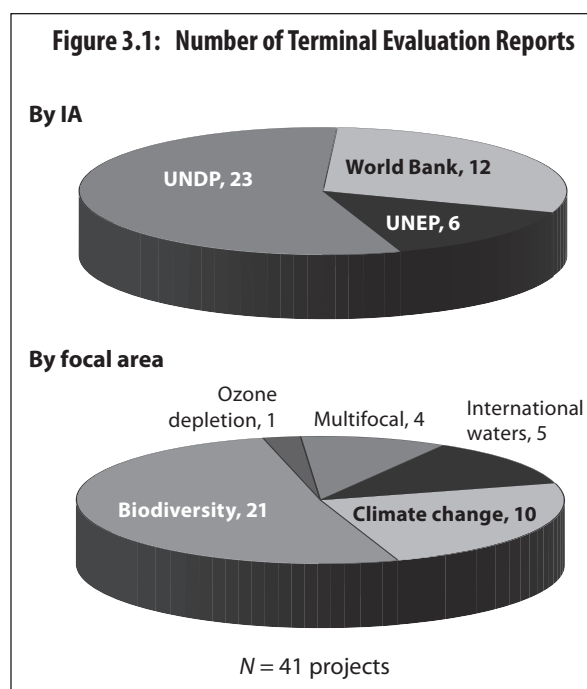
This chapter discusses verified ratings on outcomes and sustainability of the 41 projects for which terminal evaluation reports were submitted in FY 2005. Since this is the first year the GEF Evaluation Office has rated outcomes and sustainability, there is no baseline for comparison. The differences in ratings between the Implementing Agencies and the Evaluation Office are discussed in chapter 7.

Most GEF projects assessed this year seem, for the most part, to have attained their objectives. This is particularly true for World Bank and UNDP projects in all focal areas, with the exception of multifocal projects. This year's analysis also suggests that UNDP and UNEP need to give more attention to ensuring project outcome sustainability.

#### 3.1 Ratings Approach

The Evaluation Office rated project outcomes based on the level of achievement of project objectives and expected outcomes. It rated sustainability based on a set of key criteria contributing to sustainability such as financial resources, sociopolitical issues, institutional frameworks and governance, and replication. (See annex A for the actual criteria.) Of the 41 projects, there were 5 for which the Evaluation Office was unable to rate sustainability based on the information provided in the terminal evaluation reports. These included two projects for which the Evaluation Office also

had been unable to rate the outcomes. These terminal evaluation reports were excluded from the analysis. The distribution of terminal evaluations reviewed this year by focal area and IA is presented in figure 3.1.



#### 3.2 Project Outcomes

Most projects with terminal evaluation reports reviewed in 2005 (88 percent) were rated moderately satisfactory (see box 3.1) or above in their outcomes (see figure 3.2). Similarly, 95 percent of the GEF funds allocated for these 41 projects (\$260 million in total) went to projects that were



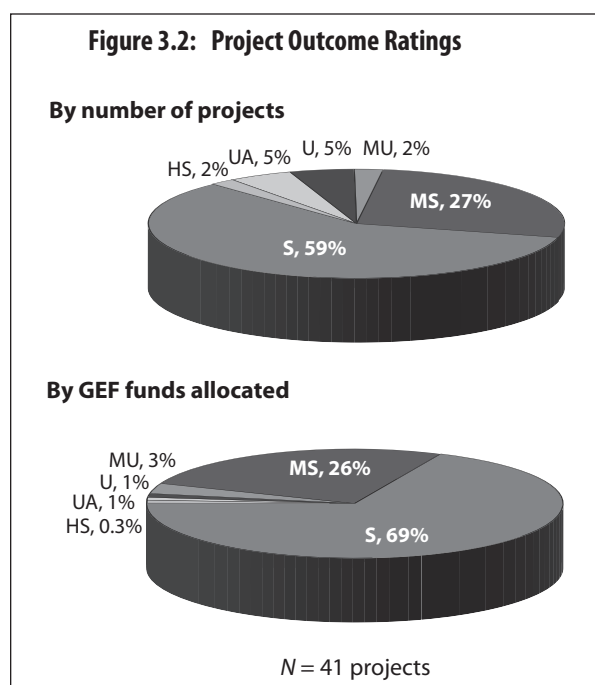
### Box 3.1: The Rating Scale

The GEF Evaluation Office uses a six-point scale for rating outcomes, sustainability, and quality of terminal evaluation reports. This scale classifies performance on a specific dimension into six gradations:

- Highly satisfactory (HS) or highly likely (HL)
- Satisfactory (S) or likely (L)
- Moderately satisfactory (MS) or moderately likely (ML)
- Moderately unsatisfactory (MU) or moderately unlikely (MU)
- Unsatisfactory (U) or unlikely (U)
- Highly unsatisfactory (HU) or highly unlikely (HU)

“Unable to assess” (UA) is used when the information contained in the terminal evaluation report did not allow for assessment.

rated moderately satisfactory or better in their outcomes. Only a small percentage (4 percent or \$10.4 million) of the GEF funds allocated for the projects reviewed in 2005 were for those rated worse than moderately satisfactory in their outcomes (see figure 3.2).



As shown in table 3.1, most projects for which the IAs submitted terminal evaluation reports in FY 2005 were rated as moderately satisfactory or above. However, the UNEP sample was quite small (six projects).

**Table 3.1: Project Outcomes by IA (number of projects)**

Rating	World Bank	UNDP	UNEP	Total
HS	0	1	0	1
S	9	13	2	24
MS	2	8	1	11
Subtotal	11	22	3	36
MU	1	0	0	1
U	0	0	2	2
HU	0	0	0	0
Subtotal	1	0	2	3
UA	0	1	1	2
Total	12	23	6	41

As shown in table 3.2, multifocal projects had the lowest outcome ratings. However, there were only four projects in that focal area reviewed this year. Note too that there was only one project reviewed in the ozone depletion (OD) focal area.

**Table 3.2: Project Outcomes by Focal Area (number of projects)**

Rating	Bio-diversity	Climate Change	Internat'l Waters	Multi-focal	OD
HS	1	0	0	0	0
S	13	5	5	0	1
MS	6	4	0	1	0
Subtotal	20	9	5	1	1
MU	1	0	0	0	0
U	0	0	0	2	0
HU	0	0	0	0	0
Subtotal	1	0	0	2	0
UA	0	1	0	1	0
Total	21	10	5	4	1

The GEF Evaluation Office will continue to assess the outcomes of UNEP and multifocal projects in future years to determine trends.

In terms of use of GEF funds, the bulk of the funds for World Bank (95 percent) and UNDP (98 percent) projects were used by projects rated satisfactory in their outcomes (see table 3.3).

**Table 3.3: Project Outcomes by IA**  
(GEF investment in million \$)

Rating	World Bank	UNDP	UNEP	Total
HS	0.0	0.8	0.0	0.8
S	110.6	64.7	1.4	176.7
MS	40.9	21.6	5.0	67.5
Subtotal	151.5	87.1	6.4	244.9
MU	8.7	0.0	0.0	8.7
U	0.0	0.0	1.7	1.7
HU	0.0	0.0	0.0	0.0
Subtotal	8.7	0.0	1.7	10.4
UA	0.0	1.7	1.3	3.0
<b>Total</b>	<b>160.2</b>	<b>88.8</b>	<b>9.4</b>	<b>258.3</b>

The majority of GEF funds in all focal areas, except multifocal, were also allocated for projects rated satisfactory in their outcomes (see table 3.4).

**Table 3.4: Project Outcomes by Focal Area**  
(GEF investment in million \$)

Rating	Bio-diversity	Climate Change	Internat'l Waters	Multi-focal	OD
HS	0.8	0.0	0.0	0.0	0.0
S	71.1	44.1	26.5	0.0	35.0
MS	33.6	32.9	0.0	1.0	0.0
Subtotal	105.5	76.9	26.5	1.0	35.0
MU	8.7	0.0	0.0	0.0	0.0
U	0.0	0.0	0.0	1.7	0.0
HU	0.0	0.0	0.0	0.0	0.0
Subtotal	8.7	0.0	0.0	1.7	0.0
UA	0.0	1.7	0.0	1.3	0.0
<b>Total</b>	<b>114.2</b>	<b>78.6</b>	<b>26.5</b>	<b>4.0</b>	<b>35.0</b>

Projects with highly satisfactory outcomes successfully achieved the project objectives and expected outcomes. For example, in Tanzania's Jozani-Chwaka Bay Biodiversity Conservation project (UNDP), strong government and community ownership was a key factor in the project's success. The project contributed to the legal establishment of the national park and effectively addressed the key threats to biodiversity conservation identified in the project document, such as resource conflicts with communities. According to the terminal evaluation report, Jozani communities are now, as a result of the project, involved in the management and conservation of resources, and 2,500 villagers benefited from savings and credit (microfinance) schemes to develop tourism-related income-generation activities. Census data show that the project interventions have contained and reversed the population decline of the red colobus monkey and reduced encroachment into the Jozani Forest.

In Tunisia's Solar Water Heating project (World Bank), adaptive management and flexibility during implementation was pivotal to success. The project properly responded to a change in the market conditions by refocusing the intervention to the needs of the residential sector as opposed to the commercial and public sectors, as had been initially planned. This change was made because there was limited demand from the latter sectors and an unexpectedly high demand from the household sector. The modified project remained clearly aimed at the grant's fundamental purpose: to encourage the substitution of fossil fuels with renewable solar energy. Project funds were fully utilized for this purpose, and there is evidence that the solar water heating units financed are operating as planned.

Projects with unsatisfactory outcomes did not achieve their objectives for several reasons; for

example, the proposed solutions did not address the underlying problems, adaptive management or solutions were not applied until late in the project's implementation, or implementation management and resource use were poor. This last was true for two UNEP projects, the regional initiative Emergency Response Measures to Combat Fires in Indonesia and to Prevent Regional Haze in South East Asia and the global Barriers and Best Practices in Integrated Management of Mountain Ecosystems. The former had some project design weaknesses, because it focused mostly on strengthening fire-fighting capacity with little recognition at the onset that fire-related issues involve more than shifts in agriculture and weather, but also entail the operations of commercial enterprises, land use changes, and climate issues. The latter was overly ambitious in scope of activities given the resources available, and thus fell short of moving from research and awareness to actionable recommendations as many proposed activities were not accomplished.

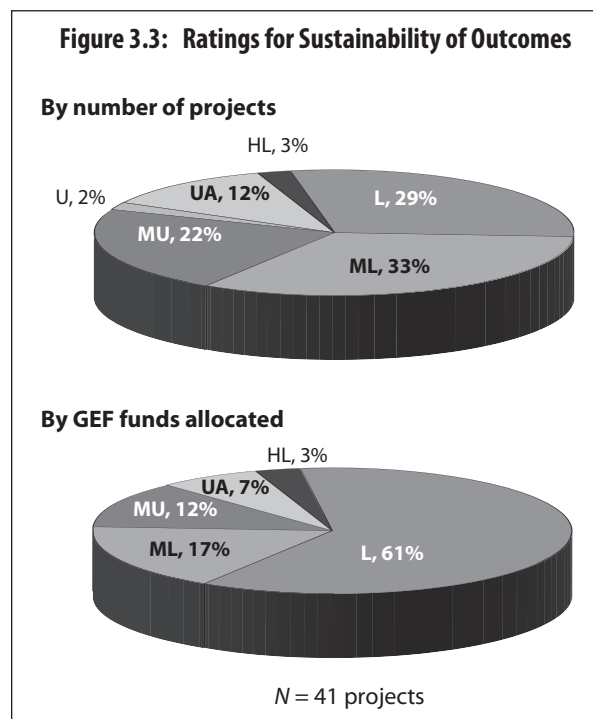
### 3.3 Sustainability of Project Outcomes

The 2005 review of terminal evaluation reports in all focal areas included rating sustainability based on an assessment of key contributing aspects and any risks that could undermine the continuation of benefits at the time of the terminal evaluation.<sup>1</sup> Four aspects of sustainability were addressed: financial, sociopolitical, institutional frameworks and governance, and replication.<sup>2</sup> Annex A provides a breakdown of the questions used to assess

<sup>1</sup>As used here, sustainability is the likelihood of continuation of project benefits after completion of project implementation ("GEF Project Cycle," GEF/C.16/Inf.7).

<sup>2</sup>Replication refers to the repeatability of the project under quite similar contexts based on the lessons and experience gained. Actions to foster replication include dissemination of results, seminars, training workshops, and field visits to project sites ("GEF Project Cycle," GEF/C.16/Inf.7).

each of the aspects of sustainability. The sustainability of outcomes by number of projects and GEF funds is shown in figure 3.3.



Although projects' overall performance in terms of sustainability is not as good as that for outcomes, at least 64 percent of the projects, and at least 81 percent of the GEF funds allocated, resulted in outcomes with a sustainability rating of moderately likely or better.

Most World Bank projects had a likely rating in sustainability (see table 3.5). Two-thirds of the UNDP projects had a moderately likely or better rating in sustainability of outcomes; the remaining third had a moderately unlikely rating. UNEP had an equal number of projects rated moderately likely or better and moderately unlikely or worse. The terminal evaluation reports for two of the six UNEP projects reviewed did not provide sufficient information on this dimension, and therefore the Evaluation Office was unable to rate the sustainability of their outcomes.



**Table 3.5: Outcomes Sustainability by IA  
(number of projects)**

Rating	World Bank	UNDP	UNEP	Total
HL	0	1	0	1
L	8	3	1	12
ML	2	10	1	13
Subtotal	10	14	2	26
MU	1	7	1	9
U	0	0	1	1
HU	0	0	0	0
Subtotal	1	7	2	10
UA	1	2	2	5
<b>Total</b>	<b>12</b>	<b>23</b>	<b>6</b>	<b>41</b>

Among focal areas, 62 percent of biodiversity projects were rated moderately likely or better in the sustainability of their project outcomes (see table 3.6). The outcome sustainability of most projects from other focal areas, except for multifocal projects, was also rated as moderately likely or better.

**Table 3.6: Outcomes Sustainability by Focal Area  
(number of projects)**

Rating	Bio-diversity	Climate Change	Internat'l Waters	Multi-focal	OD
HL	0	0	1	0	0
L	6	3	2	0	1
ML	7	4	1	1	0
Subtotal	13	7	4	1	1
MU	6	1	1	1	0
U	0	0	0	1	0
HU	0	0	0	0	0
Subtotal	6	1	1	2	0
UA	2	2	0	1	0
<b>Total</b>	<b>21</b>	<b>10</b>	<b>5</b>	<b>4</b>	<b>1</b>

Since the present analysis is based only on the 41 terminal evaluation reports reviewed in FY 2005, the resulting picture may not be as representative as desirable. As the number of reviewed projects increases in future years, a more representative assessment of sustainability will be possible.

The proportions of GEF funds assigned to projects with sustainability ratings of moderately likely or above seem to follow the distribution of number of projects. Tables 3.7 and 3.8 show the allocations and ratings by IA and focal area.

**Table 3.7: Outcomes Sustainability by IA  
(GEF investment in million \$)**

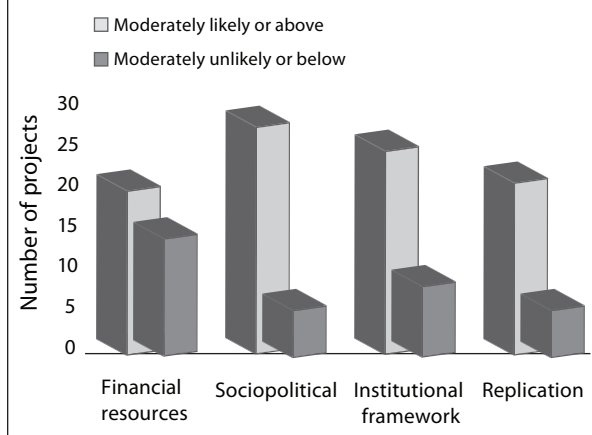
Rating	World Bank	UNDP	UNEP	Total
HL	0.0	7.4	0.0	7.4
L	140.4	14.4	0.8	155.5
ML	7.1	36.1	0.6	43.8
Subtotal	147.5	57.9	1.4	206.7
MU	4.0	27.4	0.8	32.1
U	0.0	0.0	0.9	0.9
HU	0.0	0.0	0.0	0.0
Subtotal	4.0	27.4	1.7	33.1
UA	8.7	3.5	6.3	18.5
<b>Total</b>	<b>160.2</b>	<b>88.8</b>	<b>9.4</b>	<b>258.3</b>

**Table 3.8: Outcomes Sustainability by Focal Area  
(GEF investment in million \$)**

Rating	Bio-diversity	Climate Change	Internat'l Waters	Multi-focal	OD
HL	0.0	0.0	7.4	0.0	0.0
L	55.7	60.7	4.1	0.0	35.0
ML	26.2	13.7	3.0	1.0	0.0
Subtotal	81.9	74.3	14.5	1.0	35.0
MU	18.6	0.8	12.0	0.8	0.0
U	0.0	0.0	0.0	0.9	0.0
HU	0.0	0.0	0.0	0.0	0.0
Subtotal	18.6	0.8	12.0	1.7	0.0
UA	13.7	3.5	0.0	1.3	0.0
<b>Total</b>	<b>114.2</b>	<b>78.6</b>	<b>26.5</b>	<b>4.0</b>	<b>35.0</b>

A closer examination of the factors that determine sustainability of project outcomes reveals that financial viability seems to be an area in which many projects are lagging (see figure 3.4). Financial sustainability refers to financial and economic resources available to allow for the project outcomes to be sustained once the GEF assistance

**Figure 3.4: Sustainability of Project Outcomes, by Assessment Criteria**



ends and the risk that these resources will not materialize.

Two factors were found to help increase financial sustainability: the implementation of regulations that have contributed to an increased demand for services promoted by the GEF project and enhanced private sector involvement. Some projects with likely or highly likely sustainability made use of these mechanisms. For example, in UNDP's regional project Transfer of Environmentally Sound Technologies to Reduce Transboundary Pollution in the Danube River Basin, private industry is increasing its demand for environmentally sustainable technologies and for formally recognized and accredited cleaner production techniques as the countries in which they are based move to meet more stringent European Union environmental standards. Another example of a project assessed as having good prospects for sustainability which have been increased by a strong interest on the part of the private sector

is UNDP's global initiative Removal of Barriers to the Effective Implementation of Ballast Water Control and Management Measures in Developing Countries. The World Bank's Ozone Depleting Substances Consumption Phase-out Project in Russia created a competitive capacity to attract debt and equity investment to comply with the regulatory framework for the proactive management of ozone-depleting substances consistent with international practice. An example from the biodiversity focal area is UNDP's Jozani-Chwaka Bay Biodiversity Conservation project in Tanzania, which has put in place a tourism revenue-sharing scheme that partially covers the management costs of the Jozani National Park. It also provides funds to communities for small development projects in water, sanitation, and so forth. Zanzibar is a major destination for international tourists in East Africa, and the industry has been developing rapidly. Assuming the government of Zanzibar manages tourism development appropriately, it is likely that there will be a sustainable flow of resources to cover some managerial and community costs.

Projects with moderately unlikely or worse sustainability ratings include UNEP's regional Emergency Response Measures to Combat Fires in Indonesia and to Prevent Regional Haze in South East Asia, for which many of the project follow-up activities require further funding—which may not be forthcoming; moreover, cooperation at the national and international levels has been weak. The global UNEP initiative Barriers and Best Practices in Integrated Management of Mountain Ecosystems also lacked financial and institutional sustainability.

### 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, two topics have been addressed in this chapter. The first topic is cofinancing, which here encompasses both materialization of project cofinancing and links between project cofinancing and project outcomes and sustainability. The analysis concludes that, for the most part, the GEF as a whole tends to achieve the cofinancing promised at project inception. The analysis of the links between the levels of cofinancing and outcomes and sustainability ratings was inconclusive. As the number of projects increases in the future, the GEF Evaluation Office will be able to scrutinize more evidence on this subject. The underlying issue to be addressed through such scrutiny is whether there is a ratio of cofinancing beyond which the risk of the GEF's losing leverage increases the chance of compromising the achievement of GEF objectives.

The second topic pertains to the time lag between expected and actual project closing dates and the implications for project outcomes and sustainability. The analysis indicates that, for the projects examined, outcome and sustainability ratings tend to decrease after a delay of more than two years between the expected and actual closing dates. Of the 41 projects included in this analysis, 18 (44 percent) had project completion delays of more than two years. This finding sug-

gests that closer attention to factors affecting project implementation delays during mid-term reviews might permit early detection and correction of factors affecting outcomes and their sustainability.

#### 4.1 Materialization of Cofinancing

Tracking the materialization of promised cofinancing is important because project activities are budgeted with the expectation of such materialization. For this analysis, all 116 terminal evaluation reports reviewed since 2001 were examined. However, the IAs provided actual cofinancing data only for 70 of these projects (60 percent). This limits the extent to which inferences regarding the potential factors affecting materialization of cofinancing could be made. As the quality of terminal evaluation reports improve and they disclose actual project costs, a fuller and more representative picture can be presented.

#### Total Promised Cofinancing and Actual Cofinancing

An analysis of the data provided by the 70 terminal evaluation reports that included cofinancing information indicates that promised cofinancing tended to materialize, for the most part, for these projects (see table 4.1).

**Table 4.1: Actual Compared to Promised Cofinancing**

Cofinancing	Total (million \$)	Average Cofinancing for Every GEF \$ Approved
Promised	1,900	4.4
Realized	1,770	4.1

Note: Total cofinancing promised and realized is the sum of these values for the 70 out of 116 projects reviewed from 2002 to 2005 that provided actual information on this in their terminal evaluation reports. These two values were then divided by the total GEF funds approved for these 70 projects to calculate the average funds promised for every GEF dollar approved.

### Promised and Actual Cofinancing by IA, Focal Area, and Region

A more in-depth analysis by IA, focal area, and geographical region provides some information on differences which can improve planning. The South Asia and East Asia and Pacific regions appear to have the highest ratios of promised cofinancing per GEF dollars approved (see table 4.2). In the case of East Asia and Pacific, this was driven by three very large blended climate change projects conducted by the World Bank in China and Indonesia. In South Asia, the main drivers were climate change projects in India and

Sri Lanka and a biodiversity project in India; the World Bank was the IA for all three projects. In most of the other regions, promised cofinancing was more or less equal to total GEF funding.

Interestingly, while the promised cofinancing for projects in Latin America and the Caribbean seems to be equal to the amount of GEF funds allocated to same, the region has the highest levels of actual cofinancing, with 141 percent of promised cofinancing actually materializing. This high level of cofinancing was driven mainly by some large conservation projects in Mexico, Brazil, and Bolivia, but throughout the region, most projects attracted more cofinancing than was expected at project approval. On the other hand, global projects appear to have the lowest ratios of both promised cofinancing/GEF funding and actual/promised cofinancing.

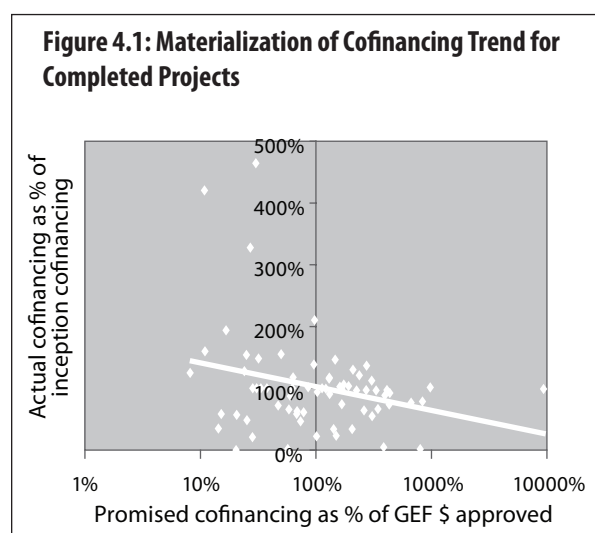
Among IAs, UNEP projects that provided information on actual costs had the lowest cofinancing ratio and also the lowest percentage of cofinancing actually materializing (see table 4.2). Of the seven

**Table 4.2: Cofinancing Ratios and Actual Cofinancing by Region, IA, and Focal Area**

Region, IA, Focal Area		Number of Terminal Evaluation Reports Reviewed	Number of Reports Providing Actual Cost Information	Promised \$ of Cofinancing per GEF \$ Approved	Promised Cofinancing That Materialized (%)
Region	Africa	30	17	2.5	76
	East Asia and Pacific	17	14	6.8	91
	Europe and Central Asia	19	9	1.0	93
	Global	11	4	0.8	66
	Latin America and the Caribbean	24	17	1.0	141
	Middle East and North Africa	9	5	1.4	98
	South Asia	6	4	5.8	98
IA	World Bank	50	39	5.0	93
	UNDP	50	25	2.5	83
	UNEP	19	7	1.8	48
Focal Area	Biodiversity	61	41	1.5	89
	Climate change	29	17	10.0	94
	International waters	16	7	1.8	94

UNEP projects providing actual cost information, three were global projects and two were in Africa, the two regions with the lowest average cofinancing actually materializing.

Among the projects reviewed, those with larger promised cofinancing as a percentage of GEF funds tended to meet their expected cofinancing goals better than did projects with smaller promised cofinancing as a percentage of GEF funds (see figure 4.1). This relationship holds true regardless of project size (full- or medium-sized projects), and is true for projects implemented by the World Bank and UNDP as well as for those in the biodiversity and climate change focal areas. The UNEP and international waters samples were not sufficiently large to permit conclusions.



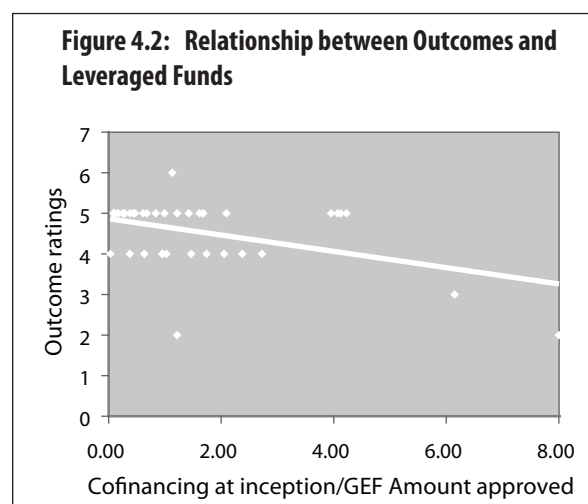
## 4.2 Relationship between Project Funding and Project Outcomes and Sustainability

The Evaluation Office examined the relationship between project funding and project outcomes and sustainability to assess whether projects with larger GEF funding and cofinancing produced better results. This analysis is based on the 41 projects whose terminal evaluation reports were reviewed

in 2005. A linear regression was carried out with GEF funds as the independent variable and the project outcomes and sustainability ratings as the dependent variable. A second linear regression was conducted with the ratio of promised cofinancing and GEF funds (that is, the cofinancing ratio) as the independent variable and the project outcomes and sustainability ratings as the dependent variable.

The analysis shows little correlation between GEF funding and outcomes and GEF funding and sustainability for the terminal evaluation reports reviewed in 2005. In other words, these projects seemed to receive sufficient GEF funds to carry out their promised activities, with some achieving their objectives more successfully than others regardless of the amount of GEF funding received.

Regarding promised cofinancing, the analysis showed that the more leveraged the project was (that is, the higher the ratio of promised cofinancing to GEF funds), the lower the outcome ratings tended to be (see figure 4.2). This correlation is driven by two outliers that have the highest cofinancing ratios and have outcome ratings of moderately unsatisfactory or worse. When these two outliers are eliminated, there is no correlation between cofinancing ratio and outcomes. The two outliers are Ghana's Natural



Resources Management project (World Bank) and the regional Emergency Response Measures to Combat Fires in Indonesia and to Prevent Regional Haze in South East Asia project (UNEP).

Since outliers may be skewing the correlation, it is necessary to examine more projects in the future. As more data (ratings) became available in the future, the Evaluation Office will be able to better assess the relationship between promised cofinancing and outcomes, and promised cofinancing and sustainability. It will also be able to determine whether the GEF should more closely monitor projects that have very high cofinancing ratios.

### 4.3 Delays in Project Completion

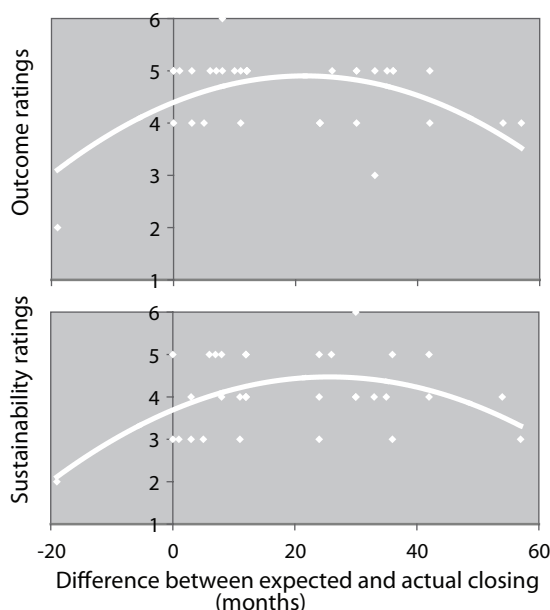
The Evaluation Office measured the time between project effectiveness and expected closing and compared it with the time between project effectiveness and actual closing to assess the implications of this delay in terms of project outcomes and sustainability.

The analysis found that projects that reached completion closed, on average, 20 months after the expected completion date: 18 (44 percent) projects closed with a delay of two years or more. The analysis showed that after two years, the quality of outcomes and their sustainability tends to decrease.<sup>1</sup> (See figure 4.3.<sup>2</sup>) The Evaluation Office

<sup>1</sup>This decline is correlated with the square of the time delay and is significant at the 95 percent confidence level (using a binomial regression).

<sup>2</sup>For figure 4.3, the outcomes and sustainability rating system was converted to a numeric scale of 1–6, with 1 being

**Figure 4.3: Effect of Delay on Project Outcomes and Sustainability**



will continue to track this relationship in future APRs, but these findings suggest that if, during mid-term reviews, the factors producing project implementation delays are detected and corrected, project outcomes and their sustainability might be improved.

highly unsatisfactory/highly unlikely and 6 being highly satisfactory/highly likely. A delay of 0 means that the project closed when expected.



# MONITORING AND EVALUATION

## 5. Assessment of Project-at-Risk Systems of GEF Partner Agencies

The assessment of the project-at-risk systems of the GEF partner agencies examines system design as reported by the respective agencies to the Evaluation Office. This assessment did not include an examination of actual internal reports to determine the degree of compliance with formal procedure. The review also identifies future oversight parameters for agency risk monitoring systems.

Many partner agencies tend to follow the World Bank approach of monitoring projects-at-risk based on a warning flag system that tracks self-rated project performance through a corporate management information system. These ratings are then aggregated and rolled up for portfolio-level reporting. The risk monitoring systems of the GEF partner agencies are evolving rapidly. Several agencies have recently installed new MISs which have a separate risk module. However, several issues still need to be addressed:

- Insufficient frequency of observations undermine the reporting power inherent in an MIS (computer reports can be generated any time, yet the underlying data are often updated only once or twice per year).
- The robustness and candor of self-assessment may be doubtful.
- Managers and staff worry about proliferation of monitoring and reporting systems,

overlap and redundancy, and staff reporting burdens.

- Only EBRD has a formal process of project-level risk validation independent of the business unit. At the World Bank, the Quality Assurance Group performs a similar function, albeit at a more aggregate level.

### 5.1 Assessment Approach

This assessment was based on information provided by GEF partner agencies and in interviews conducted with IA/EA representatives. The assessment looked at the ways in which systems are designed. It did not examine how systems are being implemented; assessment of system implementation would require auditing of actual internal reports to determine the degree of compliance with formal procedure. This would be an important element to be included in any future GEF assessment of the quality of project supervision.

Agencies reviewed include the three IAs and seven presently approved EAs, as shown in box 5.1. Five of the agencies are specialized United Nations organizations that carry out projects and provide technical assistance to developing countries; and five are development banks that provide loan finance to member countries, with GEF project financing handled in the form of grants.

### **Box 5.1: GEF Partner Agencies**

#### ***Implementing Agencies***

- UNDP
- UNEP
- World Bank

#### ***Executing Agencies***

- African Development Bank
- Asian Development Bank
- European Bank for Reconstruction and Development
- Food and Agriculture Organization
- Inter-American Development Bank
- International Fund for Agricultural Development
- UN Industrial Development Organization

## **5.2 GEF Requirements for Project Preparation and Implementation**

### **Preparation Phase**

During the project preparation process, there are several opportunities to raise risk issues and make possible clarifications or improvements before project approval and implementation. These include the internal review process of the IA or EA, review by the GEF Secretariat, the STAP roster review, the GEF Council review of work programs, and CEO endorsement. The project summary for GEF projects includes a heading for “Key Indicators, Assumptions and Risks,” and the guidance instructions indicate that the material presented under that heading is to be taken from the logframe (or results framework), which includes a column identifying project risks. The risk-related material presented in the logframe and project narrative are presumed to represent the final outcome of any issues raised during the review process, but, with one exception, there is no specific requirement that risks be discussed in depth or tracked separately during this process. The single exception is found in the “Review Criteria for GEF Full-Sized Projects,” which stipulates that, at work program inclusion, the project document must

describe the project’s logical framework, “including...risks and assumptions”; at CEO endorsement, the final project description should include “details of project activities, inputs, and related risks and assumptions.”<sup>1</sup>

The GEF Secretariat Proposal Agreement Review template, which tracks project preparation issues and specifies how, when, and by whom project-at-risk issues are to be addressed, does not include a category specific to risk.<sup>2</sup> Such information could, however, be presumed to fall under the project design heading, as it is at that point that risk-related issues should be addressed in the GEF project document. Since at pipeline entry the review process focuses on incremental reasoning, any potential shortcomings in the area of risk assessment are expected to be taken up at work program inclusion or, at the latest, by the time of CEO endorsement (final project approval).

### **Implementation Phase**

The annual project implementation review (PIR) is the standard reporting document for GEF-financed projects during their implementation phase. In most cases, the PIR represents a recapitulation of data produced by an IA/EA’s internal monitoring system, rather than a separate assessment carried out for the GEF. From this standpoint, the information available to the GEF on project and portfolio risk largely results from the aggregate strengths and weaknesses of the diverse approaches and practices of partner agencies (which also undergo frequent change, as discussed later). It is also useful to note that there are variations in the PIR form as used by different agen-

<sup>1</sup>Found at the GEF website, under Operational Policies/Templates and Guidelines/annex H ([www.thegef.org/Operational\\_Policies/Eligibility\\_Criteria/templates.html](http://www.thegef.org/Operational_Policies/Eligibility_Criteria/templates.html)), 2004, p. 3.

<sup>2</sup>Found at the GEF website, under Operational Policies/Templates and Guidelines/annex F-1 ([www.thegef.org/Operational\\_Policies/Eligibility\\_Criteria/templates.html](http://www.thegef.org/Operational_Policies/Eligibility_Criteria/templates.html)), 2003.



cies, and the form itself has evolved over time. For example, an older PIR examined for this analysis did not provide a specific field for addressing risk issues, although there was an optional descriptive assessment box in which risk issues could be presented in a narrative. In contrast, later review forms included a section on risks, which discusses risk issues; classifies them as high, substantial, modest, or low; and includes a box to report “actions taken or planned to manage High and Substantial risks.”

From another standpoint, it can be considered that risk is treated implicitly in the PIR forms. For example, poor performance by a project represents a *risk* of the project failing to achieve its objectives. An assessment of project progress toward its objectives also captures the concept of risk. The GEF review and reporting criteria for the preparation and implementation phases make frequent reference to the concepts of sustainability and replicability, and note that these concepts are inherently associated with the concept of risk. These two aspects are specifically addressed during terminal evaluations of all GEF projects; and, therefore, any disconnect between project outcomes, sustainability, and replicability can be monitored as part of the APR process.

### 5.3 Agency Reporting Systems

Most agencies’ reporting system covers two main components: financial flows, which are aggregated for internal management and accountability purposes; and some form of project performance information, which may or may not be aggregated at the portfolio level. Information technology in the form of management information systems has become the norm among IAs and EAs for agency-level monitoring of financial flows and is increasingly applied for more operational, project-level purposes as well. As MIS instruments are being

applied to project-level monitoring, much attention has been given to tools for capturing and reporting project performance in order to provide a clearer picture of overall trends for a given agency’s management and governing bodies. This typically entails considerable effort to provide meaningful aggregated information and to ensure that underlying data are reliable and consistent across reporting units and over time.

Although project-level monitoring is a long-standing practice, the details of performance monitoring vary widely in terms of frequency, criteria assessed, and so on. In almost all cases, project performance ratings are recorded as part of regular project monitoring by IA/EA staff. The supervision staff assesses the project performance within specified categories in order to maintain a record of project implementation history over time. The ratings may also be reported as numeric scores which simplifies the task of preparing aggregate reports (that is, by focal area, geographic region, or agency). Such scores are typically reported on a scale of 1 to 4, although several agencies have shifted to a six-point rating scale to better capture variation in performance. Frequency of assessment varies, but for most agencies it is typically done once or twice per year. The PIR itself is an annual document, but this does not preclude the possibility that an IA or EA conduct ratings more frequently. In recent years, ratings systems have become increasingly computerized in corporate MISs, with agency management giving more attention to results monitoring and reporting. (In several agencies, this evolution continues to be an ongoing process, and the transition to new MIS software can be difficult and time consuming, with a significant learning curve for agency staff.)

Most project reporting systems are dependent on self-rating of project performance and risks by project officers or IA/EA supervision staff. Imme-

diate line management may have a review role over such ratings, but often this function is done on a spot-check basis because of the time pressure on managers. Some agencies have identified this as a problem and have tried to find ways of strengthening procedures at two levels: (1) to ensure that ratings are entered on a timely basis by supervision staff, and (2) to engage line managers more consistently in oversight of the quality of project reporting. The World Bank improved the timeliness of supervision reporting after several years of effort, reporting only one case of a “stale” project report (in a portfolio of over 1,000 projects) by the end of FY 2005; during the 1990s, many projects had missed one or more reporting dates, and the consistency and quality of information provided often varied widely.

In the World Bank reporting system (other systems are similar in many respects), the task team leader enters judgments about project progress and other implementation issues into the MIS in several broad categories (see box 5.2). Any category rated less than satisfactory is considered a “risk flag.” The MIS then tabulates these ratings as a function of the regular internal reporting system. Depending on the number of warning flags that may be identified for a given project and the type of categories on which warning flags have been issued, a project can then be classified as “non-risky,” a “potential problem,” or an “actual problem.”

Implementation of the risk flag system has provided World Bank management with a useful tool for cross-checking the realism of task team ratings of overall project performance by looking for discrepancies between team ratings and the risk flags being tracked by the MIS. (Boxes 5.3 and 5.4 describe risk monitoring systems being implemented by and within partner agencies that emphasize independent risk monitoring.) In addition, the effectiveness or proactivity of task teams

#### **Box 5.2: World Bank Risk Rating Categories**

##### ***Performance self-rating on:***

- Implementation performance
- Likelihood of achieving the development objective
- Likelihood of achieving the global development objective (for GEF projects)
- Project management
- Financial management
- Counterpart funding
- Procurement
- Monitoring and evaluation

##### ***Corporate MIS-generated flags:***

- Project effectiveness delays
- Disbursement delays
- Country performance issues

in resolving project issues is monitored, with the MIS calculating the lag between the time when a project is identified as being at risk until the issue(s) has been resolved and the performance rating upgraded.

The World Bank currently considers one year as the maximum acceptable period for a project to remain in at-risk status; if this deadline is passed, the task team (and by inference, line management) is deemed insufficiently proactive. Corporate accountability reports (prepared by the Quality Assurance Group, which operates independently of the operational units) present regular calculations of the realism of ratings and the proactivity of supervision actions by regions and sectors. In a separate project risk rating—which existed until 2005 and was then dropped—risks could be rated as negligible, moderate, substantial, or high. At present, the task team’s judgments about project risks are implicitly reflected in the overall rating for development objective, while the MIS calculations of risk flags are used by management for monitoring at the country, regional, and portfolio-wide levels.

### **Box 5.3: Enhancing the Independence of Internal Risk Monitoring**

The monitoring approach used by the European Bank for Reconstruction and Development differs significantly from that used by the other partner agencies. The EBRD places risk analysis within a special risk management vice presidency, which reports to senior management and the Board of Directors to ensure that risk information being passed to senior management is as free as possible from potential conflicts of interest in the risk management process. This model is derived from recent developments in the commercial banking sector, arising from concerns related to fiduciary risk (such as the Sarbanes-Oxley legislation in the United States), and emphasizes transparency and accountability. In EBRD, project implementation issues are considered together with financial risks as part of an overall risk management strategy for the agency. Thus, the risk profile and complexity of a given project determine the supervision budget and schedule, rather than a standard formula for the institution. EBRD management considers all GEF projects to be complex operations regardless of scale or focal area, and treats them as inherently risky and in need of special supervision and management attention.

Other agencies also have explicit risk monitoring systems. At the time this report was being prepared, the Inter-American Development Bank, for example, was introducing a Project Alert Identification System. UNDP's new ATLAS MIS also has a separate risk module which is now being implemented. UNEP has a risk factors table which summarizes project risks identified in the project document, as well as any new ones identified during implementation. In this table, each risk is classified as low, medium, high, not applicable, or to be determined; for high risks, a separate worksheet is completed, indicating what management actions are being taken to mitigate risk. These approaches generally parallel the recent PIR approach, but there is no information on the comparability of risk judgments across agencies.

Other elements of risk also enter into the picture, notably in the area of fiduciary or safeguard risk (that is, noncompliance with agency policies or misuse of resources). All of the GEF partner agencies have a set of fiduciary policies in place in addition to those required by GEF; these are legally binding on the grantee and are supervised as a matter of course. At the World Bank, certain high-profile projects may also be classified as "corporate risk projects" during preparation, based on the potential for safeguard noncompliance and the magnitude of potential impacts (involuntary resettlement, indigenous people, safety of large dams, and so on). This is the main exception to the self-rating of risks within the World Bank; oth-

### **Box 5.4: Independence through External Risk Monitoring**

In response to the Sarbanes-Oxley legislation, private sector firms are increasingly relying on business reviews conducted by the Committee of Sponsoring Organizations (COSO) of the Treadway Commission. Some business units of the World Bank have also used COSO reviews to assess the quality of their business practice with respect to internal controls, business ethics, and corporate governance. This initiative represents another approach to tackling the institutional culture factors which are often cited as impediments to improving the candor and realism of internal reporting. As applied in the private sector, these initiatives emphasize the need for strengthened internal controls, independent validation of business units' self-assessments, and progress tracking in resolving problems. These functions can be within a corporate-wide entity, or within a business unit, but the key is autonomy of judgment and a flow of reliable information to senior management.<sup>1</sup> These approaches are quite new, however, and their actual effectiveness in changing day-to-day institutional behavior has yet to be seen.

<sup>1</sup>Financial Executives Research Foundation, "What Is COSO? Defining the Alliance That Defined Internal Control," [www.feif.org](http://www.feif.org) (2003). Also see Committee of Sponsoring Organizations of the Treadway Commission, *Enterprise Risk Management—Integrated Framework*, [www.coso.org](http://www.coso.org) (2004).

erwise, risk is understood to be managed within the business unit.

In summary, projects could be at risk due to:

- poor performance in implementation
- noncompliance with fiduciary policies
- identification of high or substantial risk factors, perhaps outside the immediate project

## 5.4 Ensuring Comparability

From the standpoint of trying to monitor the GEF's overall portfolio risk, there are basic problems of ensuring comparability, as different agencies may adopt somewhat different definitions and approaches in how they operationalize risk management. For example, the development banks, which represent half of the 10 partner agencies covered in this assessment, are lending entities with governance, supervisory, and legal instruments that differ in many ways from those of the specialized UN agencies that are the remaining GEF partners. The development banks operate in capital markets and are already giving substantial management attention to factors such as credit and market risk. In some of the development banks, these concepts are now being expanded to cover a wider range of operational risks, as has been happening for several years in the commercial banking sector.<sup>3</sup> Indeed, the concept of projects-at-risk (which is already used by the World Bank and is being introduced by the other development banks) is derived from the banking sector where nonperforming loans are a central concern of the enterprise. This important ongoing trend in the banking sector is also likely to become

increasingly relevant for the development banks (and perhaps for the other development agencies as well):

In the past, banks relied almost exclusively upon internal control mechanisms within business lines, supplemented by the audit function, to manage operational risk. While these remain important, recently there has been an emergence of *specific structures and processes aimed at managing operational risk* (Basel Committee on Banking Supervision, 2003; emphasis added).

For this assessment, an inventory card was developed which lists 17 characteristics of a project risk monitoring system (see annex B). For each of the 17 aspects, a yes response indicates presence, and a no response indicates absence, of that specific desirable element in the risk monitoring system. Five of the 17 elements are considered especially critical. The inventory card was filled out by each of the 10 GEF IAs and EAs; table 5.1 summarizes these results.

Wide variations in score, ranging from 10 to 17 yes responses, are apparent. There is a similar range in response for the five critical risk monitoring elements, with agency scores ranging from 3 to 5. On the whole, the scores are considerably higher for the development banks, which is consistent with the analysis presented in the earlier sections. Annex C presents the risk monitoring inventory disaggregated by agency.

The data generated through self-reporting is known to be affected by issues related to candor, underreporting of risks, and overestimation of performance. Recent World Bank assessments show that there is disconnect between the ratings given by task team and those given by independent reviewers such as the Quality Assurance Group in 35 to 40 percent of instances. This may be attributed to a staff perception that "problem

<sup>3</sup>Basel Committee on Banking Supervision, *Sound Practices for the Management and Supervision of Operational Risk. Bank for International Settlements* (Basel, Switzerland, 2003); and Basel Committee on Banking Supervision, *Operational Risk Management* (Basel, Switzerland, 1998).

**Table 5.1: Agency Risk-Monitoring Inventory**

Agency	Total Yes Responses	Critical Elements with Yes Responses
Asia Development Bank	17	5
African Development Bank	15	4
EBRD	17	5
Food and Agriculture Organization	12	4
Iner-American Development Bank	14	4
International Fund for Agricultural Development	12	3
UNDP	16	4
UNEP	15	4
UN Industrial Development Organization	10	3
World Bank	14	4

projects” are to be avoided, or that “managers don’t want to hear the bad news,” and observers frequently mention an incentive system that tends to downplay risks and to display a positive attitude (until problems become severe). Some institutional “churning” may also be seen in the form of frequent changes in the number or definition of rating categories, rating scales, numbers of areas being monitored by the MIS, and so on.<sup>4</sup>

These risk assessment systems are sometimes seen by the operational staff as bureaucratic paperwork, with little utility for projects and draining

scarce staff time. The task teams often complain of being unable to keep up with changes in reporting requirements and criteria.

Agencies have tried various ways to streamline these procedures, by attempting to focus reporting on key issues, reducing the amount of narrative text (which in any case presents major difficulties for portfolio roll-up purposes), and standardizing data entry fields wherever possible. However, since the need for institutional accountability is the main driver of these systems, the general trend continues toward more—rather than less—reporting, and an increasing level of attention is being given by line and agency management to the accuracy of information being reported to boards and donors.

<sup>4</sup>World Bank, *FY05 Annual Report on Portfolio Performance* (Washington, DC: 2006).



# MONITORING AND EVALUATION

## 6. Quality of Project Monitoring

---

This chapter addresses the quality of project monitoring of GEF projects. The chapter's first section continues the analysis initiated in 2004. The second section reviews the systems for quality control of M&E arrangements at the point of CEO endorsement. This material presents, for the first time, a snapshot of the functioning of the process by which project M&E is reviewed at CEO endorsement in the GEF.

There was an improvement in the quality of M&E systems for projects with terminal evaluation reports submitted to the GEF Evaluation Office in FY 2005 compared to those submitted in FY 2004. While it is premature to interpret this change as a trend, it is likely that these improvements will persist if Implementing Agencies continue to enhance their project monitoring systems. Nevertheless, for improvements to continue, it is critical that more attention be given to M&E plans during project design. The review in section 6.2 reports that 58 percent of the projects endorsed by the GEF CEO met the GEF M&E requirements at project entry. The review found that there is considerable room for improvement during project preparation, and that there are major gaps and weaknesses in the present review process. Developing and providing better guidance for the review process will lead to a more uniform understanding of M&E expectations among project reviewers. Similarly, a better and

more uniform understanding of M&E expectations will facilitate the development of good M&E plans and the enforcement of M&E standards to ensure higher compliance. In addition, investments to intensify the development of better tools and indicators will improve the measurement of GEF project results.

### 6.1 Quality of Monitoring during Project Implementation

This analysis uses 83 terminal evaluation reports—42 from FY 2004 and 41 from 2005—submitted by the IAs to the Evaluation Office. The number of projects rated as moderately satisfactory or above in terms of their monitoring during implementation increased from 39 percent in 2004 to 52 percent in 2005. This improvement is attributed to IAs' actions undertaken to address issues raised in the 2004 APR and ongoing changes by the IAs to advance the quality of monitoring. Although this improvement is important, it is not conclusive because a sizable proportion (20 percent) of the terminal evaluation reports submitted in 2005 did not provide sufficient information to assess quality of project monitoring.

#### Approach

The Evaluation Office rates the quality of project monitoring using the following criteria:



- Was an appropriate M&E system for the project put in place (including capacity and resources to implement it), and did this allow for the tracking of progress toward project objectives? The tools used might include a baseline, clear and practical indicators, and data analysis systems, or studies to assess results planned and carried out at specific points in the project.
- Was the monitoring system used effectively for project management?

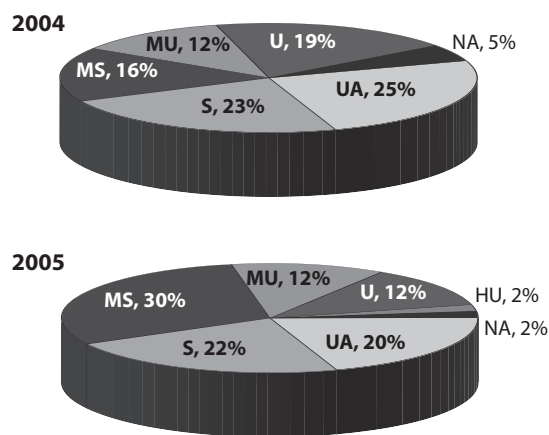
The Evaluation Office provided an overall rating on the quality of project monitoring based on individual ratings on each of these questions; each question was given an equal weight.

## Overall Findings

The Evaluation Office began rating the quality of project monitoring in 2004, which allows for comparison with the projects of 2005.<sup>1</sup> As indicated in figure 6.1, the proportion of moderately satisfactory or better projects increased from 39 percent in 2004 to 52 percent in 2005; this was largely accounted for by a reduction in the percentage of projects rated unsatisfactory (from 19 to 12 percent) and an increase in the percentage of projects rated moderately satisfactory (from 16 to 30 percent). While it is too soon to interpret this change as a trend, ongoing IA efforts to advance the quality of M&E systems (see box 6.1) are likely to be an important factor in these improvements. Thus, it could be expected that this improvement will continue in the future.

<sup>1</sup>The Evaluation Office rated quality of project M&E systems using a six-point rating scale of highly satisfactory (HS), satisfactory (S), moderately satisfactory (MS), moderately unsatisfactory (MU), unsatisfactory (U), and highly unsatisfactory (HU). Projects were rated “unable to assess” (UA) when the information contained in the terminal evaluation report did not allow for assessment. “NA” means the rating is not applicable.

**Figure 6.1: Quality of Project M&E Systems**



The proportion of terminal evaluations providing insufficient information decreased only slightly (from 25 to 20 percent), indicating that there is still significant room for improvement. The APR will continue to examine these trends in future years to assess whether this gradual improvement persists.

## Quality by IA and Focal Area

As shown in table 6.1, UNDP had the largest increase in ratings for quality of project monitoring, the performance of the World Bank was stable, and the numbers for UNEP were too small to permit any significant assessment. Because quality of monitoring has only been tracked for two years, it is still too early to draw any conclusions; however, the number of projects rated below moderately satisfactory or that have provide insufficient information to allow for assessment remains high for all IAs, accounting for 46 percent of 41 projects. Quality of monitoring by the various IAs will continue to be assessed as part of future APRs. As shown in table 6.2, climate change is the only focal area for which the quality of M&E systems worsened between 2004 and 2005.

### Box 6.1: IA Initiatives to Improve the Quality of Monitoring

The IAs have taken several measures to improve the quality of project monitoring of achievement of objectives and expected project outcomes. These have included studies to assess current monitoring practices across their portfolio of GEF projects to determine areas that need improvement. For example, UNDP conducted a study, "Status of Monitoring and Evaluation in UNDP-GEF Projects," to characterize M&E practices within and across its portfolio, based on a sample of 30 GEF projects. The study concluded that of these projects, "not one was found to be exemplary across all of the elements of its M&E system." Further, the "Quality of Indicators and Identification of Sources of Verification were almost universally weak." To address the significant challenges associated with measuring change in global environmental benefits, UNDP is piloting the use of scorecards that help reduce inconsistencies in measurement. UNDP is also looking to improve its use of repeatable independent measurements of baseline, mid-term, and end-of-project conditions.

The World Bank carried out a similar assessment of its GEF project M&E systems two years ago; consequently, its recent actions in this regard focus on training and more diligent review. The World Bank GEF team organized a one-week training workshop for task teams, attended by 30 staff members, on planning and organizing project M&E and focusing on the results framework. It also organized a workshop for GEF regional coordinators and thematic specialists in their role as reviewers of the results framework so they can better provide guidance to task teams. Additionally, the World Bank imposed much stricter criteria for the minimum standards for M&E arrangements in projects. Most World Bank regions have established an M&E help desk to provide support to task teams.

UNEP is emphasizing the identification of key outcome-level SMART indicators in the project design phase, and ensuring that activities to track these indicators are clearly specified in the M&E plan and the corresponding resources are realistically budgeted. In addition, UNEP is compiling baseline information and data, prioritizing that relevant to the outcome-level indicators. Specifically for biodiversity, UNEP is using tracking tools for strategic priority 1 (catalyzing the sustainability of protected area systems) and strategic priority 2 (mainstreaming biodiversity in production landscapes and sectors) and is awaiting further task force results concerning development of tracking tools for the other focal areas.

At the project implementation stage, UNEP's project steering committees are expected to review progress in meeting project objectives, review whether selected indicators are being monitored, and determine whether these indicators are relevant and cost effective. The committees meet on a yearly basis and include recommendations relevant to project M&E systems in their meeting minutes. The terms of reference for mid-term project reviews emphasize results monitoring; outcome indicators are reviewed again at this stage and the tracking tools validated.

**Table 6.1: Quality of Project Monitoring by IA**

Rating	World Bank		UNDP		UNEP	
	2004	2005	2004	2005	2004	2005
HS	0	0	0	0	0	0
S	10	5	0	4	0	0
MS	4	3	1	6	2	3
Subtotal	14	8	1	10	2	3
MU	0	1	4	4	1	0
U	4	0	3	4	1	1
HS	0	0	0	1	0	0
Subtotal	4	1	7	9	2	1
NA	0	0	0	0	2	1
UA	3	3	5	4	3	1
<b>Total</b>	<b>21</b>	<b>12</b>	<b>13</b>	<b>23</b>	<b>9</b>	<b>6</b>

**Table 6.2: Quality of Project Monitoring by Focal Area**

Rating	Biodiversity		Climate change		Internat'l waters		Other	
	2004	2005	2004	2005	2004	2005	2004	2005
HS	0	0	0	0	0	0	0	0
S	6	6	4	1	0	2	0	1
MS	5	9	0	1	1	2	0	0
Subtotal	11	15	4	2	1	4	0	1
MU	4	3	0	2	1	0	0	0
U	5	2	0	0	3	1	0	2
HU	0	1	0	0	0	0	0	0
Subtotal	9	6	0	2	4	1	0	2
NA	0	0	0	0	0	0	2	1
UA	4	0	4	6	2	0	1	1
<b>Total</b>	<b>24</b>	<b>21</b>	<b>8</b>	<b>10</b>	<b>7</b>	<b>5</b>	<b>3</b>	<b>5</b>

## 6.2 Systems to Ensure Quality of M&E Arrangements at Entry

This section reviews the systems for quality control of a project's M&E arrangements at the point of CEO endorsement. The objective is to determine the extent to which these quality control systems are able to ensure that GEF-financed projects meet the M&E requirements established by the GEF Council, identify any shortcomings, and identify and analyze the factors influencing these shortcomings.

This review draws its conclusions from an examination of the 74 full-sized projects that were CEO-endorsed in FY 2005; an examination of the comments provided during the review process by GEF Secretariat reviewers, GEF Council members, and STAP roster reviewers; and interviews carried out with GEF Secretariat and IA staff. Annex D presents a more detailed description of the methodology followed during the review; annex E summarizes the review's results.

To assess the quality of M&E plans at entry, an instrument was developed that measures 13 specific aspects (parameters) of M&E quality; these parameters are based on the GEF Secretariat project review criteria set forth in "GEF Project Cycle" (GEF/C.16/Inf.7) and the guidelines contained in the GEF's 2002 *Monitoring and Evaluation Policies and Procedures*. In some cases, the parameters outlined in these documents were refined to facilitate consistency and objectivity in the application of the assessment instrument.<sup>2</sup> Certain technical or operational elements were not included, as this would have required specialized technical expertise on individual projects and would have

introduced greater subjectivity into the review process.<sup>3</sup>

Since the GEF's inception, the GEF Council has, on numerous occasions, looked to strengthen M&E policies and procedures.<sup>4</sup> To streamline the project review process and in response to Council request, the GEF Secretariat in 2000 developed project review criteria that laid down the GEF requirements—including those for M&E arrangements—at various stages of the project cycle for both full- and medium-sized projects. In 2002, the GEF Secretariat published *Monitoring and Evaluation Policies and Procedures*, wherein it defined the expectations for project quality, again including those for M&E arrangements at entry. This review used the requirements applicable at the point of CEO endorsement to develop criteria to assess whether projects are in compliance with Council expectations for M&E arrangements at entry.

The parameters for assessment have been classified as either "critical"—where noncompliance indicates serious deficiencies in the M&E arrangements—or "other"; see annex D for complete descriptions. To comply with the GEF M&E expectations at entry, a project needs to be in compliance with all critical parameters and needs to perform sufficiently well on all the other parameters together. To be classified as compliant, projects were required to score at least a 2 (on a scale of 1 to 3, with 3 being the highest) on each of the critical parameters and to have an aggregate score of 26 out of a maximum of 39. Since these criteria

<sup>2</sup>The parameters that were refined further include specific and sufficient indicators, specific targets for the chosen indicators, and the specific targets being based on some assessment of the initial conditions.

<sup>3</sup>The parameters that were mentioned in the two documents but not used in the instrument include discussion of key assumptions of the project, sufficiency of M&E budget, and adaptive management.

<sup>4</sup>Decision on Agenda Item 7, Council Meeting November 1–3, 1994; Decision on Agenda Item 15—Other Business, Council Meeting May 5–7, 1999; Discussion on Agenda Item—Monitoring and Evaluation, Council Meeting May 9–11, 2001.

are also consistent with the new *GEF Monitoring and Evaluation Policy* (Evaluation Document 2006, No. 1), the findings of this study will form a baseline for monitoring and assessing the new policy's implementation.

## Project Distribution

The reviewed cohort of 74 projects mirrors the IA and focal area distribution of the GEF portfolio as a whole at the time of its third replenishment (FY 2003–06). By IA, the largest percentage were implemented by the World Bank, followed closely by UNDP; the next largest percentage were joint projects<sup>5</sup> (see figure 6.2). By focal area, biodiversity and climate change, respectively, accounted for the largest shares of projects; the smallest shares were in the persistent organic pollutants (POPs) and ozone depletion focal areas (see figure 6.2). Only projects implemented by the World Bank or UNDP, and those in either the biodiversity or cli-

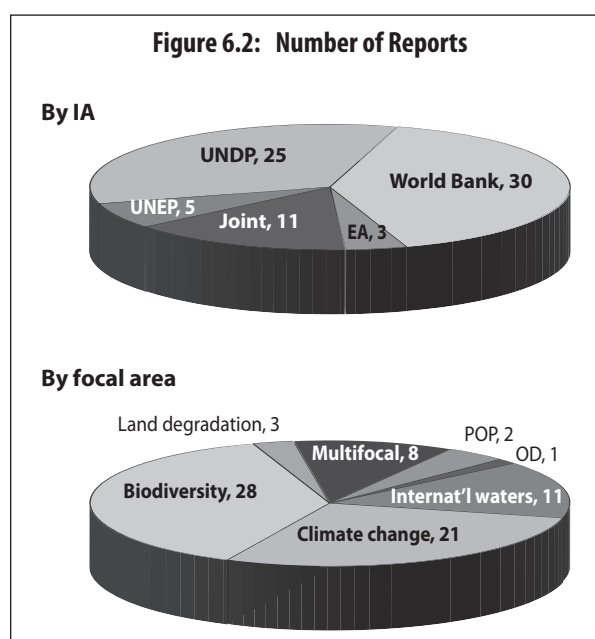
mate change focal area, have sufficient numbers to facilitate inter-group performance comparisons.<sup>6</sup>

## Findings

### Overall Project Compliance

Taking into account the respective scores of projects on each of the critical parameters as well as the cumulative scores, the assessment found that the M&E plans of 58 percent of the 74 projects met the overall M&E expectations at the point of CEO endorsement. Twenty-two percent were noncompliant with one critical parameter, 14 percent with two, and 7 percent with three. The projects in compliance with all six critical parameters had an overall score of 26 or more.

For all IAs except UNEP (whose sample was extremely small), at least half of the projects reviewed met M&E expectations (see table 6.3).



<sup>5</sup>Joint projects are those being implemented by any combination involving two or more IAs and/or EAs.

**Table 6.3: Projects with Overall Compliance by IA**

IA	Compliant Projects	
	Number	Percent
World Bank	15	50
UNDP	17	68
UNEP	2	40
Joint	7	64
EAs	2	67
<b>All agencies</b>	<b>43</b>	<b>58</b>

<sup>6</sup>Three projects in this cohort—Coral Reef Targeted Research and Capacity Building for Management (international waters), Support Program for National Capacity Self-Assessments (multifocal), and Building Capacity for Effective Participation in the Biosafety Clearing-House Project (biodiversity)—are markedly different from the mainstream projects of the GEF. Since these projects are primarily output oriented, designing appropriate outcome indicators for such projects is difficult, and a full results logframe approach also may not be an effective management tool for such projects. Nonetheless, they were included in the assessment in the interest of consistency and completeness. Further, since there are only three such projects, their inclusion does not substantially change the study's overall conclusions.

For all focal areas except land degradation, which only had one compliant project (although here too the sample was quite small), at least half of the projects met M&E expectations (see table 6.4).<sup>7</sup>

**Table 6.4: Projects with Overall Compliance by Focal Area**

Focal Area	Compliant Projects	
	Number	Percent
Biodiversity	14	50
Climate change	16	76
International waters	6	55
Land degradation	1	33
Multifocal	4	50
Ozone depletion	1	100
POPs	1	50
<b>All focal areas</b>	<b>43</b>	<b>58</b>

Overall compliance with GEF M&E expectations seems to be better in projects that have been approved recently. Within the reviewed cohort, 52 percent of the projects approved by the Council before or on June 30, 2003, comply with GEF M&E expectations; 61 percent of the projects approved after this date are in compliance. The recent measures implemented by UNDP and in the biodiversity focal area to address M&E problems seem to have led to an improvement in their overall compliance: for UNDP, the proportion of compliant projects increased from 60 to 72 percent; in biodiversity, it increased from 40 to 56 percent.<sup>8</sup> Although the level of change is not sta-

tistically significant, it is a move in the expected and desired direction. A better picture of changes in compliance over time will emerge only after the cohorts for additional years have been assessed.

### **Strengths and Weaknesses**

All or almost all projects' M&E plans are in compliance with the GEF M&E expectations in terms of specifying relevant (100 percent) and quantifiable (97 percent) indicators, in providing baseline information (92 percent), in explicitly allocating a budget for M&E (92 percent), in specifying responsibilities (99 percent) and time frames (99 percent) for M&E activities, and in specifying targets for project outputs (95 percent). Note, however, that GEF expectations regarding the provision of baseline information have been low: The present project review criteria—recognizing the difficulties and costs associated with establishing baseline conditions for very complex projects—only require projects to provide baseline information within the first year of project implementation. For compliance, a project need only promise to provide baseline information within its first year of implementation. The GEF's new 2006 M&E policy requires projects to provide baseline information at the point of work program inclusion except in those rare cases when such is not possible. Clearly, in the current situation where more than half of the projects just promise that they will provide baseline information before the end of the first year, the exception is being made more frequently than could be called "rare," and the resulting level of performance is inadequate. At present, many projects receiving Project Development Facility funds are apparently not addressing baseline data

<sup>7</sup>The three atypical projects did not comply with the M&E expectations. Dropping these from the analysis increases the proportion of projects in overall compliance from 58 percent to 61 percent. Compliance by IA rises to 52 percent for the World Bank; compliance by focal area rises to 52 percent for biodiversity, 60 percent for international waters, and 57 percent for multifocal. Thus, eliminating these projects from the analysis does not substantially change the findings.

<sup>8</sup>The World Bank has recently moved from a logframe-based monitoring system to a results framework-based monitoring system. Although the quality of the M&E plans in those projects that use a results-based framework is better

than those in the logframe-based (58 percent compared to 37 percent), the review did not provide enough evidence to show that this improvement is indeed due to the policy shift. Both types of World Bank projects are spread out over time, and their number is too small to allow for robust conclusions on this front.



needs during project preparation, although this would appear to be a highly appropriate mechanism for doing so.

The levels of compliance are lower in describing a methodology for baseline data collection (84 percent), specifying targets for objectives and outcomes that are based on an assessment of initial conditions (82 percent), discussing provisions for terminal evaluation (77 percent), and spelling out sufficient (76 percent) and specific (57 percent) indicators. More attention needs to be given to these issues.

The assessment found that the UNDP projects received higher ratings than did the World Bank's in several areas: providing a description of the methodology for determining the baseline, explicitly allocating budget to M&E activities, specifying the responsibilities and time frames for all M&E activities, and specifically mentioning that mid-term reviews and terminal evaluations would be undertaken. On all of these parameters, it is reasonable to expect that performance by other agencies also could be substantially improved if more attention were given to these M&E aspects by task teams and management during project preparation.

The World Bank and UNDP have comparable levels of performance, yet both need to improve in identifying specific and sufficient indicators, providing baseline information up front, and specifying targets based on assessment of initial conditions. In these areas, a significantly higher level of effort will be required, as there are often some fundamental technical issues to be addressed (defining the appropriate units of measurement for biodiversity or land degradation projects, dealing with long time lags in outcomes, and so on).

At the focal area level, the assessment found that climate change does better than biodiversity in

terms of providing specific, sufficient, relevant, and quantifiable indicators for project objectives and outcomes; providing baseline information on indicators up front; and specifying targets based on assessment of initial conditions. The parameters for which performance of climate change and biodiversity vary are different from those on which UNDP and the World Bank differ. While differences between these IAs appear to be due to variation in the level of management attention given to M&E arrangements at entry, the differences among the focal areas appear to be driven by more fundamental challenges (technical complexity and measurement issues) and will naturally be more difficult to address.

In recent years, the GEF focal area task forces have been attempting to define core focal area indicators and develop tracking tools to improve the quality of M&E arrangements.

- The Biodiversity Focal Area Task Force has developed a tracking tool to monitor project performance in catalyzing the sustainability of protected area systems (strategic priority 1) and mainstreaming biodiversity in production landscapes and sectors (strategic priority 2), on programmatic indicators. The task force is also searching for ways to address the M&E concerns related to capacity building for the implementation of the Cartagena Protocol on Biosafety (strategic priority 3) and generation and dissemination of best practices for addressing current and emerging biodiversity issues (strategic priority 4); in these areas, the very nature of the projects makes the design of good and cost-effective M&E plans difficult.
- The climate change focal area has been working on standardizing programmatic and portfolio-level indicators so that the projects of each strategic priority have a uniform set of indicators.



The task force is also refining review standards for M&E arrangements in projects at entry.

- The international waters focal area team has developed a framework to identify process and stress reduction. It is also in the process of defining stress-reduction indicators for nutrient reduction and groundwater projects and testing approaches to measure environmental catalytic impacts of GEF projects.
- The Land Degradation Focal Area Task Force is developing a framework to identify results indicators on the basis of specific global environmental benefits. Efforts in this focal area are also aimed at identifying best practices and analytic tools that can be applied beyond GEF projects.

The process of indicators development is complex. It requires the application of current scientific knowledge and sound technical expertise to results measurement. In developing the tracking tools used to monitor performance on biodiversity strategic priorities 1 and 2, the GEF drew on existing instruments that incorporated current scientific and technical knowledge on the subject. In other focal areas, such instruments do not exist, and the relevant knowledge must be compiled and assessed in light of the GEF's needs to track results. Some focal areas are seeking access to this expertise through partnerships. For example, the International Waters Focal Area Task Force is establishing a partnership with the groundwater initiative of UNESCO's International Hydrological Programme Working Group to develop groundwater indicators and is working with scientists at Iowa State University to develop catalytic indicators for nutrient reduction projects in the Danube Black Sea basin. The Land Degradation Focal Area Task Force is collaborating with the UN University to develop a framework to track results of sustainable land management activities. This framework

will serve as a basis for the subsequent development of GEF-specific indicators in this focal area.

Such efforts as the above need to be intensified to further the development of indicators and tracking tools. This will require corporate investments to address the technical challenges specific to each focal area, build consensus on indicators, define ways to roll up results at the portfolio level, and find ways to address issues related to attribution of outcomes to GEF initiatives.

## **Project Review Arrangements**

### ***Availability of Reviews and Comments***

In the project documents maintained in the GEF Secretariat's Project Management Information System (PMIS), GEF Secretariat reviews were included for all projects, the STAP reviews and corresponding project team responses were included for 70 projects (95 percent), but the compilation of Council member comments were only included for 32 projects (43 percent). An examination of other databases maintained by the GEF Secretariat showed that comments by Council members had been documented for at least 69 projects (93 percent). Thus, either a substantial number of projects do not include Council member comments in the subsequent versions of the project proposal, or else the documents in which the comments are recorded are not being maintained in the PMIS.

### ***GEF Secretariat Review***

The GEF Secretariat reviewers appraise projects throughout the pipeline process including at pipeline entry, work program inclusion, and CEO endorsement. They may require the project team to rework the project documents if these did not meet the GEF requirements for a particular review stage. The GEF Secretariat reviewers are expected to address M&E issues in their reviews

in a comprehensive manner. However, the guidance does not fully clarify what is expected from reviewers in terms of addressing M&E issues. The project review criteria (in “GEF Project Cycle,” GEF/C.16/Inf.7) and the 2002 publication *Monitoring and Evaluation Policies and Procedures*, despite covering M&E issues in detail, do not specify compliance standards for M&E parameters. Consequently, there is wide variation among reviewers as to how they interpret and apply M&E standards.

The GEF Secretariat reviewers pointed out at least one weakness in the M&E arrangements in 55 percent of the projects they reviewed. One would expect that they would provide more such comments on those projects that did not comply with the GEF M&E expectations, and, indeed, they did point out at least one weakness in 48 percent of these projects. However, this means that a substantial number of projects in which the M&E arrangements at the point of CEO endorsement were not in compliance with GEF M&E expectations at entry had not been commented upon by the GEF Secretariat.

No assessment could be made of the extent to which the comments by GEF Secretariat reviewers were incorporated into project documents because a task team is not required to formally document its responses to GEF Secretariat comments.

### **Council Member Comments**

The Council members comment on project proposals at the work program inclusion stage. At that stage, the level of project preparation required in terms of M&E arrangements is greater than that required at the earlier pipeline entry stage. Unlike the STAP roster reviews, where generally only one reviewer is involved, many Council members may choose to comment on a project proposal simultaneously and independently of each other. For the

purpose of this study, the compilation of Council member comments on any given project proposal has been considered a single review, since these compilations often result in a task team making improvements in the project proposal documents. In this sense, the compilations of Council member comments perform a similar function to the GEF Secretariat and STAP roster reviews.

Overall, the Council members pointed out at least one weakness in M&E arrangements in 58 percent of the projects. In projects that did not comply with the GEF M&E expectations, Council members noted at least one weakness in the M&E plans of 68 percent of such projects. Most comments by different Council members on M&E issues in any given project were consistent with each other. Thus, the Council members together appear to be thorough and consistent in flagging M&E weaknesses.

Although Council member comments are available for 69 projects, these were incorporated in the project documents maintained in the PMIS in only 32 instances.<sup>9</sup> Of these 32 projects, the Council had pointed out weaknesses in the M&E plan in 20 (63 percent). In all 20 cases, the project team claimed to have addressed all or some of the M&E issues raised by the Council members.

### **STAP Roster Review**

To date, STAP roster reviewers have not been asked to address specific M&E issues in their reviews, and the terms of reference for the STAP roster review addresses M&E only very peripherally. Consequently, in most cases, STAP reviewers do not address such M&E issues as technical feasibility of indicators, feasibility of methodology, and

---

<sup>9</sup>Comments for the other 37 projects were accessed from another central database, which exclusively maintains Council reviews. It is not known whether the task team responded to these comments, and team responses to such comments are not available.

cost effectiveness of M&E systems—even though these could be considered technical aspects of a project on which STAP roster reviewers might have considerable expertise.

STAP roster reviewers appraise the project proposal documents prior to work program inclusion. They therefore review project documents when they are at a relatively preliminary stage.

STAP roster reviewers pointed out weaknesses in the M&E arrangements of 40 percent of the projects they reviewed. For projects that did not comply with the GEF M&E requirements, STAP roster reviewers pointed out at least one weakness in the M&E plan of 39 percent of these projects. Thus, even though the STAP roster reviewers are not required to address M&E concerns in their

reviews, in many instances they do address them on their own initiative.

For almost all of the 28 projects (93 percent) in which the STAP roster reviewer pointed out a weakness in the M&E plan, the task team responded specifically to the comments. In 12 instances (43 percent), the project team claimed to have addressed all of the concerns of the STAP reviewer; in 9 (32 percent), the team claimed to have addressed some of the concerns; and in 5 (18 percent), the extent of the changes made was not described. For those proposal documents where only a few or none of the changes suggested by the STAP reviewer had been made, the project team did provide an explanation for non-incorporation. These results indicate that STAP roster reviews are taken seriously by proposal preparation teams.

# MONITORING AND EVALUATION

## 7. Quality of Terminal Evaluation Reports

---

High-quality terminal evaluation reports that provide an accurate assessment of project accomplishments and shortcomings are not only essential as a learning tool, but are also important as the building blocks for the APR's assessment of project outcomes and sustainability. The GEF Evaluation Office began rating the quality of project terminal evaluation reports in 2004, which allows for a comparison with 2005 terminal evaluation reports. Annex F presents a list of the terminal evaluation reports reviewed in FY 2005 and their respective ratings.

There has been an improvement in the quality of terminal evaluation reports submitted in FY 2005 as compared to those submitted in the previous year. Improvements are most pronounced in the terminal evaluation reports submitted by UNDP. A detailed assessment of the factors driving the quality of terminal evaluation reports shows that Implementing Agencies are addressing most of the key quality issues identified by the Evaluation Office in 2004.

This marks the first time that differences between the outcome/sustainability ratings given in the terminal evaluation reports by the IAs and the terminal evaluation reviews by the Evaluation Office have been reported in the APR. No significant differences were observed between the Evaluation Office and the World Bank's IEG ratings on a six-point scale. On average, UNEP's terminal evalua-

tion reports tend to rate performance one point higher (on a six-point rating scale) than does the Evaluation Office. Since a large number of UNDP terminal evaluation reports did not provide ratings on outcomes and sustainability, it is difficult to know the extent to which it overrates performance, if at all.

Agency terminal evaluation reports continue to be weak in terms of providing sufficient information to assess the quality of monitoring (particularly in the climate change focal area) and in reporting on actual costs including total costs, costs disaggregated at the activity level, and cofunding. UNEP's terminal evaluation reports also frequently exhibited inconsistencies between the report text and the ratings.

### 7.1 Assessment Approach

The 41 terminal evaluation reports were assessed using the following questions:

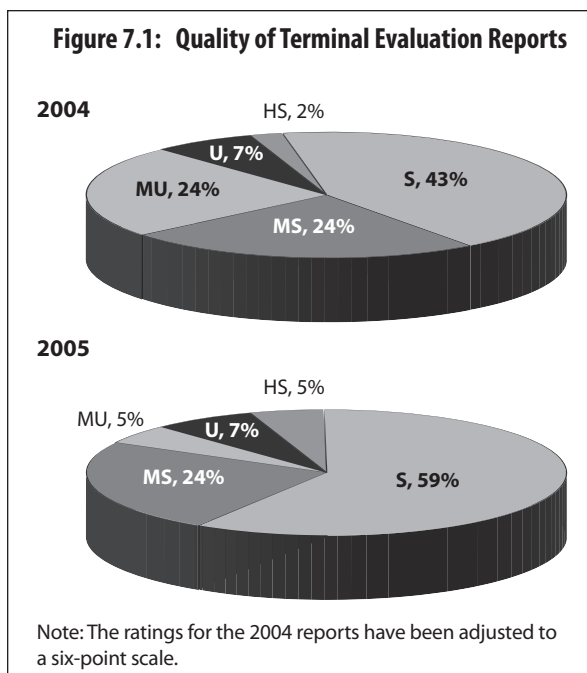
- Did the report present an assessment of relevant outcomes and achievement of project objectives in the context of the focal area program indicators, if applicable?
- Was the report consistent, the evidence complete and convincing, and the ratings substantiated when used?
- Did the report present a sound assessment of sustainability of outcomes?

- Were the lessons and recommendations supported by the evidence presented?
- Did the report include the actual project costs (total and per activity) and actual cofinancing used?
- Did the report include an assessment of the quality of the project M&E system and its use for project management?

## 7.2 Findings

The overall quality of terminal evaluation reports improved from 69 percent rated as moderately satisfactory and above in 2004 to 88 percent so rated in 2005 (see figure 7.1).

The terminal evaluation quality ratings for the World Bank and UNDP were higher in FY 2005 compared to 2004 (see table 7.1). Although the numbers for UNEP are still too small to use in drawing firm conclusions, it is worth noting that the terminal evaluation quality ratings for 2005 dropped from those in 2004. Of the 41 terminal evaluation reports reviewed in 2005, two did not provide sufficient information to assess project outcomes, and five did not provide sufficient information to make an assessment of sustainability. A higher proportion of climate change terminal



evaluations (4 out of 8 in 2004 and 6 out of 10 in 2005) continue to provide insufficient information to assess the quality of the project M&E system.

The quality of terminal evaluation reports is an area in which quick improvements can be expected, and it is likely that IA actions undertaken to address issues raised in the 2004 APR (see box 7.1) have contributed to the significant improvement in quality ratings seen so far. The 2003 issuance of clearer

**Table 7.1: Ratings on Quality of Terminal Evaluation Reports, by IA**

Rating	World Bank		UNDP		UNEP	
	2004	2005	2004	2005	2004	2005
HS	1	0	0	2	0	0
S	10	12	4	10	5	1
MS	6	0	3	9	1	2
Subtotal	17	12	7	21	6	3
MU	4	0	4	1	2	2
U	0	0	2	1	1	1
HU	0	0	0	0	0	0
Subtotal	4	0	6	2	3	3
<b>Total</b>	<b>21</b>	<b>12</b>	<b>13</b>	<b>23</b>	<b>9</b>	<b>6</b>

### Box 7.1: Changes at UNEP and UNDP to Improve the Quality of Project Terminal Evaluation Reports

UNDP and UNEP have recently undergone internal changes to improve their project evaluation processes and better address GEF needs. For example, in FY 2005, UNEP placed its Evaluation and Oversight Unit directly under the agency's executive director, which gives it greater independence from other operating units. Consistent with the new GEF M&E policy, UNEP has also adopted a six-point rating scale and requires that the GEF Evaluation Office guidelines for terminal evaluations be part of all terms of references for GEF project evaluations. Further, UNEP has made all GEF terminal evaluation reports subject to quality assessment reviews by its Evaluation and Oversight Unit and will forward these to the Evaluation Office along with the terminal evaluation reports.

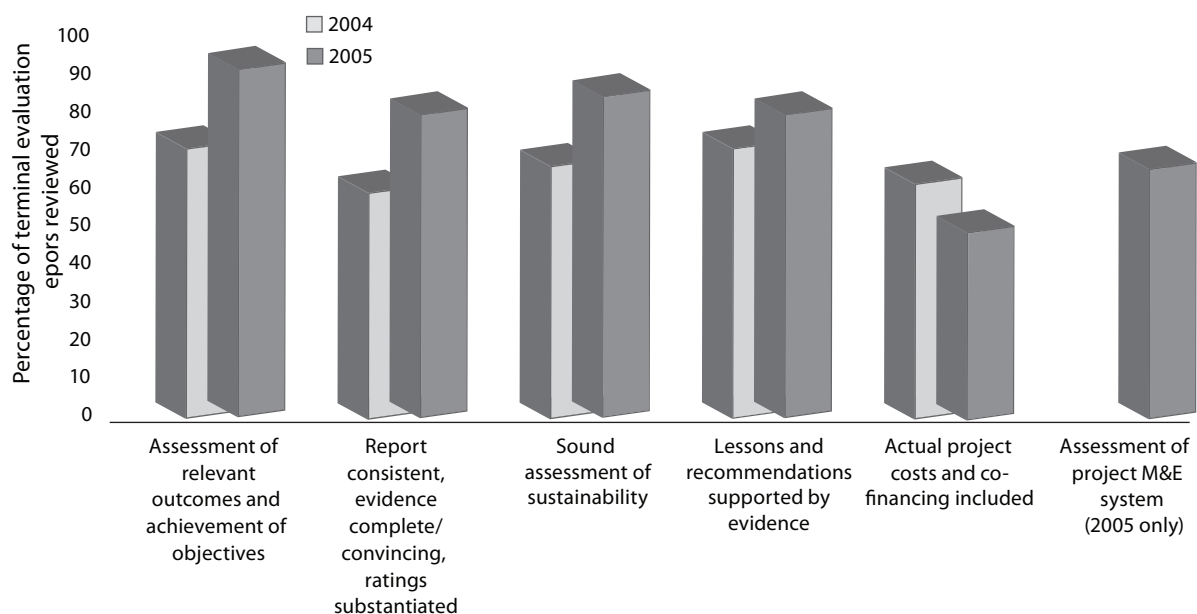
A recent UNDP study, *Status of Monitoring and Evaluation in UNDP-GEF Projects*, cited the highly decentralized nature of the agency's evaluative process as a key factor in the large variation in quality of terminal evaluation reports. Project evaluations are organized by country offices and carried out by individual consultants, and not all countries have in-residence evaluation expertise. To address these weaknesses, UNDP's GEF coordinating unit developed new project M&E guidance and tools, and—in close coordination with UNDP's regional coordination units—aggressively disseminated those instruments through a series of country workshops and over the Internet. In 2005, the UNDP Evaluation Office also assumed responsibility for evaluation of GEF-funded activities from the regional coordination units to better harmonize GEF project evaluations with UNDP's own evaluation practices.

guidelines for the preparation of terminal evaluation reports was likely another contributing factor.

A detailed assessment of the quality of terminal evaluation reports using Evaluation Office crite-

ria shows that IAs are addressing most of the key issues the Evaluation Office identified last year (see figure 7.2). Only the reporting of actual project costs including cofinancing received lower

**Figure 7.2: Percentage of Terminal Evaluation Reports Rated Moderately Satisfactory or Above, by Assessment Criteria**



Note: The ratings for the 2004 reports have been adjusted to a six-point scale.



ratings in 2005 than in 2004: Over 60 percent of the 2004 terminal evaluations reported on actual project costs; just over 50 percent were rated as providing moderately satisfactory or better information in 2005.

The ratings for the assessment of outcomes in terminal evaluation reports were higher in 2005 for all IAs (see annex G). For the World Bank and UNDP, the 2005 ratings were higher for the other criteria as well—namely report consistency, sound assessment of sustainability, and lessons supported by evidence. For these three criteria, the 2005 ratings for UNEP dropped from their 2004 levels.

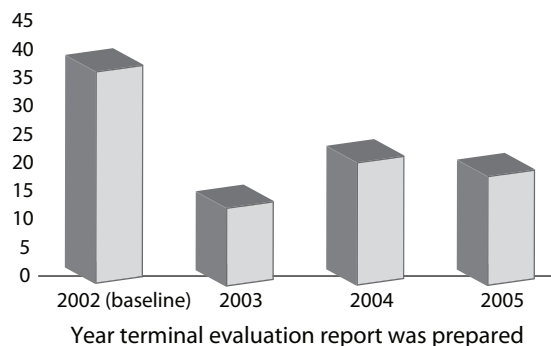
Most of the World Bank terminal evaluation reports were rated moderately satisfactory or above on their inclusion of actual project costs and cofinancing (10 of 12 in 2005); UNDP and UNEP continue to have a large proportion of projects rated less than moderately satisfactory on this criterion (15 of 23 for UNDP and 2 of 6 for UNEP in 2005).

This year, the Evaluation Office began tracking reporting on project monitoring systems as part of the quality assessment of terminal evaluations. Looking back to 2002 (when the reviews of terminal evaluation reports were first undertaken) reveals that terminal evaluation reporting on quality of project monitoring seems to have improved since the Evaluation Office issued its 2003 guidelines requesting that all terminal evaluation reports include an assessment of project monitoring systems (see figure 7.3).

### 7.3 IA Versus GEF Evaluation Office Ratings

On the whole, the IAs tend to overrate project outcomes and sustainability when compared to Evaluation Office ratings (see figure 7.4). The Evaluation Office conducted a more in-depth analysis

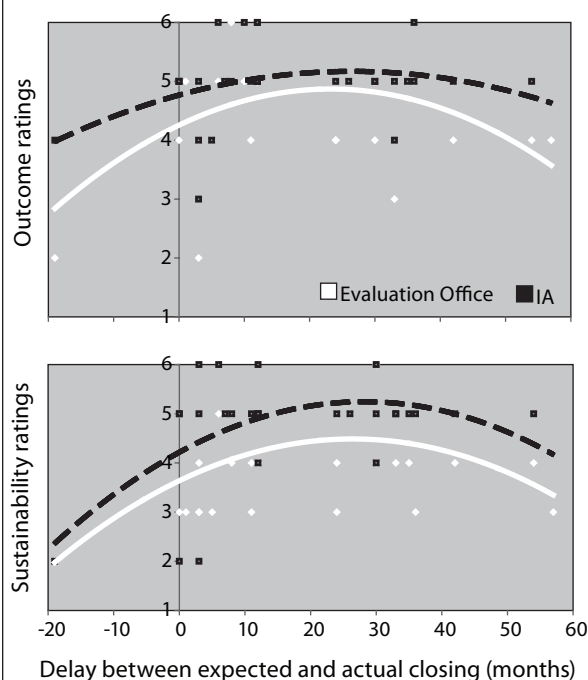
**Figure 7.3: Percentage of Terminal Evaluation Reports Not Providing Sufficient Information on Quality of Project M&E System**



to assess the differences in outcome and sustainability ratings for individual projects by the Evaluation Office and the respective IA.

Only 30 terminal evaluations reviewed in 2005 provided ratings on outcomes. When a binary rat-

**Figure 7.4: IA and GEF Evaluation Office Ratings on Outcomes and Sustainability**



ing scale is used (see box 7.2), there is a disconnect in the ratings in only two instances; in both of these, the respective IA rated project outcomes as moderately satisfactory, while the Evaluation Office rated them as moderately unsatisfactory or worse.

One of these two disconnects was for the World Bank's Natural Resources Management project in Ghana. The World Bank rated the outcomes of this project as moderately satisfactory, whereas the Evaluation Office rated them as moderately unsatisfactory. There was no disconnect in the ratings of the World Bank and Evaluation Office for the Bank's other 11 projects providing outcomes ratings or on its 10 projects rating sustainability (see table 7.2).

The second disconnect in outcome ratings on a binary scale was for UNEP's global project Barriers and Best Practices in Integrated Manage-

ment of Mountain Ecosystems. Here again, the IA accorded the project a moderately satisfactory rating on outcomes, which the Office Evaluation rated these as unsatisfactory. This rating was based on the fact that the project's terminal evaluation report indicated that the project had failed to achieve many of its outcomes and that its sustainability was not ensured beyond project closure. The project was badly implemented due to poor project management, which in fact caused UNEP to terminate the project altogether.

For UNDP projects where ratings by both the IA and the Evaluation Office are available, there is no disconnect in the binary ratings. How representative this is of UNDP's FY 2005 cohort is unclear, however. Of the 23 UNDP projects for which terminal evaluations were submitted in FY 2005, only 13 provided ratings on outcomes and 10 on sustainability. Consequently, the absence of a disconnect on a binary scale cannot be generalized to the total UNDP cohort for FY 2005.

When the six-point rating scale is used, there is little difference between the World Bank's project outcome and sustainability ratings and those of the Evaluation Office.<sup>1</sup> Although the outcome and sustainability ratings of UNDP and UNEP do not differ from those of the Evaluation Office on a binary scale (except for the one instance noted above), there are more apparent differences in the ratings on a six-point scale. On this scale, UNDP overrated the outcomes of its projects by an average of 0.4 points and overrated sustainability by an average of 0.3 points. UNEP overrated outcomes by an average of 1.0 point and sustainability by an average of 1.3 points.

Despite the magnitude of the difference in average ratings—especially for UNEP—these differences

#### **Box 7.2: Binary Rating Scale**

To calculate the rating difference for 2005 terminal evaluation reports using a binary system, the six-point rating scale used for outcomes and sustainability was converted to a binary system of 1 and 0. Thus, project outcomes rated as moderately satisfactory, satisfactory, or highly satisfactory were given a 1; and those rated moderately unsatisfactory or worse received a 0. Similarly, project sustainability ratings of moderately likely, likely, or highly likely received a 1; ratings of moderately unlikely or worse received a 0. The next step was to calculate for each project the difference between the rating provided by the GEF Evaluation Office and that provided in the terminal evaluation report (or in the terminal evaluation review done by the World Bank IEG for World Bank-GEF projects). If this difference was negative, it indicated that the terminal evaluation report provided a higher rating than that provided by the Evaluation Office. The average difference for each IA was calculated. The World Bank IEG also uses this binary system to assess the relevant difference in ratings.

<sup>1</sup>Ratings provided by the World Bank IEG have been considered for the analysis.

in outcome and sustainability ratings are not statistically significant. This is primarily because the numbers of observations where comparisons can be made are very small: 13 and 10 for UNDP and

5 and 4 for UNEP, for project outcomes and sustainability, respectively (see table 7.3). In future, when more observations will be available, better comparisons will be possible.

**Table 7.2: Average Difference between IA and GEF Evaluation Office Ratings on a Binary Scale**

Agency	Number of Terminal Evaluation Reports Reviewed	Outcomes		Sustainability	
		Number of Reports Providing Ratings	Number of Projects with Different Ratings	Number of Reports Providing Ratings	Number of Projects with Different Ratings
World Bank (IEG)	12	12	1	10	0
UNDP	23	13	0	10	0
UNEP	6	5	1	4	1
<b>Total</b>	<b>41</b>	<b>30</b>	<b>2</b>	<b>24</b>	<b>1</b>

**Table 7.3: Difference between IA and GEF Evaluation Office Ratings on a Six-Point Scale**

Agency	Number of Terminal Evaluation Reports Reviewed	Outcomes		Sustainability	
		Number of Reports Providing Ratings	Number of Projects with Different Ratings	Number of Reports Providing Ratings	Number of Projects with Different Ratings
World Bank (IEG)	12	12	2	10	3
UNDP	23	13	7	10	5
UNEP	6	5	4	4	3
<b>Total</b>	<b>41</b>	<b>30</b>	<b>13</b>	<b>24</b>	<b>11</b>

# MONITORING AND EVALUATION

## 8. Management Action Records

The GEF Council approved the format and procedures for the GEF Management Action Record at its November 2005 meeting and requested that the GEF Secretariat and GEF Evaluation Office prepare MARs, in consultation with appropriate entities, for submission to the June 2006 Council session. The format and procedures were developed in consultation with the Secretariat and IAs; at this juncture, there has been little EA involvement.

Each MAR contains columns for recommendations, management responses, and Council decisions to be filled in by the Evaluation Office (see figure 8.1). Management (that is, the GEF Secretariat and the IAs) is invited to provide a self-rating of the level of adoption of Council decisions on recommendations and add any comments as necessary. Subsequently, the Evaluation Office enters its own rating of adoption, with comments, in time for presentation to the Council. The ratings assessing progress in the adoption of Council decisions are as follows:

- **High**—fully adopted
- **Substantial**—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 of or plan for adoption, or plan and actions for adoption are in a very preliminary stage

MARs will be updated annually. After an item has been reported as fully adopted or is no longer relevant, it will be deleted from the MAR; after all items have been adopted, the MAR will be archived.

The Evaluation Office prepared draft MARs for those evaluation reports that received a management response. These seven MARs were forwarded to the Secretariat on March 17, two months prior

**Figure 8.1: Management Action Record**

Recommendation	Management Response	Council Decision	Rating in Progress of Adoption			
				Comments	GEF EO	Comments

to the Council session. The Evaluation Office requested that management input be received by April 17 to allow sufficient time to verify the assessment and draft a synthesis to be included in the present Annual Performance Report. Two MARs were received the last week of April, four more the first week of May, and the final one on May 8. The late receipt of the MARs impaired the Evaluation Office's ability to verify management's assessment of progress toward adoption of Council decisions.

In all but one case, the assessment proffered by the Evaluation Office on the adoption of Council decisions is of an indicative nature, rather than based on corroborating evidence or data. The single exception is the Evaluation Office's assessment of the adoption rate of the Council's June 2005 decision on transparency in the GEF approval process. GEF management assesses this progress as "medium," given that the work toward establishing a new database for GEF projects has begun. The Evaluation Office bases its assessment on corresponding evidence from the GEF pilot country portfolio evaluation and ongoing work in the Joint Evaluation of the GEF Activity Cycle and Modalities, as well as the ongoing consultative process. Because the experience of project proponents at the country level has not changed and informa-

tion on where projects are in the approval process is still not available, the Evaluation Office concludes that the adoption rate for the Council's decision has been "negligible" so far. The Evaluation Office believes that making information available in a transparent way is not difficult, nor does it need to rely on new database software or systems. Rather, what is needed is discipline in gathering information and presenting it in a clear way on a website.

This first presentation of the MAR has been an experiment and a learning experience. Despite earlier consultations and agreements in principle on how the MAR should be addressed, differences of opinion on how the ratings should be applied played an important role in the delay on the GEF management side in dealing with the MAR. The result was that the Evaluation Office did not have sufficient time to verify the ratings. Based on knowledge gained through other sources and evaluations, the Evaluation Office has indicated the ratings it believes would be justified. The Evaluation Office will present the MARs to GEF management again in March 2007 and is confident that management will be able to deliver its assessment of the adoption of Council decisions on evaluations in sufficient time to ensure that the Evaluation Office can verify these ratings.

# Annex A. Terminal Evaluation Report Review Guidelines

---

The assessments in the Terminal Evaluation Review will be based on the information presented in the Terminal Evaluation. If insufficient information is presented in the terminal evaluation to assess a specific issue such as, for example, the quality of the project M&E system or a specific aspect of sustainability, then the preparer of the review will briefly indicate so in that section and elaborate more if appropriate in the section of the review that addresses the quality of the terminal evaluation report. If the review preparer possesses other independent information such as, for example, from a field visit to the project, and this information is relevant to the review, then it should be included in the review only under the section on “Comments on the summary of project ratings and terminal evaluation findings.”

## A.1 Criteria for the Ratings on the Outcomes<sup>1</sup>

Based on the information in the report, the terminal evaluation review will make an assessment of the extent to which the project’s major

---

<sup>1</sup>The 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; may also include changes resulting from the intervention that are relevant to the achievement of outcomes. *Glossary of Key Terms in Evaluation and Results Based Management*, OECD, Development Assistance Committee. For the GEF, environmental outcomes are the main focus.

relevant objectives<sup>2</sup> were effectively achieved or are expected to be achieved and their relevance. The ratings on the outcomes of the project will be assessed using the following criteria:

- A. **Relevance:** In retrospect, were the project’s outcomes consistent with the focal areas/operational program strategies?
- B. **Effectiveness:** Are the project outcomes as described in the terminal evaluation report commensurable with the expected outcomes (as described in the project document) and the problems the project was intended to address (that is, original or modified project objectives<sup>3</sup>)?
- C. **Efficiency:** Include an assessment of outcomes in relation to inputs, costs, and implementation times based on the following questions: Was the project cost effective? How does the cost-time

---

<sup>2</sup>Objectives are the intended physical, financial, institutional, social, environmental, or other development results to which a project or program is expected to contribute. *Glossary of Key Terms in Evaluation and Results Based Management*, OECD, Development Assistance Committee.

<sup>3</sup>The GEF Evaluation Office is currently working with the GEF Secretariat and IAs 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.



vs. outcomes compare to other similar projects?  
Was the project implementation delayed?

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 not applicable.

## A.2 Criteria for the Rating of Impacts<sup>4</sup>

Impacts are the primary and secondary long-term effects of a development intervention. As such, they might not always be apparent at the project closing. When the impacts are apparent, the terminal evaluations are expected to report them. Special attention is required for assessing social impacts of the GEF-supported interventions.

## A.3 Criteria for the Ratings on Sustainability

Sustainability will be understood as the likelihood of continuation of project benefits after completion of project implementation.<sup>5</sup> Terminal evaluations will identify and assess the key factors required for sustainability and any risks that could undermine the continuation of the benefits at the time of the evaluation. Some of these factors (or risks) might be the presence (or absence) of stronger institutional capacities, legal frameworks, socioeconomic incentives, and public awareness. Risk factors may also include contextual circumstances or developments that are relevant to the

sustainability of outcomes. The following four aspects of sustainability will be addressed: financial, sociopolitical, institutional framework and governance, and replication and catalysis.<sup>6</sup> The following questions provide guidance to assess if the components are met:

**A. Financial resources:** What financial and economic resources will be available to allow for the project outcomes/benefits to be sustained once the GEF assistance ends (resources can be from multiple sources, such as the public and private sectors, income-generating activities, and market trends that support the project's objectives)? What is the risk that these resources will not be available, compromising the sustainability of benefits?

**B. Sociopolitical:** What is the risk that the level of stakeholder ownership is insufficient to allow for the project outcomes/benefits to be sustained? Do the various key stakeholders see it in their interest that the project benefits continue to flow? Is there sufficient public/stakeholder awareness in support of the long-term objectives of the project?

**C. Institutional framework and governance:** What institutional and technical achievements, legal frameworks, policies, and governance structures and processes are in place to allow for the project outcomes/benefits to be sustained? While responding to this question, consider if the required systems for accountability and transparency and the required technical know-how are in place. What is the risk that the institutional framework and governance may be insufficient to sustain the benefits?

<sup>4</sup>Positive and negative, primary and secondary long-term effects produced by a development intervention, directly or indirectly, intended or unintended. *Glossary of Key Terms in Evaluation and Results Based Management*, OECD, Development Assistance Committee. For the GEF, environmental impacts are the main focus.

<sup>5</sup>"GEF Project Cycle," GEF/C.16/Inf.7.

<sup>6</sup>Replication refers to repeatability of the project under quite similar contexts based on lessons and experience gained. Actions to foster replication include dissemination of results, seminars, training workshops, field visits to project sites, and so on. "GEF Project Cycle," GEF/C.16/Inf.7.

**D. Replication and catalysis:** What examples are there of replication and catalytic outcomes that suggest increased likelihood of sustainability?

**Rating System for Sustainability**

- A. Financial resources
- B. Socio-political
- C. Institutional framework and governance
- D. Replication and catalysis

A number rating 1–6 will be provided in each category according to the achievement and shortcomings with: Highly Likely = 6, Likely = 5, Moderately Likely = 4, Moderately Unlikely = 3, Unlikely = 2, Highly Unlikely = 1, and not applicable = 0. If the evaluator is unable to assess any aspect of sustainability, then it may not be possible to assess the sustainability overall. The evaluator will assess if this is the case, and this will be reported in the Annual Performance Report (APR).

Then the sustainability score of project outcomes will be:

$$\text{Sustainability rating}^7 = (A+B+C+D)/4$$

The sustainability score will be rounded and converted to the scale ranging from Highly Likely to Highly Unlikely as described above. If a criterion is rated as “not applicable” (0), then the sustainability rating will be considered as an average of the remaining ratings. For example, if B is zero, then the outcome will be the average of A, C, and D.

<sup>7</sup>Note: For terminal evaluations reviewed in FY 2005, the average of the first three ratings will be used to determine the overall sustainability ratings. Beginning next year, the last two criteria will also be used to determine the average. The reason for this was a previous agreement with the Implementing Agencies regarding the criteria to be used to assess sustainability before the two latter criteria were added.

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

Monitoring is a continuing function that uses systematic collection of data on specified indicators to provide management and the main stakeholders of an ongoing project with indications of the extent of progress and achievement of objectives and progress in the use of allocated funds. Evaluation is the systematic and objective assessment of an ongoing or completed project, its design, implementation, and results. Evaluation may involve the definition of appropriate standards, the examination of performance against those standards, and an assessment of actual and expected results. The aim is to determine the relevance and fulfillment of objectives, efficiency, effectiveness, impact, sustainability, and the worth or significance of the project. An evaluation should provide information that is credible and useful, enabling the incorporation of lessons learned into the decision-making process of both recipients and donors.<sup>8</sup>

The ratings on the quality of the project M&E systems will be assessed using the following criteria:

- A. Whether an appropriate M&E system for the project was put in place (including capacity and resources to implement it) and whether this allows for tracking of progress toward project objectives. The tools used might include a baseline, clear and practical indicators and data analysis systems, or studies to assess results that were planned and carried out at specific times in the project.
- B. Whether the M&E system was used effectively for project management.

<sup>8</sup>*Glossary of Key Terms in Evaluation and Results Based Management*, OECD, Development Assistance Committee, pp. 21 and 27.

#### **Rating System for the Quality of Project M&E Systems**

- A. Effective M&E system in place (indicators, baselines)
- B. Information used for adaptive management

A number rating 1–6 will be provided in 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 = 0. Then the quality of the terminal evaluation reports will be:

Rating on the quality of the project monitoring and evaluation system =  $(A+B)/2$

The total number will be rounded and converted to the scale of HS to HU.

### **A.5 Criteria for the Assessment of Quality of Terminal Evaluation Reports**

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

- A. 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?
- B. Was the report consistent and the evidence complete and convincing, and were the ratings substantiated when used?
- C. Did the report present a sound assessment of sustainability of outcomes?

- D. Were the lessons and recommendations supported by the evidence presented?
- E. Did the report include the actual project costs (total and per activity) and actual cofinancing used?
- F. Did the report include an assessment of the quality of the project M&E system and its use for project management?

#### **Rating System for the Quality of Terminal Evaluation Reports**

- A. Assessment of relevant outcomes and achievement of objectives
- B. Report consistent, evidence complete/convincing, and ratings substantiated
- C. Assessment of sustainability and exit strategy
- D. Lessons supported by the evidence and comprehensive
- E. Assessment of project M&E system
- F. Actual project costs (total and per activity) and actual cofinancing used

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 = 0.

Then the quality of the terminal evaluation reports will be:

Quality of the terminal evaluation reports =  $0.3*(A+B) + 0.1*(C+D+E+F)$

The total number will be rounded and converted to the scale of HS to HU.

## Annex B. Characteristics of Agency Project-at-Risk Monitoring Systems

Agency Name: _____		
Agency has project monitoring and reporting system:	Yes ____	No ____
If yes, system is: Electronic/MIS ____ Paper-based ____		
Reports required to be updated at least twice/year:	Yes ____	No ____
Report includes judgment of overall project performance:	Yes ____	No ____
<b>Report includes judgment of performance of project components:</b>	<b>Yes ____</b>	<b>No ____</b>
<b>Report includes judgment of project risk:</b>	<b>Yes ____</b>	<b>No ____</b>
Report assesses project management performance:	Yes ____	No ____
Report assesses project financial management:	Yes ____	No ____
<b>Report assesses project M&amp;E performance:</b>	<b>Yes ____</b>	<b>No ____</b>
Report tracks project disbursement history:	Yes ____	No ____
Report documents any delays in project effectiveness, key events:	Yes ____	No ____
<b>Projects with performance problems or risks are identified as at-risk or non-performing:</b>	<b>Yes ____</b>	<b>No ____</b>
Projects in risky status are flagged for special attention:	Yes ____	No ____
Report is reviewed by Agency's line management:	Yes ____	No ____
If yes, for information only ____		
If yes, for approval ____		
Report is reviewed by other Agency units:	Yes ____	No ____
<b>Follow-up on at-risk projects includes time-bound action plan:</b>	<b>Yes ____</b>	<b>No ____</b>
At-risk projects are tracked by Agency management:	Yes ____	No ____
Data on project performance and risk are aggregated for portfolio analysis:	Yes ____	No ____
Total "Yes" responses: ____		
Total Critical Elements: ____		

Note: Items in boldface are considered critical elements of risk monitoring.

## Annex C. Inventory of Project-at-Risk Monitoring Systems

Characteristic	ADB	AfDB	EBRD	FAO	IADB	IFAD	UNEP	UNDP	UNIDO	WB	Total "Yes"
Agency has project monitoring and reporting system	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	10
—If yes, system is electronic/MIS	X	X	X	X	X <sup>a</sup>	X	X	X	X	X	10
—If yes, system is paper-based											0
Reports required to be updated at least twice/year	Yes	No	Yes	Yes	No	No	Yes	Yes	Yes	No	6
<b>Report includes judgment of overall project performance</b>	<b>Yes</b>	<b>Yes</b>	<b>Yes</b>	<b>Yes</b>	<b>Yes</b>	<b>Yes</b>	<b>Yes</b>	<b>Yes</b>	<b>Yes</b>	<b>Yes</b>	<b>10</b>
<b>Report includes judgment of performance of project components</b>	<b>Yes</b>	<b>No</b>	<b>Yes</b>	<b>Yes</b>	<b>Yes</b>	<b>No<sup>b</sup></b>	<b>Yes</b>	<b>Yes</b>	<b>Yes</b>	<b>Yes</b>	<b>8</b>
Report includes judgment of project risk	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes <sup>c</sup>	10
Report assesses project management performance	Yes	Yes	Yes	No	Yes	No	Yes	Yes	No	Yes	7
Report assesses project financial management	Yes	Yes	Yes	No	Yes	No	Yes	Yes	No	Yes	7
<b>Report assesses project M&amp;E performance</b>	<b>Yes</b>	<b>No</b>	<b>Yes</b>	<b>No</b>	<b>Yes</b>	<b>Yes</b>	<b>No</b>	<b>Yes</b>	<b>No</b>	<b>Yes</b>	<b>6</b>
Report tracks project disbursement history	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	10
Report documents any delays in project effectiveness, key events	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	10
<b>Projects with performance problems or risks are identified as at-risk or non-performing</b>	<b>Yes</b>	<b>Yes</b>	<b>Yes</b>	<b>Yes</b>	<b>Yes</b>	<b>Yes</b>	<b>Yes</b>	<b>Yes</b>	<b>Yes</b>	<b>Yes</b>	<b>10</b>
Projects in risky status are flagged for special attention	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	9
Report is reviewed by Agency line management	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	10
—If yes, for information only	X			X					X		3
—If yes, for approval		X	X		X	X	X	X		X	7
Report is reviewed by other Agency units	Yes	No	Yes	Yes	No	Yes	Yes	Yes	No	No	6
<b>Follow-up on at-risk projects includes time-bound action plan</b>	<b>Yes</b>	<b>No<sup>d</sup></b>	<b>Yes</b>	<b>Yes</b>	<b>No</b>	<b>No</b>	<b>Yes</b>	<b>No</b>	<b>No</b>	<b>No<sup>e</sup></b>	<b>4</b>

Characteristic	ADB	AfDB	EBRD	FAO	IADB	IFAD	UNEP	UNDP	UNIDO	WB	Total "Yes"
At-risk projects are tracked by Agency management	Yes	Yes	Yes	No	Yes	Yes	No	Yes	Yes	Yes	8
Data on project performance and risk are aggregated for portfolio analysis	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	No	Yes	8
Total Yes responses	17	15	17	12	14	12	15	16	10	14	
Total critical elements	5	4	5	4	4	3	4	4	3	4	

ADB = Asian Development Bank; AfDB = African Development Bank; FAO = Food and Agriculture Organization; IADB = Inter-American Development Bank; IFAD = International Fund for Agricultural Development; UNIDO = UN Industrial Development Organization; WB = World Bank

Note: Items shown in bold are critical elements for risk monitoring.

- a. Not all modules of MIS were operational at time of review.
- b. Reports include narrative/qualitative information on performance.
- c. As of FY 2005, risk rating is calculated by MIS, not by project supervision staff.
- d. Only for countries with portfolio improvement plan.
- e. This is considered good practice but not required.



## Annex D. Methodology and Parameters for Reviewing M&E Arrangements at Entry

### D.1 Steps Followed

1. An APR approach paper outlining a draft framework for assessing the quality assurance system for M&E arrangements at entry was shared with the GEF Secretariat and the Implementing Agencies for their feedback.
2. Council policies and GEF guidance for project M&E systems at entry were reviewed to identify present expectations.
3. The guidance provided to GEF Secretariat project reviewers and the terms of reference for STAP roster reviewers were reviewed to determine expectations in their addressing M&E issues in their reviews.
4. The F-Test and chi square test were used to assess whether there were statistically significant differences in performance across focal areas and IAs. Where differences in performance have been reported on in this APR, they are at a 90 percent confidence level.
5. To assess the extent to which expectations for M&E arrangements at entry are being complied with, all 74 full-sized projects that received CEO endorsement during FY 2005 were examined.<sup>1</sup>
6. The project reviews by the GEF Secretariat, STAP roster reviewers, and GEF Council were examined to determine the extent to which the project review process is able to address M&E issues.
7. Interviews were conducted with the GEF Secretariat focal area leads and M&E coordinators from the IAs to further probe the issues emerging from the reviews and to identify concerns.
8. The preliminary findings were disclosed to the GEF Secretariat, focal area task forces, and IAs to verify the accuracy of the data and to identify possible methodological concerns.

### D.2 Assessment of Quality

An instrument was developed to assess the quality of M&E plans. This instrument measures 13 specific aspects (parameters) of M&E quality, which are based on the GEF Secretariat's review criteria (set forth in "GEF Project Cycle," GEF/C.16/Inf.7) and the guidelines contained in the GEF 2002 *Monitoring and Evaluation Policies and Procedures*. In some cases, the parameters out-

---

<sup>1</sup>Three projects in this cohort—Coral Reef Targeted Research and Capacity Building for Management (international waters), Support Program for National Capacity Self-Assessments (multifocal), and Building Capacity for Effective Participation in the Biosafety Clearing-House Project

---

(biodiversity)—are markedly different from mainstream GEF projects. Since such projects are primarily output-oriented, designing appropriate outcome indicators for them is difficult, and a full results logframe approach may not be an effective management tool. However, these projects were included in the assessment in the interests of consistency and completeness; since there are only three such projects, their inclusion will not substantially change the study's overall conclusions.

lined in these documents were refined to facilitate consistency and objectivity in the application of the assessment instrument.<sup>2</sup> Certain technical or operational elements were not included in this instrument, as this would have required specialized technical expertise on individual projects and would have introduced greater subjectivity into the review process.<sup>3</sup> The 13 M&E parameters are listed below.

### Critical Parameters

- Are the indicators relevant to the specified objectives and outcomes?
- Are the indicators sufficient to assess achievement of the objectives and outcomes?
- Has adequate and relevant baseline information been provided or promised?
- Has a separate budget been allocated for M&E activities?
- Have the targets been specified for the indicators for project objectives and outcomes in the logframe?
- Are the specified targets for indicators of project objectives and outcomes based on initial conditions?

### Other Parameters

- Is there at least one specific indicator in the logframe for each of the project objectives and outcomes?
- Are the indicators for project objectives and outcomes in the logframe quantifiable?

<sup>2</sup>The parameters that were refined further included specific and sufficient indicators, specific targets for the chosen indicators, and the specific targets being based on some assessment of the initial conditions.

<sup>3</sup>The parameters that were mentioned in the two documents but not used in the instrument include discussion of key assumptions of the project, sufficiency of M&E budget, and adaptive management.

- Has the baseline data collection methodology been explained?
- Have the responsibilities for M&E activities been specified?
- Have the time frames for M&E activities been specified?
- Have the performance standards (targets) been specified in the logframe for the project outputs?
- Do the project documents mention provisions for mid-term and terminal evaluations?

Information for rating the 13 assessment parameters was gathered by examining various sections of the respective project documents that mention M&E, including the logframe (or results framework), the M&E section, the annexes, and budget tables. Each project's performance on the 13 M&E parameters was then recorded and scored using the assessment instrument presented in table D.1.

The score on each individual parameter could range between 1 and 3, where 1 was the minimum possible score and 3 the maximum. A score of 2 corresponded with the minimum level of expected performance required for compliance on any given parameter. One would expect the M&E plans to score better than or at least equal to the required minimum standard.

For a project's M&E quality to be considered acceptable, it should score 2 or more on each of the five parameters classified as critical. The "other" parameters are evaluated cumulatively rather than individually; if performance on several "other" parameters is deficient, this is taken to indicate inadequate preparation of the M&E plan. The total score (summation of the scores of the project on all the parameters) required for a project to be considered acceptable is 26 or more.

**Table D.1: Instrument for Assessment of M&E Plans**

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

## Annex E. Portfolio Performance on M&E Arrangements at Entry, by Parameter

---

- **M&E Parameter #1: Is there at least one specific indicator in the logframe for each of the project objectives and outcomes?** The absence of a specific indicator for any of a project's stated objectives or outcomes implies that it will be difficult to ascertain whether that objective or outcome has been achieved. For compliance on this parameter, each of the objectives and outcomes listed in the logframe should have a corresponding indicator. While 57 percent of the projects had a specific indicator for each of the objectives and outcomes, 43 percent lacked such indicators in one or more instances.
- **M&E Parameter #2: Are the indicators relevant to the specified objectives and outcomes?** This is a critical parameter. For compliance, all or almost all of the indicators listed in the logframe are expected to be relevant to the corresponding objectives and outcomes. In instances where an indicator is not relevant, additional costs may be incurred in collecting information that is not essential. The presence of irrelevant indicators also indicates a lack of clarity regarding how various project components will help achieve the project's overall objectives. In 78 percent of the projects, all indicators were relevant; in the remaining 22 percent, almost all of the listed indicators were relevant. Thus, all the projects complied on this parameter.
- **M&E Parameter #3: Are the indicators sufficient to assess achievement of the objectives and outcomes?** The specified indicators must, when taken together, indicate the extent to which a project's overall objectives and outcomes have been achieved. This is a critical parameter and for compliance on it the listed indicators should be sufficient or largely sufficient. For 47 percent of the projects, the indicators were sufficient without any qualifications. For 28 percent—allowing for some minor omissions—the indicators were largely sufficient. Thus, 76 percent of the projects were in compliance on this parameter.<sup>1</sup>
- **M&E Parameter #4: Are the indicators for project objectives and outcomes in the logframe quantifiable?** Specifying the indicators in a quantifiable format facilitates the establishment of objective targets. For compliance on this parameter, all or at least some of the indicators should be presented in a quantifiable form. For 57 percent of the projects, all of the indicators were quantifiable, for 41 percent some of the indicators were quantifiable; this implies compliance by 97 percent of the projects. The remaining 3 percent were in noncompliance, as none of their listed indicators were in a quantifiable form.

---

<sup>1</sup>Figures do not sum to total because of rounding.

- **M&E Parameter #5: Has adequate and relevant baseline information been provided or promised?** Unless the starting point is known, it is difficult to determine progress. Therefore, this parameter has been identified as critical. Although a strong case can be made for requiring baseline information up front, the present project review criteria—recognizing the difficulties and costs associated with establishing baseline conditions for very complex projects—require projects to provide baseline information within the first year of project implementation. Therefore, for compliance on this parameter, a project should at least promise to provide baseline information within its first year of implementation. Nineteen percent of projects provided complete baseline information up front; 20 percent provided baseline information on some indicators, while promising to provide the remaining information within the first year; and 53 percent promised they would provide baseline information within the first year. In all, 92 percent of the projects were in compliance on this parameter, and 8 percent were in noncompliance.

Note, however, that the new *GEF Monitoring and Evaluation Policy* (Evaluation Document 2006, No. 1) calls for upfront provision of baseline information except in rare situations, wherein baseline information can be provided within the first year. Clearly, in the current situation, where 53 percent of projects only promise to submit baseline information within their first year, the exception is being made more frequently than could be called “rare.”

- **M&E Parameter #6: Has the baseline data collection methodology been explained?** Explaining how a baseline will be determined helps in ascertaining the methodology’s feasi-

bility.<sup>2</sup> The assessment instrument notes those cases in which a description of the baseline collection methodology was provided. Eighty-four percent of the projects were in compliance for this indicator, providing some explanation of how baselines for indicators would be determined. The remaining 16 percent of the projects did not provide any information and were thus in noncompliance.

- **M&E Parameter #7: Has a separate budget been allocated for M&E activities?** Allocation of a sufficient budget to M&E activities is essential to ensure that these activities are not stalled for want of funds. How much is “sufficient,” however, depends on such factors as the project’s size; focal area; and institutional, local, and technological context. A great degree of variation thus may be expected across projects. This parameter has been identified as critical, and for compliance a project should make explicit provision for M&E activities in its budget. In all, 92 percent of the projects explicitly allocated some funding to M&E activities, whereas 8 percent did not.
- **M&E Criterion #8: Have the responsibilities been specified for M&E activities?; M&E Criterion #9: Have the time frames been specified for M&E activities?** For sound M&E planning and implementation, it is important to specify the responsibilities and time frames for each M&E activity. For compliance on these parameters, responsibilities and time frames for at least some M&E activities should be specified. Fifty-seven percent of the projects clearly specified M&E responsibilities, 42 percent broadly specified them, and one project (1 percent) did not specify them. A similar pattern was

<sup>2</sup>Note that the technical feasibility or merits of the given methodology were not assessed, as such an assessment would require a high degree of project-specific technical expertise.



observed in terms of specifying the time frames for M&E: 57 percent of the projects specified time frames for all M&E activities, 42 percent for some, and 1 percent for none. Thus, on both these parameters, 99 percent of the projects complied with Council expectations.

- **M&E Parameter #10: Have the performance standards (targets) been specified in the log-frame for the project outputs?** Specification of targets for project outputs facilitates monitoring of resource allocation and progress of activities during project implementation. For compliance on this parameter, a project should provide targets for at least some of the outputs. Sixty percent of the projects provided targets for all project outputs, 35 percent provided them for some, and 5 percent provided no targets. Thus, 95 percent of the projects were in compliance on this parameter.
- **M&E Parameter #11: Have the targets been specified for the indicators for project objectives and outcomes in the logframe?** Whether a project achieves its desired results depends on expectations regarding the agreed indicators. Therefore, specification of targets before project launch has been identified as a critical parameter. For compliance on this parameter, the targets for at least some of the indicators should be specified. Forty-six percent of the projects specified targets for all indicators, 43 percent specified targets for some, and 12 percent specified no targets. Thus, 89 percent of the projects were in compliance on this parameter.
- **M&E Parameter #12: Are the specified targets for indicators of project objectives and outcomes based on initial conditions?** Realistic targets are not only a yardstick against which a project's performance can be assessed but can also be a source of motivation for the project

team. To be realistic, the targets should be based on some assessment of the initial conditions and on the level of change that could reasonably be expected by project end. This review did not attempt to judge whether the level of targeted change specified for a given indicator is realistic; rather, the instrument focused on whether the stated targets were based on some assessment of initial conditions. This parameter has been identified as critical, and for compliance the specified targets for at least some indicators should be based on an assessment of the initial conditions. For 23 percent of the projects, the targets for all indicators were so based; 59 percent based some of the indicators on an assessment of the initial conditions. Thus, 81 percent of the projects performed satisfactorily on this parameter, and 19 percent—which did not base any targets on an assessment of initial conditions—had unsatisfactory performance.

- **M&E Parameter #13: Do the project documents mention provisions for mid-term and terminal evaluations?** The project review criteria set forth in “GEF Project Cycle” (GEF/C.16/Inf.7) and *Monitoring and Evaluation Policies and Procedures* (2002) require projects to conduct a terminal evaluation at the time of project completion. Mid-term reviews are also encouraged so as to facilitate mid-course correction. Since all the IAs and EAs have adopted the requirement of terminal evaluations for their GEF projects and most also provide for mid-term reviews, the mention of such reviews in the project documents is more an indication of how well evaluation and review activities have been integrated into M&E plans than a signal of whether these activities will actually be conducted (it is assumed that they will). For compliance on this parameter, a project should indicate that it plans to conduct at least the terminal evaluation. Seventy-three percent of the



projects mentioned that they would conduct both a mid-term review and terminal evaluation; another 5 percent noted that they would

conduct a terminal evaluation. Thus, 78 percent of the projects are in compliance on this parameter.

## Annex F. Ratings for Terminal Evaluation Reports Reviewed, by Project

Project Name	Implementing Agency	Outcomes	Sustainability	Quality of Terminal Evaluation Report							Quality of Project M&E System			Year Effective or Prodop Signature	Year IA Prepared Terminal Evaluation Report
				Overall rating	Outcomes/Objectives	Report Consistency	Sustainability	Lessons	Actual costs and Cofinancing	M&E System	Overall Rating	Effective M&E System in Place	Information Used for Management		
Belize: Conservation and Sustainable Use of the Barrier Reef Complex	UNDP	MS	MU	S	S	S	S	S	U	S	MU	MU	U	1999	2004
Brazil: Brazilian Biodiversity Fund	WB	S	L	S	S	S	S	S	MS	S	MS	MS	MS	1996	2005
Côte d'Ivoire: Control of Exotic Aquatic Weeds in Rivers and Coastal Lagoons to Enhance and Restore Biodiversity	UNDP	MS	MU	MS	S	MS	S	S	U	MU	MS	MS	UA	1995	2004
Cuba: Priority Actions to Consolidate Biodiversity Protection in the Sabana-Camagüey Ecosystem	UNDP	S	ML	S	S	MS	S	S	HS	S	S	HS	MS	1999	2004
Democratic People's Republic of Korea: Conservation of Biodiversity in Mt. Myohyang	UNDP	S	MU	S	S	S	S	S	MS	S	MS	MS	MS	2000	2004
Ghana: Natural Resources Management	WB	MU	UA	S	MS	S	MS	HS	MS	S	MU	MU	MU	1999	2003
Guatemala: Integrated Biodiversity Protection in the Sarstun-Motagua Region	UNDP	S	ML	HS	HS	S	HS	HS	S	HS	U	U	U	1997	2004
India: Ecodevelopment	WB	MS	L	S	S	S	S	S	S	MS	S	S	S	1996	2004
Indonesia: Coral Reef Rehabilitation and Management	WB	S	ML	S	S	S	MS	S	S	MS	MS	U	HS	1998	2005
Lebanon: Strengthening of National Capacity and Grassroots In-Situ Conservation for Sustainable Biodiversity Protection	UNDP	MS	ML	MU	S	U	U	S	HU	S	HU	HU	UA	1995	2004
Nepal: Arun Valley Sustainable Resource Use and Management Pilot Demonstration Project	UNEP	S	ML	MS	MS	MS	MS	S	S	MS	MS	U	S	2001	2004
Regional: Desert Margin Program—Phase I	UNEP	MS	UA	MS	S	MS	MS	MS	MS	MS	MS	MS	MS	2002	2004
Regional: Environment and Information Management	WB	S	MU	S	MS	S	S	S	S	S	S	S	S	1998	2004
Regional: Land Use Change Analysis as an Approach for Investigating Biodiversity Loss and Land Degradation	UNEP	S	L	S	HS	S	S	S	MU	MU	MS	MS	UA	2001	2004
Regional: Reducing Biodiversity Loss at Cross-Border Sites in East Africa	UNDP	S	L	MS	S	MU	S	S	U	HS	U	U	S	1998	2004
Regional: Inventory, Evaluation and Monitoring of Botanical Diversity in Southern Africa: A Regional Capacity and Institution Building Network SABONET	UNDP	S	MU	S	MS	S	HS	MS	S	MU	MS	U	S	1998	2005
Sri Lanka: Conservation of Medicinal Plants	WB	S	L	S	S	S	HS	MS	S	S	MS	MU	S	1998	2004
Sudan: Conservation and Management of Habitats and Species, and Sustainable Community Use of Biodiversity in Dinder National Park	UNDP	MS	MU	MS	MS	MS	S	S	HU	MS	MU	MU	MU	2000	2005

Project Name	Implementing Agency	Outcomes	Sustainability	Quality of Terminal Evaluation Report							Quality of Project M&E System			Year Effective or Prod Signature	Year IA Prepared Terminal Evaluation Report
				Overall rating	Outcomes/Objectives	Report Consistency	Sustainability	Lessons	Actual costs and Cofinancing	M&E System	Overall Rating	Effective M&E System in Place	Information Used for Management		
Tanzania: Development of Jozani-Chwaka Bay National Park, Zanzibar Island	UNDP	HS	L	MS	S	MS	S	MS	HU	MS	S	MS	S	2000	2003
Vietnam: Creating Protected Areas for Resources Conservation in Vietnam Using a Landscape Ecology Approach	UNDP	S	ML	HS	HS	HS	S	HS	U	S	S	MS	HS	1999	2005
Yemen: Conservation and Sustainable Use of the Biodiversity of Socotra Archipelago	UNDP	S	ML	MS	S	MS	S	MS	MU	MU	S	MS	S	1997	2003
Chile: Reduction of Greenhouse Gases	UNDP	UA	UA	U	U	HU	U	U	MU	MU	UA	UA	UA	1995	2003
China: Efficient Industrial Boilers	WB	S	L	S	HS	S	S	S	MU	MU	UA	UA	MS	1997	2004
India: Optimizing Development of Small Hydel Resources in the Hilly Regions	UNDP	MS	ML	S	HS	HS	S	MS	HU	MS	MU	MU	U	1994	2005
Peru: Renewable Energy Systems in the Peruvian Amazon Region	UNDP	MS	MU	S	MS	S	MS	MS	HS	S	MU	MS	MU	2001	2005
Poland: Coal to Gas Conversion	WB	MS	L	S	S	S	S	S	MS	MS	S	S	S	1995	2004
Regional (Egypt, Palestinian Authority): Energy Efficiency Improvements and Greenhouse Gas Reductions	UNDP	S	ML	MS	MS	MS	S	MS	U	MU	UA	UA	S	1998	2004
Russia: Capacity Building to Reduce Key Barriers to Energy Efficiency in Russian Residential Buildings and Heat Supply	UNDP	S	ML	S	S	S	S	MU	U	S	MS	MS	MS	1997	2005
Tunisia: Barrier Removal to Encourage and Secure Market Transformation and Labeling of Refrigerators	UNDP	MS	ML	MS	MS	MS	S	MS	U	U	UA	UA	UA	1999	2004
Tunisia: Solar Water Heating	WB	S	L	S	S	S	MS	S	MS	U	UA	UA	S	1995	2004
Ukraine: Climate Change Mitigation through Energy Efficiency in Municipal District Heating, Stage 1	UNDP	S	UA	MS	S	MS	S	MS	U	U	UA	UA	UA	2002	2004
Global: Removal of Barriers to the Effective Implementation of Ballast Water Control and Management Measures in Developing Countries	UNDP	S	HL	S	S	S	S	S	S	S	MS	MS	S	2000	2005
Poland: Rural Environmental Project	WB	S	ML	S	S	S	S	S	MU	S	S	S	S	2000	2004
Regional: SAP for the International Waters of the Pacific Small Islands and Development States	UNDP	S	MU	MS	MS	MU	S	U	S	S	U	U	HU	2000	2004
Regional: Transfer of Environmentally Sound Technologies to Reduce Transboundary Pollution in the Danube River Basin	UNDP	S	L	S	HS	HS	HS	MU	MU	HS	MS	MS	MS	2001	2005
Regional: Western Indian Ocean Islands Oil Spill Contingency Planning Project	WB	S	L	S	S	S	S	S	S	S	S	S	S	1999	2004
Global: Barriers and Best Practices in Integrated Management of Mountain Ecosystems	UNEP	U	U	MU	MS	HU	MU	MU	MS	MS	U	U	N/A	2002	2004
Global: Technology Transfer Networks - Phase I: Prototype Set-Up & Testing and Phase II: Prototype Verification & Expansion (SANET)	UNEP	UA	UA	MU	MS	HU	HU	HU	MS	MS	N/A	U	N/A	2002	2003
Regional: Building Wider Public and Private Constituencies for the GEF in Latin America and the Caribbean: Regional Promotion of Global Environment Protection through the Electronic Media	UNDP	MS	UA	S	MS	S	S	S	HS	MS	U	U	U	2001	2004
Regional: Emergency Response Measures to Combat Fires in Indonesia and to Prevent Regional Haze in South East Asia	UNEP	U	MU	U	MU	U	HU	MU	HU	U	UA	UA	MS	1998	2003
Russia: Ozone Depleting Substance Consumption Phaseout	WB	S	L	S	S	S	S	S	MS	HU	UA	UA	S	1996	2004

## Annex G. Quality of Terminal Evaluation Reports, by IA and Assessment Criteria

**Table G.1: Report Presented Assessment of Relevant Outcomes and Achievement of Project Objectives**

Rating	World Bank		UNDP		UNEP	
	2004	2005	2004	2005	2004	2005
Highly satisfactory	2	1	1	4	2	1
Satisfactory	8	9	4	11	5	1
Moderately satisfactory	7	2	3	7	0	3
Subtotal	17	12	8	22	7	5
Moderately unsatisfactory	1	0	3	0	0	1
Unsatisfactory	3	0	2	1	2	0
Highly unsatisfactory	0	0	0	0	0	0
Subtotal	4	0	5	1	2	1
<b>Total</b>	<b>21</b>	<b>12</b>	<b>13</b>	<b>23</b>	<b>9</b>	<b>6</b>

**Table G.2: Report Consistent, Evidence Complete and Convincing, Ratings Substantiated**

Rating	World Bank		UNDP		UNEP	
	2004	2005	2004	2005	2004	2005
Highly satisfactory	2	0	0	3	1	0
Satisfactory	6	12	3	8	4	1
Moderately satisfactory	7	0	4	8	0	2
Subtotal	15	12	7	19	5	3
Moderately unsatisfactory	3	0	2	2	3	0
Unsatisfactory	2	0	4	1	1	1
Highly unsatisfactory	1	0	0	1	0	2
Subtotal	6	0	6	4	4	3
<b>Total</b>	<b>21</b>	<b>12</b>	<b>13</b>	<b>23</b>	<b>9</b>	<b>6</b>

**Table G.3: Report Presented Sound Assessment of Sustainability of Outcomes**

Rating	World Bank		UNDP		UNEP	
	2004	2005	2004	2005	2004	2005
Highly satisfactory	2	1	2	3	1	0
Satisfactory	6	8	4	17	3	1
Moderately satisfactory	6	3	2	1	4	2
Subtotal	14	12	8	21	8	3
Moderately unsatisfactory	6	0	3	0	0	1
Unsatisfactory	0	0	2	2	1	0
Highly unsatisfactory	1	0	0	0	0	2
Subtotal	7	0	5	2	1	3
<b>Total</b>	<b>21</b>	<b>12</b>	<b>13</b>	<b>23</b>	<b>9</b>	<b>6</b>

**Table G.4: Lessons and Recommendations Supported by Evidence Presented**

Rating	World Bank		UNDP		UNEP	
	2004	2005	2004	2005	2004	2005
Highly satisfactory	2	1	1	2	0	0
Satisfactory	11	10	6	9	5	2
Moderately satisfactory	4	1	1	8	2	1
Subtotal	17	12	8	19	7	3
Moderately unsatisfactory	2	0	3	2	1	2
Unsatisfactory	2	0	1	2	1	0
Highly unsatisfactory	0	0	1	0	0	1
Subtotal	4	0	5	4	2	3
<b>Total</b>	<b>21</b>	<b>12</b>	<b>13</b>	<b>23</b>	<b>9</b>	<b>6</b>

**Table G.5: Report Included Actual Project Costs and Cofinancing**

Rating	World Bank		UNDP		UNEP	
	2004	2005	2004	2005	2004	2005
Highly satisfactory	4	0	0	3	0	0
Satisfactory	9	5	1	4	3	1
Moderately satisfactory	6	5	4	1	1	3
Subtotal	19	10	5	8	4	4
Moderately unsatisfactory	2	2	1	3	0	1
Unsatisfactory	0	0	5	8	3	0
Highly unsatisfactory	0	0	2	4	2	1
Subtotal	2	2	8	15	5	2
<b>Total</b>	<b>21</b>	<b>12</b>	<b>13</b>	<b>23</b>	<b>9</b>	<b>6</b>

## Annex H. Management Response

---

### H.1 Introduction

We welcome the *GEF Annual Performance Report 2005*, prepared by the GEF Evaluation Office. The 2005 APR presents its findings under three main subheadings: results, processes, and monitoring and evaluation. Under these subheadings, the report provides an assessment of (1) project outcomes and sustainability, (2) delays in project completion, (3) materialization of cofinancing, and (4) quality of the M&E arrangements at the point of CEO endorsement.

According to the report, the assessment of project outcomes, project sustainability, and delays in project completion relies on an analysis of 41 projects, for which the terminal evaluations were submitted by the Implementing Agencies to the Evaluation Office in FY 2005. For assessment of the materialization of cofinancing, all the 116 terminal evaluations submitted after January 2001 were considered. Of these, 70 terminal evaluations (60 percent) provided information on actual materialization of cofinancing. The assessment of quality of the M&E arrangements at the point of CEO endorsement is based on the 74 full-sized projects that were CEO-endorsed in FY 2005.

The report acknowledges that the findings presented have several limitations due to the small number of projects for some agencies, inadequate data in some cases, reliance on self-reporting by

the agencies, and uncertainties in the process of verification of terminal evaluation reports submitted by agencies.

Nevertheless, the discussion of the issues assessed in the APR 2005 provides a series of useful insights that can contribute to portfolio management at the GEF.

### H.2 Results

We welcome the report's conclusion that most of the completed GEF projects that were assessed this year have acceptable performance in terms of outcomes and sustainability. However, the figures for performance ratings for many of these projects indicate a higher achievement level than simply having "acceptable performance." According to the report, 88 percent of the 41 GEF projects reviewed in FY 2005 were rated moderately satisfactory or above in their outcomes. In terms of the effectiveness of the use of GEF funds, 95 percent of the \$260 million allocated to the projects reviewed in FY 2005 went to projects that achieved moderately satisfactory or better outcomes. In terms of sustainability, 80 percent of the allocated GEF funds were for projects with a sustainability rating of moderately likely or better.

### H.3 Processes

We are pleased that the projects that were examined have realized almost all cofinancing prom-



ised at project inception. However, we are concerned that the exceptions to this positive picture are global projects and those in Africa.

The report states that excessive delay in project completion is associated with lower performance in terms of outcomes and sustainability. However, it also states that this association does not imply causality because excessive delay in project completion is more likely to be a symptom than an underlying cause affecting outcomes and sustainability. The report notes that the Evaluation Office will further analyze the underlying causes in other evaluations such as the Joint Evaluation of the GEF Activity Cycle and Modalities, as well as future Annual Performance Reports, to ascertain the extent and the specific forms in which project delay affects project outcomes and sustainability. We look forward to the outcome of this further analysis.

## H.4 Monitoring and Evaluation

We are encouraged by the report's conclusion that the overall quality of monitoring of GEF projects is showing signs of improvement. This is an indication that efforts made by the GEF Secretariat and the GEF partner agencies have begun to pay off. The actions taken by the agencies to address weaknesses in project monitoring systems have led to improvements. However, we acknowledge that this is an ongoing process and agree that there is still room for further improvement. As the monitoring responsibility at the portfolio level has been shifted to the Secretariat, to be undertaken in coordination with the agencies, we are working on developing a Results Management Framework for the GEF, for Council review at its December 2006 meeting, with the aim of identifying appropriate units of accountability for results, and associated tools and practices.<sup>1</sup>

---

<sup>1</sup>The policy recommendations under discussion for the fourth replenishment of the GEF Trust Fund direct the GEF

According to the APR 2005, a substantial proportion of projects did not meet the 2003 minimum M&E requirements "at entry" and would not have met the new minimum M&E requirements of the new M&E policy. As this is a technical project design issue for which the GEF agencies are accountable, there have been efforts made to address the situation have been made at this level. For example, the World Bank has been providing monitoring and results management training to its staff who are involved in reviewing as well as designing and implementing GEF projects. UNDP has also substantially improved its M&E guidance and practices. For example, in the biodiversity focal area, UNDP has issued improved guidance on indicators, monitoring and evaluation—including effective budgeting for it. UNDP also scrutinizes these M&E elements carefully in its internal review processes. UNEP has produced a number of tools, including a revised internal project review process, and checklists and guidance for staff to ensure that the M&E standards are met at entry.

The report also states that there are gaps in the present project review process and that M&E concerns are, consequently, not being adequately addressed. We agree that there exists room for improvement in reviewing M&E elements in project design. We also agree with the APR recommendation that the GEF Secretariat should redraft project review guidelines and standards to ensure compliance with the new M&E minimum requirements. In fact, this work is already under

---

Secretariat, GEF agencies, and GEF Evaluation Office to develop a common set of quantitative and qualitative indicators and tracking tools for each focal area to be used consistently in all projects with a view to facilitating aggregation of results at the country and program levels and assessment of GEF transformational impact. A complete Results Management Framework is to be developed by the GEF Secretariat and brought forward for Council consideration by the end of 2006.

way to incorporate minimum requirements for M&E more clearly into the GEF project review criteria, including provision of more guidance to Secretariat program managers for reviewing M&E design in project documentation. The STAP is also considering ways and means to enhance the contribution of STAP roster reviews during the process. In addition, the GEF Secretariat will consider modifying the Proposal Agreement Review template used for project reviews by adding a section that addresses the candor and realism of the risk assessments, as suggested by the APR 2005. However, the Secretariat will try to do this without using language that presupposes lack of honesty or transparency.

The report states that although focal area task forces are developing portfolio-level indicators and tracking tools, these tools are not yet developed enough to adequately address the need to measure project-level results. However, the report does not clearly differentiate between the level of progress made by different focal areas and the different levels of ability to measure portfolio-level results.

In this respect, we would like to highlight that the biodiversity focal area has made important strides with the portfolio monitoring system which it has developed and is currently implementing. In fact, this year, in addition to submitting project implementation reviews for individual biodiversity projects, the GEF agencies were also requested to submit tracking tools for GEF-3 projects under Strategic Priority One (Catalyzing Sustainability of Protected Area Systems) and Strategic Priority Two (Mainstreaming Biodiversity in Production Landscapes and Sectors) that were part of the PIR 2005 cohort.

The tracking tools are central to the portfolio monitoring system that has been established by the GEF Secretariat and the agencies in the bio-

diversity focal area. The system, which was developed for application at the start of GEF-3, allows for key project-level indicators to be rolled up to the level of the biodiversity portfolio in order to present a consolidated picture of portfolio-level coverage and outcomes. The portfolio monitoring system will continue to be implemented in the GEF-4 period.

Meanwhile, the international waters, climate change, persistent organic pollutants, and land degradation focal areas are similarly undertaking activities to identify program-level indicators and strategies to roll up project-level indicators to the program level. These initiatives are expected to be completed by December 2006, in congruence with the completion of the development of the GEF Results Management Framework.

We agree with the conclusion that although focal area task forces have made significant progress in developing indicators and tracking tools at the portfolio level, there remain some technical difficulties to be overcome to adequately address the need to measure and aggregate results at the portfolio level. Reflecting the APR 2005 recommendation that the Secretariat should support focal area task forces with corporate resources to develop indicators and tracking tools to measure the results of GEF operations in the various focal areas, a request is being made for a *Special Initiative for Results Management* in the FY 2007 Corporate Budget. This activity would be in line with the ongoing efforts to develop a GEF Results Management Framework.

The report asserts that the present project-at-risk systems at the partner agencies of the GEF vary greatly and may have to address issues such as insufficient frequency of observations, robustness and candor of assessments, overlap and redundancy, and independent validation of risk. While we agree that there is always room for improve-

ment, GEF agencies have been making progress in addressing most of these issues. For example, the project-at-risk system in the World Bank is already well developed, having been in place since 1996. It has been revised from time to time based on reviews undertaken by the World Bank's Quality Assurance Group. Similarly, UNDP has improved its risk management system, both in terms of reporting and central monitoring. Risk management is now conducted using the risk module in ATLAS, UNDP's corporate enterprise resource platform for project financial management. The model contributes to achieving results and impacts by allowing systematic and early project risk identification and analysis, and by facilitating risk monitoring and improving adaptive management.

We are pleased with the assessment that overall quality of terminal evaluations is improving. We agree that there are still some areas where improvements are necessary, and we expect that the FY 2006 APR will reflect a further improvement as a result of the additional measures set in place by the agencies during FY 2005.<sup>2</sup> The GEF Secretariat will work with the GEF agencies to make sure that these improvements are realized.

## H.5 Conclusions

The APR 2005 is a welcome assessment of the current status of project outcomes and sustainability, delays in project completion, materialization of

---

<sup>2</sup>Refer to the Management Action Record for the 2004 Annual Performance Report for agency actions to improve the quality of terminal evaluations.

cofinancing, and quality of the M&E arrangements at CEO endorsement, based on an analysis of terminal evaluation reports submitted by GEF agencies.

In developing a Results Management Framework, the GEF Secretariat, together with focal area task forces, aims to take into consideration the conclusions and recommendations of the APR 2005 and complement the APR exercise, in the future, with a serious effort at portfolio-level monitoring of outcomes and, if possible, impacts. The establishment of the independent GEF Evaluation Office has provided an opportunity for the GEF Secretariat to take leadership in the area of portfolio-level monitoring. It is envisioned that under the GEF Results Management Framework which will build on the results management systems already in place in the agencies, there will be a "division of labor" among GEF entities, where IAs can be responsible for project-level quality and monitoring of their respective portfolios, while the GEF Secretariat can concentrate on GEF-wide program and portfolio-level performance, strategic issues, and portfolio monitoring across agencies. Such a division of labor (repeatedly called for in various M&E studies) will also help improve quality at entry and streamline the project cycle by avoiding overlap of the review functions at project entry. The annual PIR exercise will be revised and improved in line with these goals.

The GEF Secretariat is working to advance such thinking, working with the GEF agencies, the focal area task forces, and the STAP, and to develop a GEF Results Management Framework that will reflect these principles.







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

[www.theGEF.org](http://www.theGEF.org)