



Global Environment Facility

GEF/ME/C.28/2/Rev.1
May 11, 2006

GEF Council
June 6-9, 2006

Agenda Item 4

GEF ANNUAL PERFORMANCE REPORT (2005)

(Prepared by the GEF Evaluation Office)

Recommended Council Decision

The Council, having reviewed the document GEF/ME/C.28/2 *Annual Performance Report 2005* endorses its recommendations and request that the GEF Evaluation Office report on the follow-up of the following three decisions:

- 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.
- 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 the GEF operations in the various focal areas.
- The GEF Secretariat reviewers should appraise the candor and realism of project risk assessment in the project reviews.

GEF partner agencies need to continue to follow-up on the recommendations made in last year's APR regarding the need to improve terminal evaluation reports.

TABLE OF CONTENTS

EXECUTIVE SUMMARY.....	1
I. MAIN CONCLUSIONS AND RECOMMENDATIONS.....	4
1.1 Introduction.....	4
1.2 Main Conclusions	5
A. Results.....	5
B. Processes	6
C. Monitoring and Evaluation.....	7
1.3 Recommendations	10
1.4 Issues for the Future	11
II. SCOPE AND METHODOLOGY	12
2.1 Scope	12
2.2 Methodology	13
A. RESULTS	
III. PROJECT OUTCOME AND SUSTAINABILITY	16
3.1 Approach	16
3.2 Project Outcomes	17
3.3 Sustainability of Project Outcomes.....	19
B. PROCESSES	
IV. PROCESSES AFFECTING ATTAINMENT OF PROJECT RESULTS: THE MATERIALIZATION OF CO-FINANCING AND PROJECT IMPLEMENTATION DELAYS	22
4.1 Materialization of Co-financing.....	22
4.2 Relationship Between Project Funding and Project Outcomes and Sustainability..	24
4.3 Time Lags in Implementation Completion.....	25
C. MONITORING AND EVALUATION	
V. PROJECT-AT-RISK SYSTEMS OF GEF PARTNER AGENCIES	27
5.1 Approach	27
5.2 GEF Requirements for Project Preparation and Implementation	28
5.3 Agency Reporting Systems	29
5.4 Challenges in Ensuring Compatibility Across Agencies	31
5.5 Independent Monitoring.....	33
VI. QUALITY OF PROJECT MONITORING.....	35
6.1 Quality of Monitoring During Project Implementation.....	35
6.2 Review of the Systems to Ensure Quality at Entry of M&E Arrangements.....	38

VII. QUALITY OF TERMINAL EVALUATION REPORTS	47
7.1 Approach	47
7.2 Findings	48
7.3 Difference in Ratings by Implementing Agencies and Evaluation Office	50
VIII. MANAGEMENT ACTION RECORDS	54
Appendix 1: Ratings for the Achievement of Objectives, Sustainability of Outcomes and Impacts, Quality of Terminal Evaluation Reports and Project M&E Systems	56
Appendix 2: Methodological Brief Used for the Review of Monitoring Arrangements At Entry.....	61
Appendix 3: Performance of the Portfolio on M&E Arrangements at Entry on 13 Parameters	65
Appendix 4: Agency Project-at-Risk Monitoring Inventory Card	68
Appendix 5: Project-at-Risk Inventory.....	69
Appendix 6: List of Terminal Evaluation Reports Reviewed and GEFEO Ratings.....	71
Appendix 7: Quality of Terminal Evaluation Reports by Implementing Agency and Assessment Criteria	75

Executive summary

1. This is the second Annual Performance Report (APR) that the Evaluation Office presents since the GEF Council approved the transfer of responsibility for monitoring to the implementing agencies and the GEF Secretariat. This has allowed the Office to focus more on assessing results of the GEF activities and overseeing monitoring and evaluation operations across the GEF. The higher quality of terminal evaluations submitted by the implementing agencies in FY 2005 also allowed the Office to assess the extent to which the projects are achieving their objectives. Furthermore, processes that affect project results and M&E arrangements are reported on in the APR.
2. The findings presented have several limitations. Most findings are based on the terminal evaluation reviews, which are based on the information provided by terminal evaluation reports. This introduces uncertainty into the verification process, which is mitigated by incorporating in the terminal evaluation reviews any pertinent information that has been independently gathered through other evaluations. The Office is also seeking to improve the independence of terminal evaluation reports by more fully involving the central evaluation units of the partner agencies in the process.
3. Project outcomes and sustainability is one of the topics addressed this year for the first time. A high proportion of recently terminated projects, both in terms of the number of projects and the financial resources allocated to these projects, were rated as marginally satisfactory or higher. This is in itself a positive finding, although at the moment based on a limited number of projects. A more representative assessment of the result of GEF projects will be possible as information on more projects becomes available for analysis in the coming years. It should also be mentioned that deficient project and program monitoring across the GEF system hampers efforts to aggregate results. Only by putting in place robust M&E systems at the project and program levels, will the GEF will be able to demonstrate the extent of its contributions towards addressing critical global environmental problems.
4. The APR contains the following conclusions:
 - a. Most of the completed GEF projects that were assessed this year have acceptable performance in terms of outcomes and sustainability.
 - b. Projects that were examined have realized almost all co-financing promised at the project inception, except for global projects and those in Africa.
 - c. Excessive delay in project completion is associated with lower performance in terms of outcomes and sustainability.
 - d. The quality of monitoring is showing signs of improvement. However, there is significant room for further improvement.
 - e. 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.

- f. There are gaps in the present project review process. Consequently, M&E concerns are not being adequately addressed.
 - g. 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.
 - h. Overall quality of terminal evaluations is improving. However, there are still some areas where major improvements are necessary.
5. The following recommendations are formulated:
- a. 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.
 - b. The GEF Secretariat should support Focal Area Task Forces with corporate resources to develop indicators and tracking tools to measure the results of the GEF operations in the various focal areas.
6. The Evaluation Office will issue guidelines on the minimum requirements and how they will be evaluated in the coming APRs. The Office will carry out another assessment of the M&E quality assurance systems in the coming years to follow up on the progress in implementation of the 2006 GEF M&E Policy. It will also give attention to M&E during project implementation to ensure that the GEF M&E requirements are being honored both at entry and during project execution.
7. In future assessments of the project-at-risk systems of the partner agencies, the Office will include an assessment of the actual internal reports to determine the degree of compliance with the formal procedures of the project-at-risk system design.
8. The present analysis of the links between the promised level of co-financing and outcome or sustainability is inconclusive. While the analysis of the full set of projects does show an inverse relationship between levels of co-financing and outcome or 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 co-financing could be associated with a higher risk of a project losing sight of the GEF objectives. As the number of projects with terminal evaluations increases, it will be possible to draw more robust inferences.
9. The first exercise to track the rate of adoption of Council decisions on evaluation reports through the **Management Action Records** (MARs) has been a mixed experience, which will need to be improved the next time the MARs will be presented to Council in June 2007. Differences of interpretation on how adoption should be rated caused delays on the GEF Management side, which meant that the Office received the MARs too late to verify the ratings of Management. On the basis of other evaluations and insights through

the consultative process, the Office has indicated how it perceives the rate of adoption so far. The Office is confident that with the experience gained during the process it will be possible to present verified ratings to Council in June 2007. The MARs have been published as an information document for Council (GEF/ME/C.28/Inf.2, May 2006).

10. On one rating, verification was possible. In the MAR of the 2004 APR, Management assesses as “medium” the rate of adoption of the Council decision in June 2005 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 Management Information System of the GEF. Given the evidence that the Office has gathered in the field visits of the Country Portfolio Evaluation and the Joint Evaluation of the GEF Activity Cycle and Modalities, the Office has been able to verify this assessment and it has downgraded the rate of adoption to “negligible”. Information on where projects are in the approval process is still not available in a systematic way. For project proponents on the country level, nothing has changed since the decision of Council in 2006. The Country Portfolio Evaluation in Costa Rica proposes to Council to reinforce its decision of last year – the MAR on the 2004 APR underscores the need for this reinforcement.

CHAPTER I: MAIN CONCLUSIONS AND RECOMMENDATIONS

1.1 Introduction

11. This is the second Annual Performance Report (APR) that the GEF Evaluation Office (the Office) presents since the GEF Council approved the transfer of responsibility for monitoring to the implementing agencies and the GEF Secretariat. This transfer of responsibilities has allowed the Office to focus more on assessing results of the GEF activities and overseeing monitoring and evaluation operations across the GEF. The higher quality of terminal evaluations submitted by the implementing agencies in FY2005 also allowed the Office to include in the APR an assessment of the extent to which the GEF projects are achieving their objectives. This APR presents a detailed account of some aspects of project results, of processes that may affect project results, and of monitoring and evaluation arrangements across the GEF system.

12. This is the first time that the APR includes an assessment of project outcomes, of project sustainability, of delays in project completion, of materialization of co-financing, and of quality of the M&E arrangements at the point of CEO endorsement. For the assessment of project outcomes, project sustainability and delays in project completion 41 projects were considered, for which the terminal evaluations were submitted by the Implementing Agencies to the Office in FY2005. Altogether, the GEF had invested \$260 million in these 41 projects. For assessment of the materialization of co-financing all the 116 terminal evaluations submitted after January 2001 were considered. Of these, 70 (60%) terminal evaluations provided information on actual materialization of co-financing. The GEF has altogether invested \$380 million in these 70 projects and has been able to leverage an additional amount of \$1,770 million in the form of co-financing. For assessment of quality of the M&E arrangements at the point of CEO endorsement, the 74 full size projects that were CEO endorsed in FY 2005 were considered. The GEF has altogether approved an investment of \$535 million in these projects. This APR also contains for the second time an assessment of the quality of project monitoring, and of quality of terminal evaluation reports, for which 83 terminal evaluations were considered, of which 41 were submitted in FY 2005 and 42 in FY 2004. This allowed comparisons between the performances during these two years. The GEF had invested about \$460 million in these 83 projects.

13. The findings presented have several limitations. Most findings are based on the terminal evaluation reviews, which are based on the information provided by terminal evaluation reports. This introduces uncertainty into the verification process. The Office seeks to mitigate this uncertainty by incorporating in its terminal evaluation reviews any pertinent information that has been independently gathered by the Office through other evaluations. The Office is also seeking to improve the independence of terminal evaluation reports by more fully involving the central evaluation units of the partner agencies in the process. The assessment on project-at-risk systems of the partner agencies is based on self reporting by the agencies.

14. On many issues, on which performance is being reported in the APR, information is presently available only for FY 2005. Comparisons between years will become possible in future APRs. For assessment of the quality of terminal evaluations the data is available for FY 2004 and 2005. Although this allows comparisons between the performances in these two years, it does not allow 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.

15. Project outcomes and sustainability is one of the topics addressed this year for the first time by the APR. A high proportion of the operations, both in terms of the number of projects and the financial resources allocated to these projects, were rated as marginally satisfactory or higher. This is a very positive finding. Nonetheless, a more authoritative assessment of the result 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 of outcome and sustainability, deficient project and program monitoring across the GEF system hampers efforts to aggregate results. Only by putting in place robust M&E systems at the project and program levels, will the GEF will be able to demonstrate the extent of its contributions towards addressing critical global environmental problems.

16. Council approved the procedure and format to be followed for the **Management Action Records** (MARs) concerning the rate of adoption of Council decisions on evaluation reports in November 2005. In the sections on Monitoring and Evaluation the MARs are reported on, since they show the level of learning of the GEF on the basis of evaluation reports. The MARs themselves will be posted as Information Document GEF/ME/28/Inf.2.

1.2 Main Conclusions

A. Results

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

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

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

18. *Sustainability of project outcomes.* The Office rated sustainability based on four key criteria. These are: financial resources; socio-political issues; institutional framework and governance; and, replication. The key findings are:

- Seventy six percent of the projects were rated moderately likely (ML) or above in Sustainability. Of the 23 UNDP projects that were assessed, seven (30%) were in the moderately unlikely (MU) category – just below the level where project performance could be considered acceptable. This presents an opportunity for improvement.
- In terms of GEF funds, 80% of the allocated funds were for projects with a sustainability rating of moderately likely (ML) or better.
- Among the criteria used to determine sustainability, projects tend to be the weakest in terms of financial viability.

The differences in ratings between the Implementing Agencies and the Office can be found under conclusion 8.

B. Processes

Conclusion 2: The projects that were examined have realized almost all co-financing promised at the project inception, except for global projects and those in Africa.

19. The analysis of co-financing included 116 projects for which terminal evaluation reports, completed after January 2001, had been submitted. Of these, 70 (60%) terminal evaluations provided information on actual co-financing realized. The key findings of this assessment are:

- Most of the projects achieved the co-financing promised at inception. On average, projects promised 4.4 dollars per GEF dollar and achieved 4.1 dollars per GEF dollar.
- The projects with higher promised co-financing as a percentage of GEF funds tend to meet the expected co-financing better than projects with lower promised co-financing as a percentage of GEF funds.
- Latin America and the Caribbean (LAC) region has the highest level of actual co-financing with 141 % of promised co-financing actually materializing. The lowest levels of actual co-financing as a percentage of promised co-financing are found among global projects (66%) and projects in Africa (76%).

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

20. The analysis of 41 projects reviewed by the Office in FY2005 shows that outcome and sustainability ratings tend to be lower for the projects with completion delays greater than 24 months. This association, however, 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 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.

C. Monitoring and Evaluation

Conclusion 4: The quality of monitoring is showing signs of improvement. However, there is significant room for further improvement.

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

- Compared to FY 2004, there is an improvement in the quality of project monitoring systems in FY 2005. The number of projects with MS or better rating increased from 39% in 2004 to 52% in 2005.
- The actions taken up by the Implementing Agencies (IAs) to address weaknesses in project monitoring systems have led to improvements. However, with project monitoring systems of 24% of the projects being rated as MU or worse and 20% of the terminal evaluations submitted to the Office not providing sufficient information to rate project M&E, there is still a lot of 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 new minimum M&E requirements of the new M&E policy.

22. An assessment of the compliance of the projects with the minimum M&E requirements at CEO endorsement, in which the M&E arrangements of all the 74 full size projects that were CEO Endorsed in 2005 were assessed, shows:

- Fifty eight percent of projects comply with the 2003 minimum requirements for M&E arrangements at the point of CEO Endorsement.
- Among the IAs, the UNDP projects have better ratings than the World Bank on some compliance parameters, whereas among the focal areas Climate Change projects have better ratings than Biodiversity on some compliance parameters. The differences in the ratings between the agencies are caused by the level of attention of management to M&E issues. The differences in the ratings between focal areas are caused by the level of technical difficulties encountered when monitoring.

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

23. The major gaps and weaknesses in the review process:

- At present there is insufficient guidance for the GEF Secretariat reviewers to adequately and consistently address M&E issues;
- Standards applied by the GEF Secretariat reviewers vary;

- The 2003 minimum requirements for M&E were interpreted in a variety of ways especially as regards the identification of baseline data;
- Although focal area task forces are developing project level indicators and tracking tools, these tools are not yet developed enough to adequately address the need to measure project level results;
- Focal area task forces have made significant progress in developing indicators and tracking tools, nevertheless technical difficulties have to be overcome to adequately address the needs to measure and aggregate results.

Conclusion 7: 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.

24. The assessment of the project-at-risk systems of the GEF partner agencies addresses only the issue of system design as reported by the respective agency to the Office. This assessment did not examine of the actual internal reports to determine the degree of compliance with the formal procedure. The key findings of this assessment are:

- Many IA/EAs' monitor projects-at-risk are using a 'warning flag' system which tracks self-rated project performance through a corporate Management Information System. These ratings are aggregated and rolled-up for portfolio-level reporting. The project-at risk assessment systems of the partner agencies that are development banks generally have most of the desirable characteristics, whereas others partner agencies may lack many of them
- This assessment identified the following issues:
 - Insufficient frequency of observations undermines the reporting power inherent in a Management Information System – computer reports can be generated any time, yet the underlying data are often only updated once per year;
 - It may be difficult to ensure robustness and candor of self-assessment ;
 - Managers and staff worry about proliferation of monitoring and reporting systems, overlap or redundancy, and staff reporting burdens; and,
 - Most agencies lack formal arrangements for independent validation of the self reported project-at-risk assessment. Only 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 evaluations is improving. However, there are still some areas where major improvements are necessary.

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

- Compared to FY 2004 there has been a marked improvement in the overall quality of terminal evaluations in FY 2005; especially the terminal evaluations submitted by the UNDP and the World Bank.
- A detailed assessment of the factors driving the quality of terminal evaluation reports using the Office criteria shows that Implementing Agencies are addressing most of the key quality issues that were identified last year.
- There is little difference in outcomes and sustainability ratings given by the Evaluation Office and by the Implementing Agencies when a binary scale is used.¹ When comparing the ratings on the six point scale, while there is no difference between the Office ratings and the World Bank's Independent Evaluation Group ratings, UNEP tends to rate its projects a point higher than the Office. Since many of the terminal evaluations submitted by UNDP did not provide outcomes and sustainability ratings robust inferences cannot be drawn about the overall reliability of the ratings of its terminal evaluations.
- The terminal evaluations still continue to be weak in assessing the quality of monitoring (especially terminal evaluations from the Climate Change Focal area). They also frequently fail to report on the actual costs including the total costs, a breakdown per activity of the GEF financing and co-financing by other sources. Thus, despite improvement in the overall quality of the terminal evaluations submitted by the Implementing Agencies there are gaps in the provided information.

Management Action Records

26. The first exercise to track the rate of adoption of Council decisions on evaluation reports through the **Management Action Records** (MARs) has been a mixed experience, which will need to be improved the next time the MARs will be presented to Council in June 2007. Differences of interpretation on how adoption should be rated caused delays on the GEF Management side, which meant that the Office received the MARs too late to verify the ratings of Management. On the basis of other evaluations and insights through the consultative process, the Office has indicated how it perceives the rate of adoption so far. The Office is confident that with the experience gained during the process it will be possible to present verified ratings to Council in June 2007. The MARs have been published as an information document for Council (GEF/ME/C.28/Inf.2, May 2006).

27. On one rating, verification was possible. In the MAR of the 2004 APR, Management assesses as “medium” the rate of adoption of the Council decision in June 2005 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 Management Information System of the GEF. Given the evidence that the Office has gathered in the field visits of the Country Portfolio Evaluation and the Joint Evaluation of the GEF Activity Cycle and Modalities, the Office has been able to

¹ By classifying observations with moderately satisfactory or better ratings as an acceptable level of performance and observations with moderately unsatisfactory or worse ratings as an unacceptable level of performance, the six point scale can be converted into a binary scale.

verify this assessment and it has downgraded the rate of adoption to “negligible”. Information on where projects are in the approval process is still not available in a systematic way. For project proponents on the country level, nothing has changed since the decision of Council in 2005. The Country Portfolio Evaluation in Costa Rica proposes to Council to reinforce its decision of last year – 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.

28. Lack of guidance has been identified as a problem causing reviewers to apply their own perspective rather than a common concern to meet the minimum requirements. Considerations should also be given to a more clearly defined role for STAP roster reviewers in the assessment of scientific and technical aspects of project indicators.

29. The GEF Secretariat should modify the ‘Proposal Agreement Review’ template used for project reviews by adding a separate section for “Candor and Realism of the Risk Assessment.” This will ensure that any risk related issue that is flagged in any stage of the review is followed-up on during the 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 the GEF operations in the various focal areas.

30. In recent years Focal Area Task Forces have taken up various actions to develop tools that are necessary to measure the environmental results of the GEF Operations. However, there is still a need for intensification of the present efforts for further development of tools such as indicators and tracking tools. This will require corporate investments to address the technical challenges specific to each focal area, to build consensus on indicators, to define ways to roll-up results at the portfolio level and to find out ways to address issues related to attribution.

31. The on-going work of the Implementing and Executing Agencies to improve the quality of terminal evaluations should continue. The terminal evaluations provided by the implementing agencies still have major information gaps. They are weak in terms of assessing project monitoring systems, and in reporting the actual project costs including the total costs, a breakdown per activity of GEF funding and co-funding. UNDP needs to fully engage its central evaluation group in to the process and UNEP needs to further enhance the independence of their central evaluation group to improve the quality of terminal evaluations and address differences in ratings. Progress in this area will be tracked through the Management Action Record of the previous APR and in assessments of future APRs.

1.4 Issues for the Future

32. The findings on the quality of M&E arrangements at entry confirm the importance of the new minimum requirements for M&E of the new M&E policy. The new policy asks projects to provide adequate baseline information on indicators at the point of work program inclusion in all cases, barring exceptions. The Office will issue guidelines on the minimum requirements and how they will be evaluated in the coming APRs. The Office will carry out another assessment of the M&E quality assurance systems in the coming years to follow up on the progress in implementation of the 2006 GEF M&E Policy. It will also give attention to M&E during project implementation to ensure that the GEF M&E requirements are being honored both at entry and during project execution.

33. The review of the project-at-risk monitoring systems of the partner agencies of the GEF shows that there is a need to enhance accounting and validation in existing agency framework. This is particularly the case for those frameworks that depend almost exclusively on self assessment by management, most prevalent among the partner agencies other than the 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 Office will include an assessment of the actual internal reports to determine the degree of compliance with the formal procedures of the project-at-risk system design.

34. The present analysis of the links between the promised level of co-financing and outcome or sustainability is inconclusive. While the analysis of the full set of projects does show an inverse relationship between levels of co-financing and outcome or 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 co-financing could be associated with a higher risk of a project losing sight of the GEF objectives. As the number of projects with terminal evaluations increases, it will be possible to draw more robust inferences.

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

CHAPTER II: SCOPE AND METHODOLOGY

2.1 Scope

36. The APR provides an annual presentation of the performance of the completed projects of the GEF, 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, whereas others could be 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. This overview consists of the Office's assessment of the extent to which the completed projects, for which the terminal evaluations were submitted in FY 2005, achieved expected outcomes and sustainability of outcomes. The APR will continue to report annually on attainment of objectives and outcomes.
- An analysis of the materialization of project co-financing by region and Implementing Agency along with an analysis of the links between project co-financing and project outcome and sustainability. The Office will continue to report on these issues on an annual basis.
- An analysis of correlation between project implementation delays and project outcomes and sustainability. The Office will continue to report on these issues on an annual basis.
- An assessment of the quality of project monitoring, which involves an examination of 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 continue to annually report the quality of project monitoring at completion. Reporting on the quality assurance systems for project M&E arrangements at CEO endorsement will be done biennially.
- An inventory of the present risk monitoring practices of the GEF Implementing and Executing Agencies, which reports on the approaches used by the GEF partner agencies to track risk. This inventory also identifies the strengths and weaknesses of the current risk monitoring systems of the partner agencies. The APR will report on the risk monitoring systems of the partner agencies biennially.
- An assessment of the quality of terminal evaluation reports submitted by the Implementing Agencies to the Office in FY 2005. This assessment is presented annually and it provides information broken down by focal area and implementing agency. This year APR also reports on quality of the terminal evaluations reporting on M&E during implementation.

2.2 Methodology

37. Project terminal evaluation reports submitted by Implementing and Executing Agencies to the Office form the core information source for a large portion of the APR, particularly for the topics that are reported annually. For this reason ensuring the reliability of terminal evaluation reports is critical. The Office seeks to assess and strengthen its reliability in several ways.

38. The Office reviews terminal evaluation reports to determine the extent to which reports address all the objectives and outcomes promised in the project document, to evaluate the reports' internal consistency and to verify that ratings are properly substantiated by the evaluation's findings. Terminal evaluation reports are reviewed by the Office staff using a set of detailed guidelines to ensure that uniform criteria are used by the reviewers during the review process (see Appendix 1 for details). When deemed appropriate, a reviewer may propose to up grade or down grade project ratings in the terminal evaluation report. The reviews and the proposed ratings modifications are subsequently examined by the senior evaluation officer in the respective focal area who confirms or rejects the initial reviewer's conclusions and ratings. When projects are down rated below moderately satisfactory for outcomes or below moderately likely for sustainability, a second senior evaluation officer in the Office also examines the review to ensure that the new ratings are justified. When terminal evaluation reports provide insufficient information to make an assessment or verify the Implementing Agency ratings on outcomes, sustainability or quality of project M&E systems, the Office classifies the projects as "Unable to Assess" and excludes it from any further analysis on the respective dimension.

39. The review process described above has several limitations. It is ultimately based on the information provided by terminal evaluation reports. Full verification of these reports could probably be ascertained through field verification. The Office seeks to mitigate this uncertainty by incorporating in its terminal evaluation reviews any pertinent information that has been independently gathered by the Office as part of other evaluations. The Office will test several approaches to targeted field verification. For example, setting aside time for field verification of projects during country visits carried out in the context of other thematic evaluations. The Office will also carry out full field evaluations when the findings of the terminal evaluation review or targeted field verification for a project deem an independent evaluation necessary.

40. Another way to address the reliability concerns pertaining to terminal evaluations is to work with GEF partner agencies to more fully engage the 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 Office. The Independent Evaluation Group (IEG) of the World Bank conducts desk reviews and verification of all implementation completion reports, which are produced by management. IEG also carries out field verifications for 25% of the World Bank operations. The Office has monitored IEG ratings over the last two years and has found only minor differences in ratings given by the Office and by the IEG. Therefore, the Office will use IEG's validation of terminal evaluation reports and, where necessary, will

complement these with a relatively minor effort to address the GEF specific information needs. During FY 2005 UNDP and UNEP took steps to more directly involve their central evaluation groups in the evaluation of GEF projects. In the case of UNEP steps were taken to strengthen the independence of its central evaluation group. The Office will continue its dialogue with GEF partner agencies. Meanwhile, the Office will continue to review terminal evaluation reports and verify their ratings.

41. An important issue related to the reliability of terminal evaluation reviews as a major source of information to the APR is whether the Office is able to access the terminal evaluations of completed projects in a timely manner. To ensure this, the Office has put in place a system to track the submission of terminal evaluations. The Office has created a database of terminal evaluations expected in any given year. Information on this database is sent to the Implementing Agencies for verification. Agencies are expected to submit terminal evaluations for completed projects or new dates for terminal evaluations for extended projects. This tracking system includes all GEF projects with an original completion since January 2001. This pertains to the majority of GEF projects. Starting FY 2005 the Office is also keeping track of the time between completion of project implementation and submission of terminal evaluations and between terminal evaluation completion and submission. The analysis of the 2005 data shows that on average terminal evaluations were received by the Office 7.8 months after their completion and 10.5 months after completion of project implementation. This average is well within the 12 months limit set in the new GEF M&E Policy. However, 11 terminal evaluations (27% of the total), had been submitted a year after completion of project implementation with the breakdown being: 3 of 12 for the World Bank, 7 of 23 for the UNDP and 1 of 6 for the UNEP.

42. Special reviews were carried out on the systems of quality assurance for M&E arrangements at CEO endorsement and on the status of project-at risk systems. Appendix 2 and 3 provide a description of the methodology used in these assessments.

43. The 2005 APR presents an indicative picture of how the projects whose terminal evaluations were reviewed in 2005 performed. The 'F' test and Chi Square test were used to assess differences between groups of projects and the findings reported as significant are at 90% or higher confidence level. The regression analysis was used to assess magnitude and direction of change associated with different variables. Nonetheless, the information obtained so far place some important limitations to 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 times during implementation, the numbers are still relatively small to draw firm conclusions. In the case of the assessment of project monitoring at completion a large proportion of terminal evaluations failed to provide sufficient information, so a significant proportion of those projects are not included in the analysis. Data for two years do not permit the Office to infer trends. These limitations will diminish in the coming years as implementing and executing agencies submit more terminal evaluation reports that comply with the GEF Terminal Evaluation Guidelines. As the GEF project portfolio matures, an increasing number of terminal evaluation reports will also permit a more in-depth analysis. Larger

and more reliable data sets will allow the Office to meaningfully assess progress and to make comparisons among agencies and focal areas.

44. The preliminary findings of this report were presented and discussed on various occasions with the GEF Secretariat, and Implementing and Executing Agencies, including focal area task forces meetings that took place during November and December 2005 and the Interagency Meeting held in Washington, DC, in January 2006. Individual reviews of project terminal evaluations and the results for quality of projects M&E at entry were also shared with the implementing agencies and the GEF Secretariat for factual verification.

RESULTS

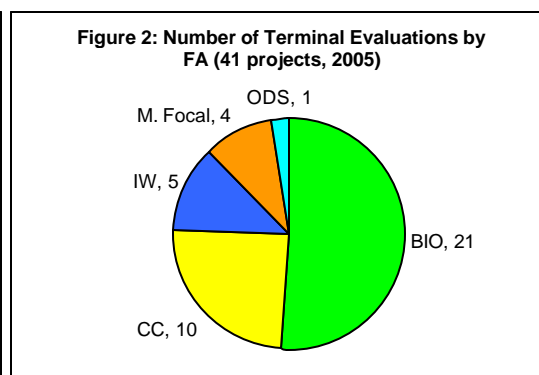
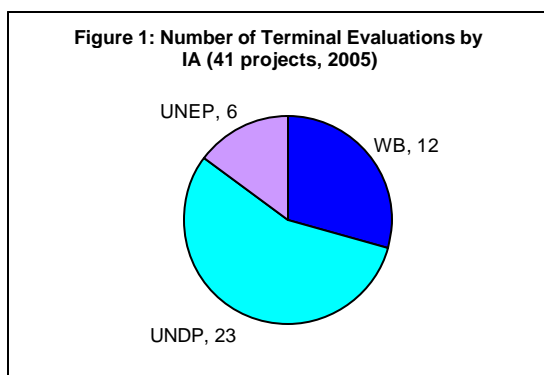
CHAPTER III: PROJECT OUTCOME AND SUSTAINABILITY

45. This chapter discusses verified ratings on outcomes and sustainability of the 41 projects for which terminal evaluations were submitted in FY 2005. Since this is the first year the GEF Evaluation Office rates outcomes and sustainability there is no baseline for comparison. Most GEF projects assessed this year seem to have for the most part attained their objectives. This is particularly true for World Bank and UNDP projects in all focal areas with an exception of multi focal projects. This year's analysis also suggests that UNDP and UNEP need to give more attention to ensuring project outcome sustainability.

46. The differences in ratings between the Implementing Agencies and the Evaluation Office are discussed in chapter 7.

3.1 Approach

47. The Office rated the project outcomes based on the level of achievement of the project objectives and expected outcomes. The Office rated sustainability based on a set of key criteria that contribute to sustainability such as financial resources, socio-political issues, institutional frameworks and governance, and replication. Of the 41 projects, there were five evaluations for which the office was unable to rate sustainability based on the information provided in the terminal evaluations. These also included two projects for which the Office had been unable to give the outcomes ratings. These terminal evaluations were excluded from the analysis. The distribution of terminal evaluations reviewed this year by Focal Area and Implementing Agency is presented in figures 1 and 2.



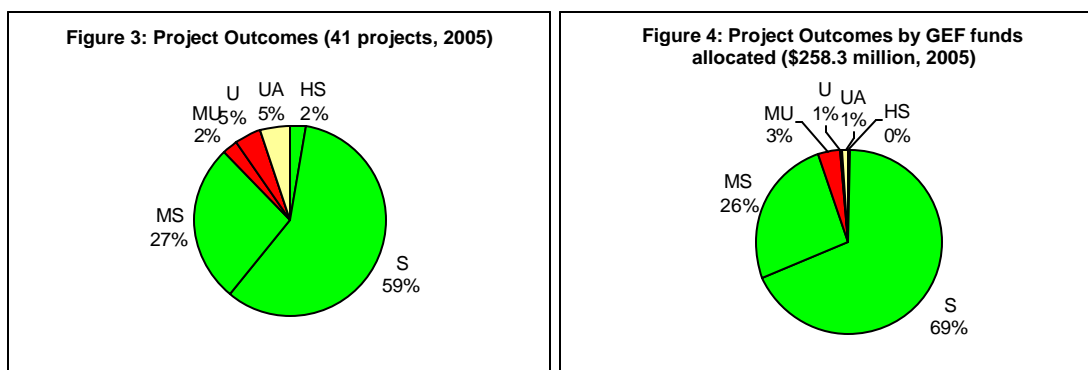
Box 1: The Rating Scales

The GEF Evaluation Office has used a six point scale for rating outcomes, sustainability, and quality of terminal evaluations. The six-point rating 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), and Highly Unsatisfactory (HU) or Highly Unlikely (HU). Unable to assess

(UA) has been used when the information contained in the terminal evaluation did not allow an assessment.

3.2 Project Outcomes

48. Most projects (88 % of 41 projects) with terminal evaluations reviewed in 2005 were rated MS (see Box 1) or above in their outcomes (see figure 3). Similarly, most GEF funds allocated to these projects (95%) were rated as MS or above in their outcomes. Only a small percentage (5% of \$260 million) of the GEF funds allocated for projects reviewed in 2005 were for projects rated below MS in their outcomes (Figure 4).



Outcomes by Implementing Agency and Focal Area

49. As indicated in Table 1, most projects for which the World Bank and UNDP submitted terminal evaluations in FY 2005 were rated as moderately satisfactory or above. As indicated in Table 2, multi-focal projects have the lowest outcome ratings. However, there were only 4 of these projects this year. More information about these projects is provided below in the last paragraph of this section. The GEF Evaluation Office will continue to assess outcomes of UNEP and multi-focal projects in future years to determine trends.

Table 1: Project outcomes by IA (Number of projects)

Rating	WB	UNDP	UNEP	Total
HS	0	1	0	1
S	9	13	2	24
MS	2	8	1	11
<i>Subtotal</i>	<i>11</i>	<i>22</i>	<i>3</i>	<i>36</i>
MU	1	0	0	1
U	0	0	2	2
HU	0	0	0	0
<i>Subtotal</i>	<i>1</i>	<i>0</i>	<i>2</i>	<i>3</i>
UA	0	1	1	2
Total	12	23	6	41

Table 2. Project outcomes by FA (Number of projects)

Rating	BD	CC	IW	MF	ODS
HS	1	0	0	0	0
S	13	5	5	0	1
MS	6	4	0	1	0
<i>Subtotal</i>	<i>20</i>	<i>9</i>	<i>5</i>	<i>1</i>	<i>1</i>
MU	1	0	0	0	0
U	0	0	0	2	0
HU	0	0	0	0	0
<i>Subtotal</i>	<i>1</i>	<i>0</i>	<i>0</i>	<i>2</i>	<i>0</i>
UA	0	1	0	1	0
Total	21	10	5	4	1

50. In terms of the use of GEF funds, the bulk of the funds for World Bank (95 %) and UNDP (98%) projects were used by projects rated satisfactory (S) in their outcomes (Table 3). The bulk of GEF funds in all focal areas were also allocated for projects rated satisfactory in their outcomes except multi-focal (Table 4).

**Table 3: Project outcomes by IA
(GEF investment in million \$)**

Rating	WB	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
<i>Subtotal</i>	<i>151.5</i>	<i>87.1</i>	<i>6.4</i>	<i>244.9</i>
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
<i>Subtotal</i>	<i>8.7</i>	<i>0.0</i>	<i>1.7</i>	<i>10.4</i>
UA	0.0	1.7	1.3	3.0
Total	160.2	88.8	9.4	258.3

**Table 4: Project outcomes by FA
(GEF investment in million \$)**

Rating	BD	CC	IW	MF	ODS
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
<i>Subtotal</i>	<i>105.5</i>	<i>76.9</i>	<i>26.5</i>	<i>1.0</i>	<i>35.0</i>
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
<i>Subtotal</i>	<i>8.7</i>	<i>0.0</i>	<i>0.0</i>	<i>1.7</i>	<i>0.0</i>
UA	0.0	1.7	0.0	1.3	0.0
Total	114.2	78.6	26.5	4.0	35.0

51. Projects with highly satisfactory outcomes successfully achieved the project objectives and expected outcomes. For example, in the *Tanzania - Jozani-Chwaka Bay Conservation Project* (UNDP), strong government and community ownership was a key factor in the success. The project contributed to the legal establishment of the national park and effectively addressed the key threats to biodiversity conservation that were identified in the project document such as resource conflicts with communities. According to the terminal evaluation, as a result of the project, Jozani communities are now involved in the management and conservation of resources and 2500 villagers benefited from savings and credit (micro-finance) schemes to develop tourism related income generation activities. The census data showed that the project interventions have contained and reversed the decline in the population of Red Colobus Monkey and reduced encroachment into the Jozani Forest.

52. In the *Tunisia - Solar Water Heating* (World Bank) project, adaptive management and flexibility during implementation was pivotal to success. The project properly responded to the change in the market conditions by refocusing the intervention to the needs of the residential sector as opposed to the commercial and public sector as had been initially planned. This change was made because there was limited demand from commercial and public sectors and an unexpectedly high demand from the household sector. The modified project remained clearly directed at the underlying purpose of the grant: 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.

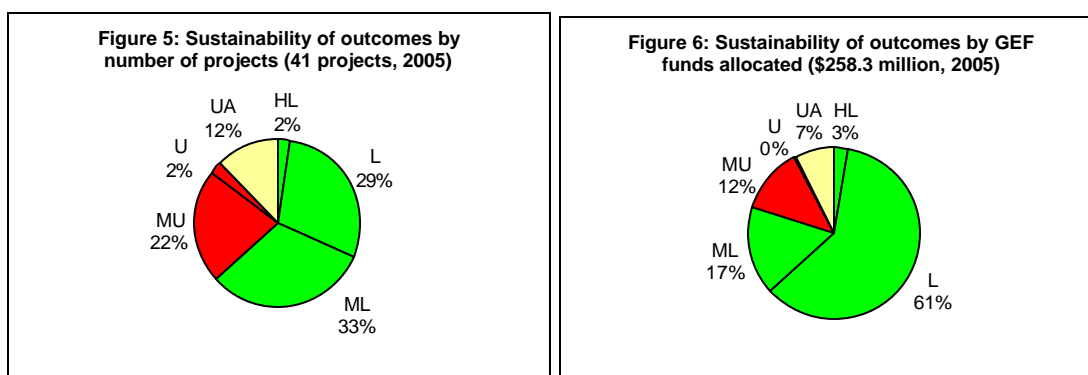
53. Projects with unsatisfactory outcomes did not achieve their objectives for multiple reasons such as, for example, the proposed solutions not addressing the underlying problems, adaptive management or solutions not done until late in the project implementation, or poor implementation management and resource use. This was so in the case of *Regional - Emergency Response Measures to Combat Fires in Indonesia and to Prevent Regional Haze in South East Asia* (UNEP) and *Global - Barriers and Best Practices in Integrated Management of Mountain Ecosystems* (UNEP) project. The first project had some project design weaknesses because it focused mostly on strengthening

fire fighting capacity while realizing later that fire related issues were more complex than shifting agriculture and weather and involved commercial companies, land use changes and climate issues. The second was over ambitious in the 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

54. The review of terminal evaluations in 2005 (in all focal areas) included rating sustainability² based on an assessment of key contributing aspects and any risks that could undermine the continuation of the benefits at the time of the terminal evaluation. The following four aspects of sustainability were addressed: financial, socio-political, institutional frameworks and governance, and replication³. Appendix 1 provides a breakdown of the questions to assess each of the aspects of sustainability.

55. The sustainability of outcomes by number of projects and GEF funds is presented in Figures 5 and 6:



56. Although the overall performance of projects in terms of sustainability is not as good as that for outcomes, still at least 64% of the projects and at least 81% of the GEF funds allocated resulted in outcomes with a sustainability rating of Moderately Likely (ML) or better.

Sustainability of Outcomes by Implementing Agency and Focal Area

57. Among the Implementing Agencies, most World Bank projects had a Likely rating in sustainability (Table 5). Two thirds of the UNDP projects had a ML or better rating in sustainability of outcomes; the remaining third had MU rating. UNEP had equal number of projects rated ML or better and as MU or worse. The terminal evaluations of two projects from UNEP did not provide enough information on this dimension and therefore the Office was unable to rate the sustainability of outcomes for them.

² Sustainability will be understood as the likelihood of continuation of project benefits after completion of project implementation. GEF Project Cycle. GEF/C.16/Inf.7. October 5, 2000.

³ 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, etc. GEF Project Cycle, GEF/C.16/Inf.7, October 5, 2000.

58. Among focal areas, 62% of biodiversity projects were rated ML or better in the sustainability of their project outcomes (Table 6). The outcome sustainability of most projects from other focal areas, except for multi-focal projects, was also rated as ML or better. Since the present analysis is based on the 41 Terminal Evaluations reviewed in FY 2005, the picture portrayed by it may not be as representative as would be desirable. However, as the number of reviewed projects increases a more representative assessment of sustainability will be possible.

Table 5: Outcomes sustainability by IA (Number of projects)

Rating	WB	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
Total	12	23	6	41

Table 6: Outcomes sustainability by FA (Number of projects)

Rating	BD	CC	IW	MF	ODS
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
Total	21	10	5	4	1

59. Regarding the use of the GEF funds, as mentioned previously, the proportions of funds assigned to projects with sustainability ratings of Moderately Likely (ML) or above seem to follow the distribution of the number of projects. Tables 7 and 8 show the allocations and ratings by Implementing Agencies and Focal Areas.

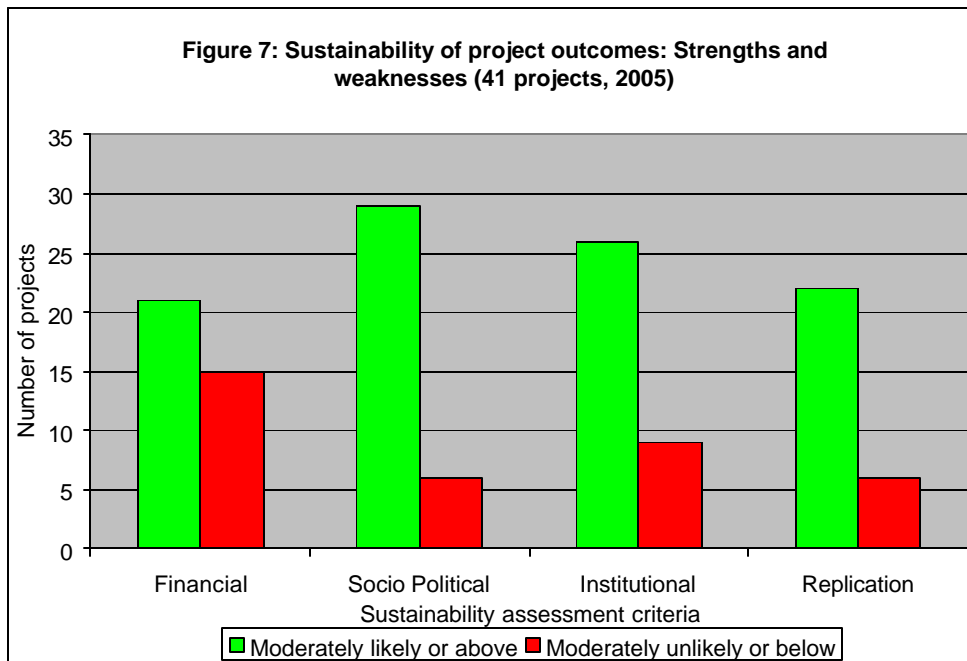
Table 7: Outcomes sustainability by IA (GEF investment in million \$)

Rating	WB	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
Total	160.2	88.8	9.4	258.3

Table 8: Outcomes sustainability by FA (GEF Investment in million \$)

Rating	BD	CC	IW	MF	ODS
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
Total	114.2	78.6	26.5	4.0	35.0

60. A closer examination of the factors that determine sustainability of projects outcomes, as indicated in Figure 7, reveals that financial viability seems to be an area where many projects are lagging. Financial sustainability refers to financial and economic resources available to allow for the project outcomes to be sustained once the GEF assistance ends and the risk that these resources will not materialize.



61. Two aspects that were found in some projects with Likely (L) or Highly Likely (HL) sustainability and that helped increase financial sustainability was the implementation of regulations that have contributed to increase in demand for services promoted by the GEF project and enhanced private sector involvement. Some examples are the *Regional - Transfer of Environmentally Sound Technologies (TEST) to Reduce Transboundary Pollution in the Danube River Basin* (UNDP) where the private sector (i.e. industries) in the region is increasing its demand for environmentally sustainable technologies and for formally-recognized and accredited cleaner production techniques as the countries have to meet more stringent EU environmental standards. The *Global - Removal of Barriers to the Effective Implementation of Ballast Water Control and Management Measures in Developing Countries* (UNDP) is another example of a project that has been assessed as having good prospects for sustainability increased by a strong interest of the private sector. Also the *Ozone Depleting Substances Consumption Phase-out Project* in Russia (World Bank) created the competitive capacity to attract debt and equity investment to comply with the regulatory framework for the proactive management of ozone depleting substances issues consistent with the international practice. Another example from the Biodiversity focal area is the *Jozani-Chwaka Bay Conservation Project* in Tanzania (UNDP). The project has a tourism revenue sharing scheme in place which partially covers management costs of the Jozani National Park. Furthermore, it does provide funds to the communities for small development projects (water, sanitation etc). Although financial sustainability is dependent on tourism at present, 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 then it is likely that there will be a sustainable flow of resources to cover some managerial and community costs.

62. Projects with moderately unlikely or worse sustainability ratings include the *Regional - Emergency Response Measures to Combat Fires in Indonesia and to Prevent Regional Haze in South East Asia* (UNEP), for which many of the project follow-up activities require further funding which may not be forthcoming and co-operation at national and international level has been weak. The *Global - Barriers and Best Practices in Integrated Management of Mountain Ecosystems* (UNEP) project also lacked financial and institutional sustainability.

PROCESSES

CHAPTER IV: PROCESSES AFFECTING ATTAINMENT OF PROJECT RESULTS: THE MATERIALIZATION OF CO-FINANCING AND PROJECT IMPLEMENTATION DELAYS

63. The specific topics addressed in this chapter vary from year to year. This year two topics have been addressed. One topic is co-financing, including ‘materialization of project co-financing’ and ‘links between project co-financing, and project outcome and sustainability.’ The analysis concludes that for the most part the GEF as a whole tends to achieve the co-financing that was promised at project inception. The analysis of the links between the levels of co-financing 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 issue. The underlying issue that would be addressed through such scrutiny is whether there is a point of co-financing beyond which the risk of the GEF losing leverage increases the chance of compromising the achievement of GEF objectives.

64. The second topic addressed in this chapter is the analysis of the time lags between expected and actual project closing dates and its implications for project outcomes and sustainability. This 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 the actual closing date. This finding is significant for this group of projects because 18 (44%) out of the 41 projects included in the analysis had implementation completion delays of more than two years. This finding suggests that closer attention to factors affecting project implementation delays during midterm reviews might permit early detection and correction of factors affecting outcomes and their sustainability.

4.1 Materialization of Co-financing

Approach

65. This section intends to assess whether promised co-financing has materialized. Tracking this indicator is important because project activities are budgeted with the expectation of promised co-financing materializing. For this analysis, all 116 terminal evaluations reviewed since 2001 were examined. However, the IAs provided actual co-financing data only for 70 projects (60%). Therefore, this limits the extent to which inferences regarding the potential factors affecting materialization of co-financing could be made. As the quality of terminal evaluations improve and these reports disclose the actual projects costs, a more fuller and representative picture could be presented.

Total Promised and Actual Co-financing

66. The total co-financing promised and the total co-financing approved was calculated by adding these for all 70 projects. Each of these two totals was divided by the total GEF funding for the 70 projects. An analysis of the data provided by these 70

terminal evaluation reports indicates that promised co-financing tends to materialize for the most part for these projects (Table 9).

Table 9: Actual Compared to Promised Co-financing for Projects with Terminal Evaluation Reports Reviewed from 2002 to 2005⁴

Total Co-financing Promised (in million \$)	1,900
Total Co-financing Realized (in million \$)	1,770
Average \$ Promised for Every GEF \$ Approved	4.4
Average \$ Realized for Every GEF \$ Approved	4.1

Promised and actual Co-financing by Implementing Agency, Focal Area and Region

67. A more in depth analysis by Implementing Agency, Focal Area and geographical region provides information on some differences between those that can improve planning. East Asia and the Pacific together with South Asia seem to have the highest ratios of promised co-financing per GEF dollars approved (Table 10). In the case of East Asia and the Pacific, this was driven by three very large blended World Bank climate change projects 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; World Bank was the Implementing Agency for all these three projects. The available information indicates that global projects could have the lowest ratios of promised as well as actual Co-financing (as a percentage of the promised co-financing).

68. Interestingly, while the promised co-financing for Latin America and the Caribbean seem to be equal to the amount of GEF funds assigned for projects there, the region has the highest levels of actual co-financing with 141% of promised co-financing actually materializing. This was driven mainly by some large conservation projects in Mexico, Brazil, and Bolivia, but most projects in Latin America and the Caribbean attracted more co-financing than expected.

Table 10: Co-financing Ratios and Actual Co-financing by Region

Region ⁵	Total number of terminal evaluations considered	Total number of terminal evaluations providing actual cost information	Promised \$ of Co-financing per GEF Approved \$	Percentage of Promised Co-financing Actually Materializing
AFR	30	17	2.5	76
EAP	17	14	6.8	91
ECA	19	9	1.0	93
GLO	11	4	0.8	66
LAC	24	17	1.0	141
MENA	9	5	1.4	98
SA	6	4	5.8	98

⁴ Total co-financing promised and realized was the sum of these values for the 70 out of 116 projects that provided actual information in the terminal evaluations. These two values were then divided by the total GEF funds approved for these 70 projects to calculate the average funds (US\$) promised for every GEF dollar approved.

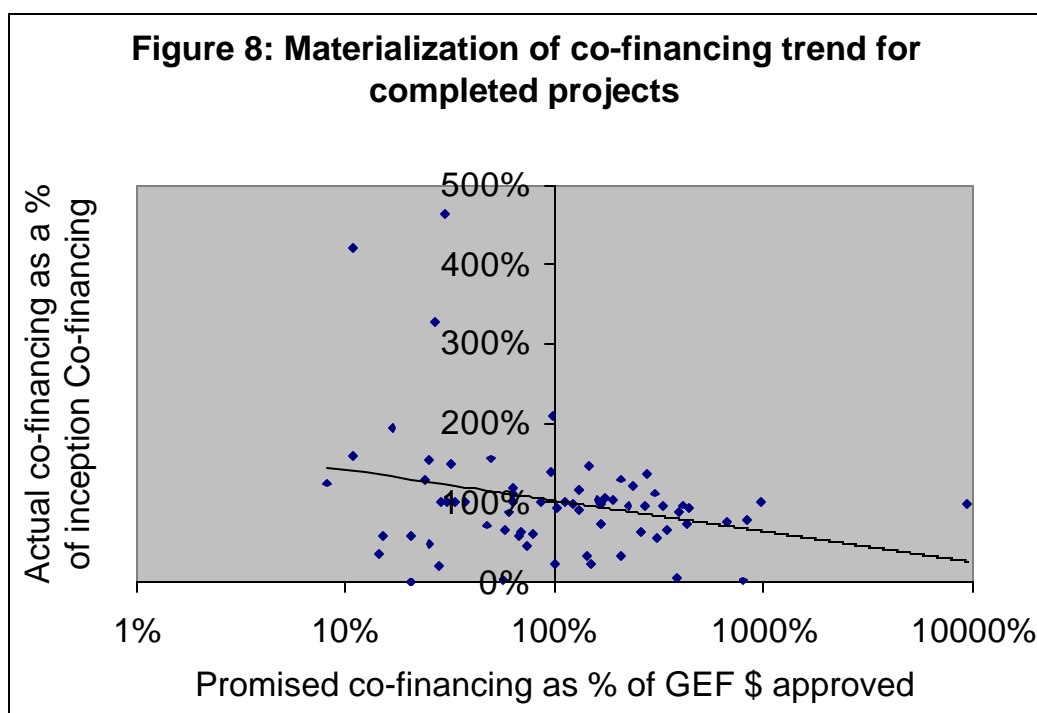
⁵ Regions shown are: Africa (AFR), East Asia and the Pacific (EAP), Europe and Central Asia (ECA), Global (GLO), Latin America and the Caribbean (LAC), Middle East and Northern Africa (MENA), and South Asia.

69. Among Implementing Agencies, UNEP projects that provided information on actual costs had the lowest co-financing ratio and also the lowest percentage of co-financing actually materializing (Table 11). Of the seven UNEP projects providing actual cost information, three were global projects and two were in Africa, the two regions with the lowest average co-financing actually materializing.

Table 11: Co-financing Ratios and Actual Co-financing by IA and FA

Particulars	Total number of terminal evaluations considered	No. of terminal evaluations Providing Actual Costs Information	Promised \$ of Co-financing per GEF Approved \$	Percentage of Promised Co-financing Actually Materializing
WB	50	39	5.0	93
UNDP	50	25	2.5	83
UNEP	19	7	1.8	48
BIO	61	41	1.5	89
CC	29	17	10.0	94
IW	16	7	1.8	94

70. Among the projects reviewed those with larger promised co-financing as a percentage of GEF funds tend to meet the expected co-financing better than projects with smaller promised co-financing as a percentage of GEF funds (Figure 8). This relationship holds true regardless of the project size (i.e., Medium Size Projects or Full Size Projects), and holds true for World Bank, and UNDP as well as for Bio and CC projects. For IW and UNEP the sample was not large enough to draw conclusions.



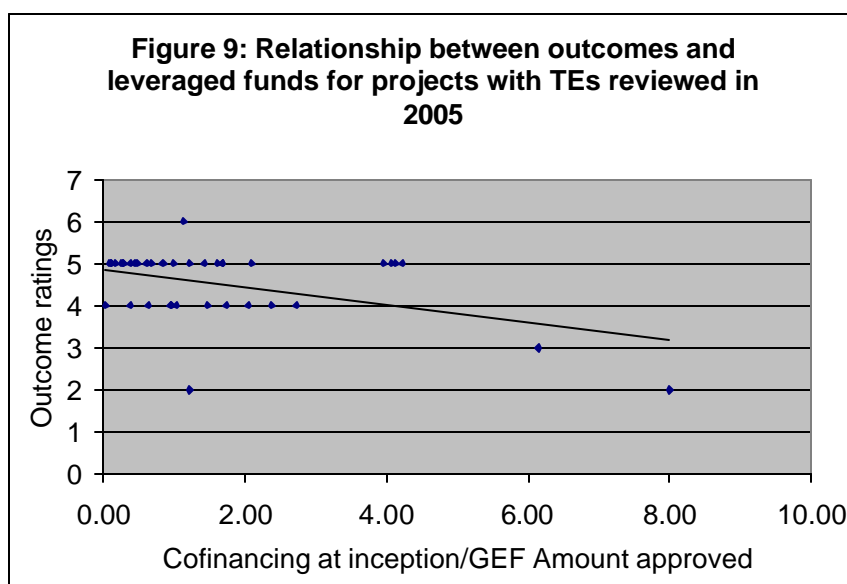
4.2 Relationship Between Project Funding and Project Outcomes and Sustainability

71. The Office examined the relationship between project funding, outcomes and sustainability to assess whether projects with larger GEF funding and co-financing produced better results. This analysis is based on the 41 projects whose terminal evaluations were reviewed in 2005. Linear regressions were carried out with GEF funds as independent and the project outcomes and sustainability ratings as dependent variable. Another linear regression was carried out with the ratio of promised co-financing and GEF funds (i.e. co-financing ratio) as an independent variable and project outcomes and sustainability ratings as a dependent variable.

72. According to the analysis, there is little correlation between GEF funding and outcomes, and GEF funding and sustainability for the terminal evaluations reviewed in 2005. In other words, projects reviewed in 2005 seem to get enough GEF funds to carry out the activities that they promise with some achieving their objectives more successfully than others regardless of the amount of GEF funding received.

73. Regarding promised co-financing, the analysis showed that the more leveraged the project was (i.e. the higher the ratio of promised co-financing to GEF funds) the lower the outcome ratings tended to be (see figure 9). This correlation is driven by two outliers that have the highest co-financing ratios and have outcome ratings MU or worse. When these two outliers are eliminated, then there is no correlation between co-financing ratio and outcomes. The two outliers are the Ghana - 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).

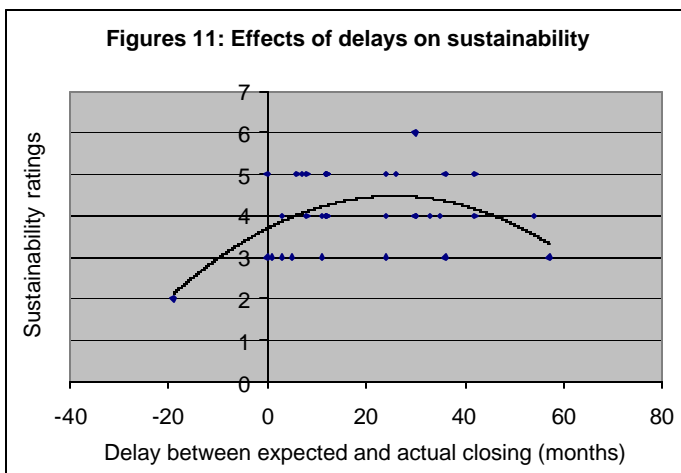
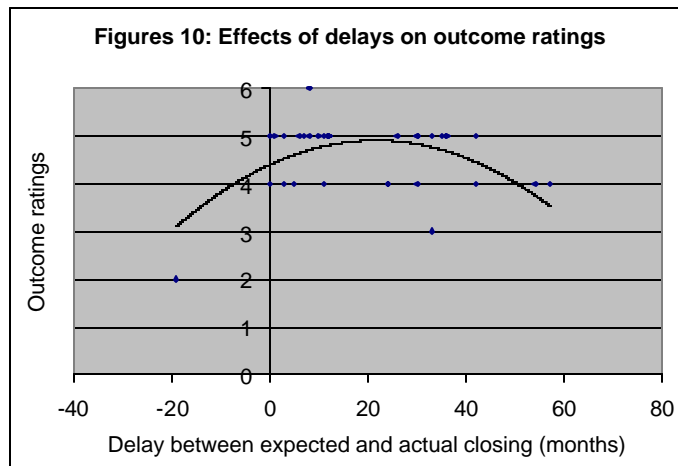
74. Since outliers may be skewing the correlation, it is necessary to examine more projects in the future. As more data (i.e. ratings) became available in future, the Office will be able to better assess the relationship between promised co-financing and outcomes, and promised co-financing and sustainability.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



4.3 Time Lags in Implementation Completion

75. The 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 in terms of outcomes and sustainability.

76. The analysis found that projects that reached completion closed on average 20 months after the expected completion date: 18 (44%) projects closed with a delay of two years or more. In addition, data shows that after two years, the quality of outcomes and their sustainability tends to decrease⁶ (figure 10⁷ and 11). The Office will continue to track this relationship in the future APRs, but these findings suggest that closer attention to factors affecting project implementation delays during midterm reviews might permit early detection and correction of factors affecting outcomes and their sustainability.



⁶ This decline is correlated with the square of the time delay and is significant at 95% confidence level (using a binomial regression).

⁷ For Figures 10 and 11, the outcomes and sustainability six-scale rating system was converted to a scale of 1 through 6, with 1 being HU and 6 being HS for outcomes, and 1 being HU and 6 being HL for sustainability. A delay of zero means that the project closed when expected.

MONITORING AND EVALUATION

CHAPTER V: PROJECT-AT-RISK SYSTEMS OF GEF PARTNER AGENCIES

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

78. 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 MIS system. These ratings are then aggregated and 'rolled-up' for the portfolio-level reporting. The risk monitoring systems of the GEF partner agencies are evolving rapidly. Several of the partner agencies have recently installed new Management Information Systems (MIS) that 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 is often updated only once or twice per year);
- 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; and
- 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 Approach

79. This assessment was based on information provided by GEF agencies and in interviews conducted with Implementing and Executing agency representatives. The assessment looks at the ways in which systems are designed. The assessment did not examine the way the systems are being implemented because the assessment didn't audit the actual internal reports to determine the degree of compliance with formal procedure. This would be an important element to be included in any future assessment by the GEF of the quality of project supervision.

Box 2: GEF Partner Agencies

Implementing Agencies

- United Nations Development Program (UNDP)
- United Nations Environment Program (UNEP)
- World Bank (IBRD)

Executing Agencies

- African Development Bank (AFDB)
- Asian Development Bank (ADB)
- European Bank for Reconstruction and Development (EBRD)
- Food and Agriculture Organization (FAO)
- Inter-American Development Bank (IADB)
- International Fund for Agricultural Development (IFAD)
- United National Industrial Development Organization (UNIDO)

80. Agencies reviewed include the three implementing agencies and seven presently approved executing agencies, as shown in (Box 2). Five of the agencies are specialized United Nations organizations which carry out projects and provide technical assistance to developing countries, and five are development banks which provide loan finance to member countries, with GEF project financing handled in the form of grants.

5.2 GEF Requirements for Project Preparation and Implementation

Preparation Phase

81. During the project preparation process there are several opportunities for risk issues to be raised and possible clarifications or improvements to be made 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 the 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 exception is found within the GEF review criteria for full-sized projects, which stipulate that at Work Program Inclusion, the project document must describe the project logical framework, "*including...risks and assumptions,*" while at CEO Endorsement the final project description should include "*details of project activities, inputs, and related risks and assumptions...*" (Annex H, rev. March 2004, Review Criteria for GEF Full-Sized Projects, p.3.).

82. The GEF Secretariat Proposal Agreement Review template (Annex F-1, rev. Dec. 2003), which tracks project preparation issues and specifies how, when, and by whom the project-at-risk related issues are to be addressed, does not include a category specific to risk. Though this can be presumed to fall under the heading for Project Design as it is there where the 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

83. The annual Project Implementation Report (PIR) is the standard reporting document for GEF-financed projects during the 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 agencies, and the form itself

has evolved over time. For example, a UNEP PIR review 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, a UNDP PIR review form includes 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.*"

84. At another level, 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. It should also be pointed out that the GEF review and reporting criteria, during preparation as well as implementation phases, make frequent reference to the concepts of sustainability and replicability, and 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 disconnects between project outcomes, sustainability and replicability can be monitored as part of the Annual Performance Report process.

5.3 Agency Reporting Systems

85. Most agencies' reporting systems cover two main components: financial flows, which are aggregated for internal management and accountability purposes, and some form of project performance information as well, which may or may not be aggregated at the portfolio level. Information technology in the form of management information systems (MIS) has become the norm among Implementing Agencies and Executing Agencies 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.

86. Monitoring of project performance varies in details (frequency, criteria assessed, etc.) yet project-level monitoring is a longstanding practice. In almost all cases, project performance ratings are recorded as part of regular project monitoring by staff of the IA or EA. 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 (i.e. by focal area, geographic region, or agency). Typically such scores are reported on a scale of 1 to 4, although several agencies have now shifted to a 6 point rating scale in order to capture variation in performance in better manner. The frequency of assessment varies, but for most agencies this is typically once or twice per year. The PIR itself is an annual document, but this does not preclude the possibility that an IA or EA conducts ratings more frequently. In recent years such systems have become increasingly computerized in the form of a corporate MIS system, with agency management giving more attention to results monitoring and reporting. (In several

agencies this is an ongoing process, and the transition to new MIS software can often be difficult and time-consuming, with a significant learning curve for agency staff.)

87. A key point is that most of the project reporting systems are dependent on self-rating of project performance and risks by project officers or supervision staff of the IA/EA. Immediate 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, (i) to ensure that ratings are entered on a timely basis by supervision staff, and (ii) 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 FY05; during the 1990s many projects had missed one or more reporting dates, and the consistency and quality of information provided often varied widely.

88. Taking the World Bank reporting system as an example (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. Though there have been numerous changes in software and formatting since the early 1990s, the essential elements have remained broadly similar (See Box 3).

Box 3: 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 & Evaluation.

Corporate MIS system generated flags

- Project effectiveness delays
- Disbursement delays
- Country performance issues

89. Any category rated less than satisfactory is considered a 'risk flag.' The MIS can then tabulate these ratings as a function of the regular internal reporting system. Depending on the number of warning flags which 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' project. The 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 task team ratings and risk flags being tracked by the MIS. In addition, the effectiveness or 'pro-activity' of task teams in resolving project issues is monitored, with the MIS calculating the time lag between the moment when a project is identified as being at-risk until the issue(s) have been resolved and the performance rating has been upgraded.

90. The World Bank currently considers one year as the maximum period acceptable for a project to remain in at-risk status, failing which the task team (and by inference, the

line management) is deemed to be 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 pro-activity 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 country, regional and portfolio-wide levels.

91. Other agencies also have explicit risk monitoring systems. For example, at the time this report was being prepared, IADB was introducing a ‘Project Alert Identification System.’ UNDP’s new ‘ATLAS’ MIS system 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 Low, Medium, High, Not Applicable or To Be Determined, and in the case of High Risks, a separate worksheet is to be completed indicating what management actions are being taken to mitigate that risk. These approaches generally parallel the recent PIR approach, but there is no information on the comparability of risk judgments across agencies.

92. Other elements of risk also enter into the picture, notably in the domain of fiduciary or safeguard risk (i.e. non-compliance with agency policies or misuse of resources). All of the partner agencies have a set of fiduciary policies in place, in addition to those required by GEF, which are legally binding on the grantee and are supervised as a matter of course. In the World Bank, certain high-profile projects may also be classified as ‘corporate risk projects’ during preparation, based on the potential for ‘safeguard non-compliance’ and the magnitude of potential impacts (involuntary resettlement, indigenous people, safety of large dams, etc.). This is the main exception to the self-rating of risks within the World Bank; otherwise risk is understood to be managed within the business unit.

93. Projects could be at-risk due to:

- Poor performance in implementation;
- Non-compliance with fiduciary policies; and,
- Identification of high or substantial risk factors, perhaps outside the immediate project.

5.4 Challenges in Ensuring Compatibility across Agencies

94. From the standpoint of trying to monitor overall portfolio risk of the GEF, there are basic problems of ensuring comparability, as in practice different agencies may adopt somewhat different definitions and approaches in how they operationalize risk management. For example, the development banks or IFIs (International Financial Institutions), which represent half of the ten partners covered in this assessment, are lending entities with governance, supervisory, and legal instruments which differ in many

ways from the specialized UN agencies which comprise the remaining GEF partners. The IFIs operate in capital markets and are already giving substantial management attention to factors such as credit and market risk. In some of the IFIs 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 (see Basel Committee 1998 and 2003). Indeed, the concept of ‘projects at risk’ (already used by the World Bank and being introduced by other International Financial Institutions) is derived from the banking sector where non-performing 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 IFIs (and perhaps for 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*⁸. (Basel Committee, 2003; emphasis added).

95. For this assessment, an inventory card was developed which lists 17 characteristics of a project risk monitoring system (see Appendix 4). 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 for an effective risk monitoring system. The inventory card was filled-out by each of the ten GEF Implementing and Executing Agencies; Table 12 below presents a tabulation of these results. Wide variation in scores for agencies, ranging from 10 ‘Yes’ responses to 17 (out of a total of 17 questions), are apparent. Some of the agencies report that meet all of the critical risk monitoring elements, in which scores ranged from 3 to 5 (out of a total of 5 critical elements). On the whole the scores are considerably higher for the development banks, which is consistent with the analysis presented in the earlier sections.

Table 12: Agency Risk-Monitoring Inventory⁹

Agency	Total “YES” responses	Critical Elements with Yes Responses
ADB	17	5
AFDB	15	4
EBRD	17	5
FAO	12	4
IADB	14	4
IFAD	12	3
UNDP	16	4
UNEP	15	4
UNIDO	10	3
World Bank	14	4

⁸ Basel Committee on Banking Supervision. February 2003. *Sound Practices for the Management and Supervision of Operational Risk*. Bank for International Settlements. Also see Basel Committee on Banking Supervision. September 1998. *Operational Risk Management*.

⁹ Appendix 5 presents the risk monitoring inventory disaggregated by agency.

96. The data generated through self-reporting is known to be affected by issues related to candor, under-reporting of risks, and overestimation of performance. Recent World Bank assessments show disconnect in the range of 30-40% between ratings given by task teams, and those of independent reviewers such as the Quality Assurance Group (QAG). This may be attributed to a staff perception that ‘problem projects’ are to be avoided, or that “managers don’t want to hear the bad news,” and observers frequently mention an incentive system which 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, the numbers of areas being monitored by the MIS, etc¹⁰.

97. These systems are sometimes seen by the operational staff as little more than bureaucratic paperwork – feeding the beast – with little utility for projects and a drain on scarce staff time. The task teams often complain of being unable to keep up with the changes in reporting requirements and criteria. Agencies have tried various ways to streamline these procedures, by trying to focus on reporting on key issues, reducing the amount of narrative text (which in any case presents major difficulties for portfolio “rollup” purposes), and standardizing data entry fields wherever possible. However, since institutional accountability need 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.

5.5 Independent Monitoring

98. The approach used by the European Development Bank (EBRD) differs in a significant way from that used by the other agencies. The EBRD places risk analysis within a special risk management vice-presidency, which reports to senior management and the Board, 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 (e.g. the Sarbanes-Oxley legislation in the U.S.), 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 coefficient for the institution. Significantly, EBRD management considers all GEF projects to be complex operations regardless of scale or focal area, and treats them as inherently risky needing special supervision and management attention.

99. COSO reviews (Committee of Sponsoring Organizations of the Treadway Commission), which are increasingly used by private sector firms in response to Sarbanes-Oxley, have also been conducted in some business units of the World Bank, to assess the quality of business practice with respect to internal controls, business ethics,

¹⁰World Bank 2006. *FY05 Annual Report on Portfolio Performance*.

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 (COSO reviews, Enterprise Risk Management Framework), independent validation of business units’ self-assessments, and tracking of progress in resolving problems. These functions can be within a corporate-wide entity, or within the business unit, but the key is autonomy of judgment and a flow of reliable information to senior management¹¹. These approaches are quite new, however, and their actual effectiveness in changing day-to-day institutional behavior is yet to be seen.

¹¹ Financial Executives Research Foundation. April 2003. “What is COSO? Defining the Alliance that Defined Internal Control.” www.fei.org. Also see Committee of Sponsoring Organizations of the Treadway Commission, September 2004. *Enterprise Risk Management—Integrated Framework*. www.coso.org.

CHAPTER VI: QUALITY OF PROJECT MONITORING

100. This Chapter addresses the quality of project monitoring of the GEF projects. The first part of the Chapter continues with the project monitoring analysis initiated in 2004. The second part of the chapter reviews the systems for quality control of Monitoring and Evaluation (M&E) arrangements at the point of CEO endorsement. This second part of the chapter 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.

101. There was an improvement in the quality of project M&E systems for projects with terminal evaluations 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 project monitoring systems. Nevertheless, for improvements to continue it is critical that more attention is given to M&E plans during project design. The review in the second part of this chapter reports that 58% of the projects endorsed by the CEO meet the GEF M&E requirements at project entry. The review found that there is considerable room for improvement during project preparation; 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 the M&E expectations among the project reviewers. This can sharpen the criteria for baseline information required at the point of CEO endorsement, and accord better definition and enforcement of the M&E standards so as to ensure higher compliance. Investments to intensify the development of the right tools and indicators will improve the measurement of results of the GEF projects.

6.1 Quality of Monitoring During Project Implementation

102. This analysis uses 83 terminal evaluations, 42 from FY 2004 and 41 from 2005, submitted by the Implementing Agencies to the Office. The number of projects rated as marginally satisfactory or above in monitoring during implementation increased from 39% in 2004 to 52% in 2005. This improvement is attributed to IAs' actions undertaken to address issues raised in the 2004 report and the ongoing changes within the IAs to advance the quality of monitoring. While this improvement is important, it is not conclusive because a sizable percentage (20 percent) of terminal evaluations submitted in 2005 did not provide sufficient information to assess quality of project monitoring.

Approach

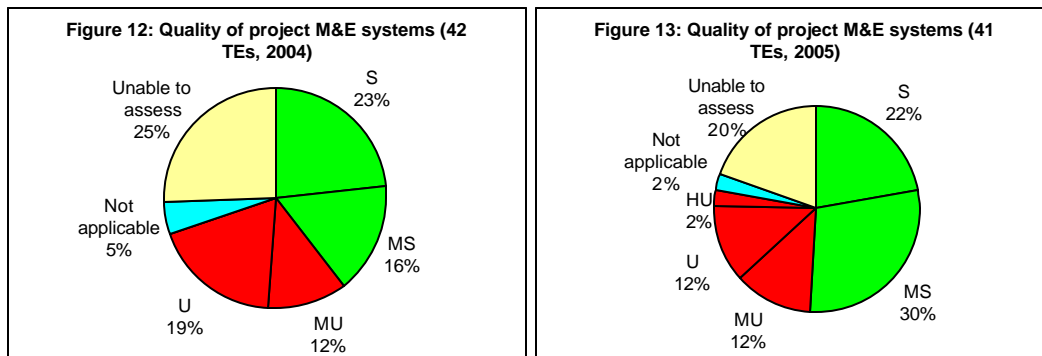
103. The Office rates the quality of project monitoring using the following criteria:

- Whether an appropriate M&E system for the project was put in place (including capacity and resources to implement it) and whether this allowed for tracking of progress towards projects 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 times in the project.
- Whether the monitoring system was used effectively for project management.

104. The Office provided ratings to each of these questions that had equal weights (50/50) on the overall rating on the quality of the project monitoring.

Overall Findings

105. The Office began rating¹² the quality of project monitoring in 2004, which allows a comparison with the projects of 2005. As indicated in Figures 12 and 13, the proportion of the MS or better projects increased from 39% in 2004 to 52% in 2005 with the biggest gains being in the reduction of projects rated unsatisfactory from 19% to 12% and the increase in projects rated MS from 16% to 30%. While it is too soon to interpret this change as a trend, ongoing efforts of the IAs to advance the quality of M&E systems (see box 4) are likely to be an important factor in these improvements, thus it could be expected that this improvement may continue in future.



106. The proportion of terminal evaluation reports providing insufficient information only had a slight decrease (from 25% to 20%) which indicates that there is still significant room for improvement. The APR will continue examining the trends in future years to assess whether the gradual improvement trend continues.

¹² The GEF Evaluation Office rated the quality of project M&E systems based on a six-point rating scale as follows: Highly Satisfactory (HS), Satisfactory (S), Moderately Satisfactory (MS), Moderately Unsatisfactory (MU), Unsatisfactory (U), and Highly Unsatisfactory (HU). Unable to assess (UA) was used when the information contained in the terminal evaluation did not allow an assessment.

Box 4. Implementing Agency Initiatives to improve the quality of monitoring

The Implementing Agencies have taken several measures to improve the quality of project monitoring of the achievement of objectives and expected project outcomes. This has included studies to assess the current monitoring practices across their portfolio of the GEF projects to determine the areas that need improvement. For example, UNDP indicated that they conducted a study on the “Status of Monitoring and Evaluation in UNDP-GEF projects” to characterize M&E practices within and across the portfolio, based on a sample of thirty GEF projects. The study indicated that “of these projects not one was found to be exemplary across all of the elements of its M&E system.” The study found that the “*Quality of Indicators*) and *Identification of Sources of Verification* were almost universally weak.” While there are significant challenges associated with measuring change in global environmental benefits, UNDP-GEF indicates that it is piloting the use of scorecards which help to reduce inconsistencies in measurement. UNDP-GEF also indicates that it is improving on use of repeatable independent measurements of baseline conditions, mid-term conditions, and end of project conditions.

Having carried out a similar assessment of M&E systems of the GEF projects two years ago, the main actions taken by the World Bank are training and more diligent review. The World Bank-GEF team organized a one week training workshop for task teams, attended by 30 staff, on planning and organizing M&E in projects and specially focusing on the Results Framework. They also organized a workshop for GEF Regional Coordinators and Thematic Specialists in their role as reviewers of the Results Framework so they can be better able to provide guidance to task teams. The level of review was also raised as the World Bank imposed much stricter criteria for the minimum standards for M&E arrangements in projects. Most of the World Bank regions have established an M&E help desk to provide support to task teams on M&E.

UNEP indicated that at the project design phase, they are emphasizing the identification of key SMART indicators at the outcome level and are also ensuring that activities needed to track the SMART outcome indicators are clearly specified in the M&E plan and that resources needed to track the indicators are realistically budgeted. In addition, UNEP is compiling baseline information and data prioritizing the information relevant to the outcome-level indicators. Specifically for biodiversity, UNEP indicated that they are using the tracking tools for Strategic Priority 1 (Catalyzing Sustainability of Protected Areas) and Strategic Priority 2 (Mainstreaming Biodiversity in Production Landscapes and Sectors) and are awaiting the results of the other Task Forces concerning development of tracking tools for the other focal areas.

At the project implementation stage, UNEP indicated that project steering committees are expected to review progress in meeting project objectives, review whether selected indicators are being monitored and whether they are actually relevant and cost-effective. Project steering committees meet on a yearly basis and reflect recommendations relevant to project M&E systems in the minutes of the meetings. In addition, the terms of reference of mid-term project reviews emphasize results monitoring; again, a review of the outcome indicators is done at this stage and the tracking tools are validated.

Quality of Monitoring by Implementing Agencies and Focal Areas

107. As indicated in Tables 13 and 14, UNDP had the largest increase in the ratings for the quality of the project monitoring, and the performance of World Bank on the other hand was stable. The numbers for UNEP were too small to allow any significant assessment. It is still early to draw any conclusions on the quality of monitoring because this has been tracked for only two years, but the number of projects rated under marginally satisfactory or with insufficient information to allow assessment remains high for all Implementing Agencies (46% of 41 projects). This will continue to be assessed as part of the future APRs.

Table 13: Monitoring Quality by IA (Projects)

Rating	WB		UNDP		UNEP	
	04	05	04	05	04	05
Year						
HS	0	0	0	0	0	0
S	10	5	0	4	0	0
MS	4	3	1	6	2	3
<i>Sub Total</i>	<i>14</i>	<i>8</i>	<i>1</i>	<i>10</i>	<i>2</i>	<i>3</i>
MU	0	1	4	4	1	0
U	4	0	3	4	1	1
HU	0	0	0	1	0	0
<i>Sub Total</i>	<i>4</i>	<i>1</i>	<i>7</i>	<i>9</i>	<i>2</i>	<i>1</i>
NA	0	0	0	0	2	1
UA	3	3	5	4	3	1
Total	21	12	13	23	9	6

Table 14: Monitoring quality by Focal Areas (projects)

Rating	BD		CC		IW		Others	
	04	05	04	05	04	05	04	05
Year								
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
<i>Sub Total</i>	<i>11</i>	<i>15</i>	<i>4</i>	<i>2</i>	<i>1</i>	<i>4</i>	<i>0</i>	<i>1</i>
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
<i>Sub Total</i>	<i>9</i>	<i>6</i>	<i>0</i>	<i>2</i>	<i>4</i>	<i>1</i>	<i>0</i>	<i>2</i>
NA	0	0	0	0	0	0	2	1
UA	4	0	4	6	2	0	1	1
Total	24	21	8	10	7	5	3	5

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

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

109. This review draws its conclusions from the examination of the 74 full size projects that were CEO endorsed in FY 2005; an examination of the comments provided during the review process by the GEF Secretariat reviewers, the GEF Council members, and the STAP roster reviewers; and interviews carried out with GEF Secretariat and Implementing Agency staff. Appendix 2 presents a more detailed description of the methodology followed during the review. To assess the quality of M&E plans at entry an instrument was developed. This instrument measures 13 specific aspects of M&E quality, which are based on the Review Criteria of the GEF Secretariat (2000) and the guidelines contained in the M&E Policies and Procedures (2002).¹³ In some cases, parameters outlined in these documents were refined to facilitate consistency and objectivity in the application of the assessment instrument.¹⁴ Certain technical or operational elements were, however, not included as this would have required specialized technical expertise on individual projects and also would have introduced greater subjectivity into the review process (see Box 2).¹⁵

110. Since the inception of the GEF, the GEF Council has at numerous occasions given attention to the need to strengthen policies and procedures for Monitoring and Evaluation

¹³ See Appendix 2 A.

¹⁴ 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.

¹⁵ The parameters that were mentioned in the two documents but not used in the instrument include discussion on key assumptions of the project, sufficiency of M&E budget; and adaptive management.

(M&E).¹⁶ In order to streamline the project review process and also as a response to the request of the Council, the GEF Secretariat developed Project Review Criteria (2000) which laid down the GEF requirements, including those for the M&E arrangements, at various stages of the project cycle for both full size and medium size projects. In January 2002 the GEF Secretariat published “Monitoring and Evaluation Policies and Procedures” wherein it defined the expectations for project quality including those for M&E arrangements at entry. This review uses the requirements applicable at the point of CEO Endorsement to develop the criteria to assess whether the projects are in compliance with the Council expectations for M&E arrangements at entry. The parameters for assessment have been classified as Critical Parameters – where non compliance indicates serious deficiencies in the M&E arrangements – or Other Parameter (see Box 5). To comply with the GEF M&E expectations at entry a project needs to be in compliance with all the critical parameters and needs to perform sufficiently well on all the parameters together (see Appendix 2). To be classified as compliant, on a scale of one to three (greater the better) the projects were required to score at least two on each of the critical parameters and were required have an aggregate score of 26 out of the maximum possible 39. Since these criteria are also consistent with the New GEF Policy on M&E (2006), the findings of this study will form a baseline for monitoring and assessment of the implementation of the new policy.

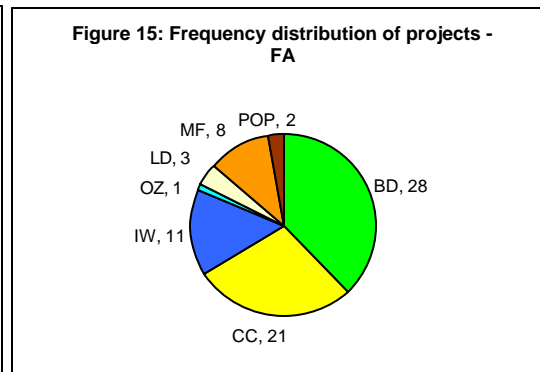
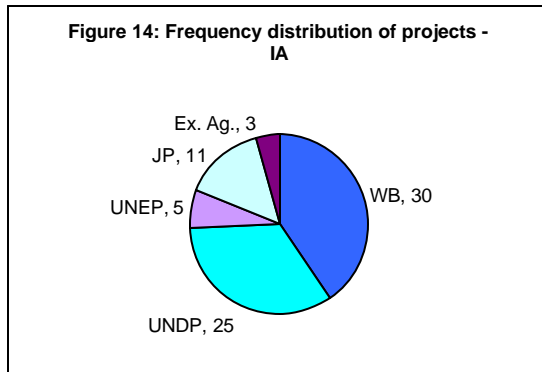
Frequency Distribution of the Projects

111. Of the 74 projects reviewed, the World Bank is the implementing agency in 30 projects (41%), UNDP in 25 (34%), and UNEP in five (7%) projects. Of the remainder, 11 (15%) are Joint Projects¹⁷ (JPs) and three (4%) are being implemented by EAs (Figure 14). Among the focal areas, biodiversity has 28 (38%) projects, Climate Change 21 (28%), International Waters 11 (15%), Multi Focal Areas eight (11%), Land Degradation three (4%), Persistent Pollutants two (3%), and Ozone Depletion has one (1%) project (Figure 15). Thus, broadly speaking among the implementing agencies only the World Bank and UNDP, and among the focal areas only Biodiversity and Climate Change, have sufficient numbers to facilitate inter group performance comparisons¹⁸.

¹⁶ 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).

¹⁷ Joint Projects refer to projects that are being jointly implemented by any combination involving two or more IAs or EAs or both.

¹⁸ Three projects that were part of the studied cohort – the “Coral Reef Targeted Research and Capacity Building for Management” (International Waters), the “Support Program for National Capacity Self-Assessments” (Multi Focal Areas), and the “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 may also not be an effective management tool for such projects. However, a main reason to include these projects for assessments is to be consistent with the Terminal Evaluation Reviews (terminal evaluation reviews). Further, since there are only three such projects their inclusion will not substantially change the overall conclusions of the study.



Findings

Overall Project Compliance

112. Taking into account the respective scores of projects on each of the critical parameters as well as the cumulative scores, the study found that the M&E plans of 58% of the projects meet the overall M&E expectations at the point of CEO Endorsement. Twenty two percent of the projects were non compliant in one critical parameter, 14% in two and 7% in three parameters. The projects that were in compliance with all the critical parameters also had an overall score of 26 or more.¹⁹

Table 15: Projects with Overall Compliance – Implementing Agencies

Implementing Agency	Compliant
World Bank	15 (50%)
UNDP	17 (68%)
UNEP	2 (40%)
Joint Projects	7 (64%)
Executing Agencies	2 (67%)
All Agencies	43 (58%)

113. Among the implementing agencies, 68% of UNDP projects, 50% of World Bank Projects and 64% of Joint Projects meet the M&E expectations (Table 15). Among the focal areas, 76% of the Climate Change projects, 55% of International Waters, 50% of Biodiversity and 50% of Multi Focal Areas projects meet the M&E expectations at CEO Endorsement (Table 16).

¹⁹ All the three atypical projects, which have been identified earlier in the methodology section, didn't comply with the M&E expectations. If these projects are dropped from the analysis, the projects in overall compliance with the M&E expectations increase from 58% to 61%. Similarly, among the IA's and focal areas the projects with overall compliance will change to 52% for World Bank, 52% for Biodiversity, 60% for International Waters, and 57% for Multi Focal Areas. Thus, dropping the three markedly different projects from the analysis does not substantially change the import of the findings.

Table 16: Projects with Overall Compliance – Focal Area

Focal Area	Compliant
Biodiversity	14 (50%)
Climate Change	16 (76%)
International Waters	6 (55%)
Land Degradation	1 (33%)
Multi Focal Areas	4 (50%)
Ozone Depletion	1 (100%)
Persistent Pollutants	1 (50%)
All Focal Areas	43 (58%)

114. The overall compliance with the GEF M&E expectations seems to be better in the projects that have been approved recently. For example, within the reviewed cohort while 52% of the projects approved by the council before or at June 30, 2003, comply with the GEF M&E expectations, 61% of the projects approved after June 30, 2003, are in compliance. The recent measures implemented by the UNDP and by the Biodiversity focal area to address the M&E problems seem to have led to an improvement in their overall compliance performance: for UNDP the number of compliant projects improved from 60% to 72%, whereas for Biodiversity it improved from 40% to 56%.²⁰ Although, the level of change is not statistically significant it is in the expected and desired direction. A better picture of the changes in compliance level over time will emerge only after the cohorts for other years have been assessed.

Strengths and Weaknesses

115. All or almost all M&E plans are in compliance with the GEF M&E expectations in specifying relevant (100%) and quantifiable (97%) indicators, in providing baseline information (92%), in explicitly allocating a budget for M&E (92%), in specifying responsibilities (99%) and time frames (99%) for M&E activities, and in specifying targets for project outputs (95%). Out of these, the GEF bar for the expectations on baseline information has been low and, thus, the compliance standards may not adequately address the need for adequate baseline information.²¹ If the provisions of the new M&E policy (2006) on baseline information – which requires projects to provide adequate baseline information at the point of work program inclusion in almost all the

²⁰ The World Bank has recently made a shift from the log-frame based monitoring systems to the results framework based monitoring system. While quality of the M&E plans in the projects that used results based framework are better than those in the log-frame based (58% compared to 37%), this study does not provide enough evidence to show that this improvement is indeed due to the policy shift. Both types of World Bank projects – those that use results based framework or else use logframe approach – are spread over time and the number of World Bank projects is small to allow robust conclusions on this front.

²¹ Although there is a strong case for requiring the baseline information to be provided upfront, keeping in mind the difficulties and costs involved in establishing baseline conditions for very complex projects, the Project Review Criteria required at the time of the review requires the projects to provide baseline information within the first year of project implementation. Therefore, for compliance on this parameter a project just had to promise to provide baseline information within the first year of project implementation. Although 92% of the projects are in compliance on this parameter, 53% just promise conducting a baseline survey and provide no baseline information upfront. The new M&E policy of GEF (2006) requires projects to provide baseline information upfront except in “rare” situations wherein baseline information could be provided within the first year. Clearly, presently the exception is being made in more instances than what can be called “rare.”

cases – is taken into account, the present level of performance where more than half of the projects just promise that they will provide baseline information before the end of the first year of project implementation is inadequate. At present, many projects receiving PDF funds are apparently not addressing baseline data needs during project preparation though this would appear to be a highly appropriate mechanism for doing so.

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

117. This study found that the UNDP projects rated better than the World Bank projects 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 midterm reviews and terminal evaluations will be undertaken. On all these parameters, it is reasonable to expect that performance by other agencies could also be substantially improved if more attention is given to these M&E aspects by task teams and management during project preparation.

118. The World Bank and UNDP have comparable performance yet both need to improve in: identifying specific and sufficient indicators; providing baseline information upfront; 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 (i.e., defining the appropriate units of measurement for biodiversity or land degradation projects, dealing with long time lags in outcomes, etc.). Already a substantial investment has been made in establishing sets of core indicators for some focal areas such as biodiversity. As indicated below, in other focal areas efforts on their way to develop indicators must be treated with greater urgency.

119. At the level of focal areas, the study found that Climate Change does better than Biodiversity in terms of: providing specific, sufficient, relevant and quantifiable indicators for project objectives and outcomes; in providing baseline information on indicators up front; and, in specifying targets that are based on an assessment of initial conditions. The parameters on which performance of Climate Change and Biodiversity vary are, however, different from those on which UNDP and World Bank differ. While differences between the World Bank and the UNDP 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.

120. In recent years the GEF Focal Area task forces have been making continued efforts to define core focal area indicators and to develop tracking tools to improve the quality of M&E arrangements.

- The Biodiversity Focal Area task force has developed a tracking tool to monitor performance of the projects of ‘Catalyzing Sustainability of Protected Areas (Strategic Priority I);’ and, ‘Mainstreaming Biodiversity in Production Landscapes and Sectors (Strategic Priority II)’, on programmatic indicators. The Biodiversity Focal Area team is also grappling with finding ways to address the M&E concerns related to ‘capacity building for the implementation of Cartagena Protocol on Bio Safety (Strategic Priority III)’ and ‘generation and dissemination of best practices for addressing current and emerging biodiversity issues (Strategic Priority IV),’ where due to the inherent nature of the projects designing good and cost effective M&E plans is 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. They are also refining the 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.
- 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 also seek to identify best practices and tools of analysis that can be applied beyond GEF projects.

121. The process of indicators development is, however, complex. It requires the application of current scientific knowledge and sound technical knowledge to results measurement in the GEF. In the case of the tracking tools that are now used in the biodiversity focal area to monitor performance in Strategic Priorities I and II the GEF drew on instruments that had already been developed incorporating current scientific and technical knowledge on the subject. In other focal areas such instruments do not exist and there is a need to compile and assess the relevant knowledge in light of GEF’s needs to track results. Some focal areas are seeking access to this expertise through partnerships. For example, IW Focal Task Force is establishing a partnership with the International Hydrological Programme Working Group ground water initiative of UNESCO for the development of groundwater indicators and is working with the scientists of Iowa State University to develop catalytic indicators for nutrient reduction projects in the Danube Black Sea basin. Similarly the Land Degradation 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 for the land degradation focal area.

122. This said there is still a need for intensification of the present efforts for further development of tools such as project level indicators and tracking tools. This will require corporate investments to address the technical challenges specific to each Focal Area, to build consensus on indicators, to define ways to roll-up results at the portfolio level and to find out ways to address issues related to attribution.

Project Review Arrangements

Availability of Reviews and Comments

123. In the project documents maintained in the Project Management & Information System (PMIS) the GEF Secretariat reviews were included for all the projects, 70 (95%) projects included the STAP reviews (along with the project team responses). In comparison, only 32 (43%) projects included the compilation of the comments by the Council members. An examination of other databases maintained by the GEF Secretariat showed that comments by Council members had been documented for at least 69 (93%) projects. Thus, either a substantial number of projects do not include the comments by Council members 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 system. Due to this, in addition to the PMIS system, other databases that maintain comments by Council members were also accessed.

The GEF Secretariat Review

124. The GEF Secretariat reviewers appraise projects throughout the pipeline process including pipeline entry, work program inclusion, and CEO endorsement. The GEF Secretariat reviewers are expected to address the M&E issues in their reviews in a comprehensive manner. They may require the project task team to rework the project documents if the documents did not meet the GEF requirements for the stage of review. The guidance for the GEF Secretariat reviewers, however, is not fully adequate in clarifying what is expected from reviewers in terms of addressing M&E issues. 'Project Review Criteria', 2000, and 'M&E Policies and Procedures', 2002, despite covering M&E issues in detail, does not specify the compliance standards for M&E parameters. As a result, there is wide variation in the manner in which the GEF Secretariat reviewers interpret and apply the M&E standards.

125. The GEF Secretariat reviewers pointed out at least one weakness in the M&E arrangements in 55% of the projects in their reviews. For the projects that do not comply with the GEF M&E expectations – where one would also expect more comments pointing out M&E weaknesses – the GEF Secretariat reviewers pointed out at least weakness in 48% of the projects. Thus, a substantial number of projects where the M&E arrangements at the point of CEO Endorsement were not in compliance with the GEF M&E expectations at entry, had not been commented upon by the GEF Secretariat.

126. An assessment of the extent to which the comments by the GEF Secretariat reviewers were incorporated into the project document could not be done because a task team generally responds to their comments primarily by way of incorporating the recommended changes in the subsequent versions of the project documents. The task team is not required to formally document its responses to the GEF Secretariat comments.

The Council Member Comments

127. The Council members comment on the 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 Pipeline Entry, an earlier stage in the

pipeline. Unlike the STAP roster reviewers, where generally only one reviewer is involved in a review, many Council members may choose to comment on a project proposal simultaneously and independent of each other. For the purpose of this study the compilation of the comments by the Council members on any given project proposal has been considered as a single ‘review.’ The rationale for this is that these compilations of the Council member comments often result in a task team making improvements in the project proposal documents. Thus, it is in this sense that the compilations of comments by the Council members perform a role similar to the reviews by the GEF Secretariat and the STAP roster reviewers.

128. Overall, the Council members pointed out at least one weakness in the M&E arrangements in 58% of the projects. For the projects that do comply with the GEF M&E expectations – where one would expect more comments – at least one weakness in the M&E plan of a project was pointed out by the Council members in 68% of 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.

129. Although Council reviews are available for 69 projects, these reviews were incorporated in the project documents maintained in the PMIS in only 32 instances²². Of these 32 projects, the Council had pointed out weaknesses in the M&E plan in 20 (63%). In all 20 cases the project team claimed to have addressed all or some of the M&E issues raised by the Council members.

The STAP Roster Review

130. The STAP roster reviewers have so far not been asked to address specific M&E issues in their reviews. The terms of reference for the STAP roster review addresses M&E in a very peripheral manner. Consequently, in most cases STAP Reviewers do not address M&E issues such as technical feasibility of indicators, feasibility of methodology, and cost-effectiveness of M&E systems, even though these could be considered as the technical aspects where the STAP roster reviewers have a comparative advantage.

131. The STAP roster reviewers appraise the project proposal documents before the work program inclusion. They, therefore, get to review project documents when they are still in a more preliminary stage.

132. In their project reviews, the STAP Reviewers pointed out weaknesses in M&E arrangements of 40% of the projects. For the projects that did not comply with the GEF M&E requirements, the STAP roster reviewer pointed out at least one weakness in the M&E plan in 39% projects. Thus, even though the STAP roster reviewers are not required to address the M&E concerns in their reviews, in many instances they do address them *suo moto*.

²² The Council Reviews for remaining 37 projects were accessed from another central database that exclusively maintains Council Reviews. It is not known whether the task team responded to these reviews. In any case the responses of the task team for such reviews are not available.

133. Of the 28 projects where the STAP reviewer pointed out a weakness in the M&E plan, in almost all instances (93%) the task team specifically responded to these comments. In 12 (43%) instances the project team claimed to have addressed all of the concerns of the STAP reviewers, and in nine (32%) they claimed to have addressed some of the concerns, whereas in five (18%) the extent of the changes was not described. In all the instances where only a few or none of the changes suggested by the STAP reviewer had been made, the project team did provide an explanation (i.e., justified the decision not to make any change despite the STAP reviewer's suggestion). These results appear to show that STAP Roster reviews are taken seriously by preparation teams and even in cases where the suggested change is not accepted, a justification for this decision is provided.

CHAPTER VII: QUALITY OF TERMINAL EVALUATION REPORTS

134. High quality terminal evaluations that provide accurate assessment of accomplishments and shortcomings of projects are not only essential as a learning tool, they are also important because they are the building blocks of the assessment of outcomes and sustainability in the APR. The GEF Evaluation Office began rating the quality of project Terminal Evaluation Reports in 2004 which allows a comparison with 2005 terminal evaluations. Appendix 6 presents a list of the terminal evaluations reviewed in FY 2005 and their respective ratings.

135. There has been an improvement in the quality of terminal evaluations submitted in FY 2005 compared to those submitted in the previous year. Improvements are most prominent in the terminal evaluations of 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 that had been identified in 2004.

136. It's for the first time that differences between the outcome/sustainability ratings given in the Terminal Evaluations by the Implementing Agencies and the Terminal Evaluation Reviews by the Office have been reported in the APR. No significant differences are observed between the Office and the World Bank's Independent Evaluation Group ratings on a six point scale. On average, the UNEP Terminal Evaluation Reports tend to rate performance a point (on a six point rating scale) higher than the Office. Since a large number of UNDP Terminal Evaluations did not provide ratings on outcomes and sustainability, it is difficult to know the extent to which it over rates performance.

137. Terminal evaluations 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 the actual costs including the total costs, costs disaggregated at the activity level, and co-funding. UNEP's evaluation reports also frequently exhibited inconsistencies between terminal evaluation text and ratings.

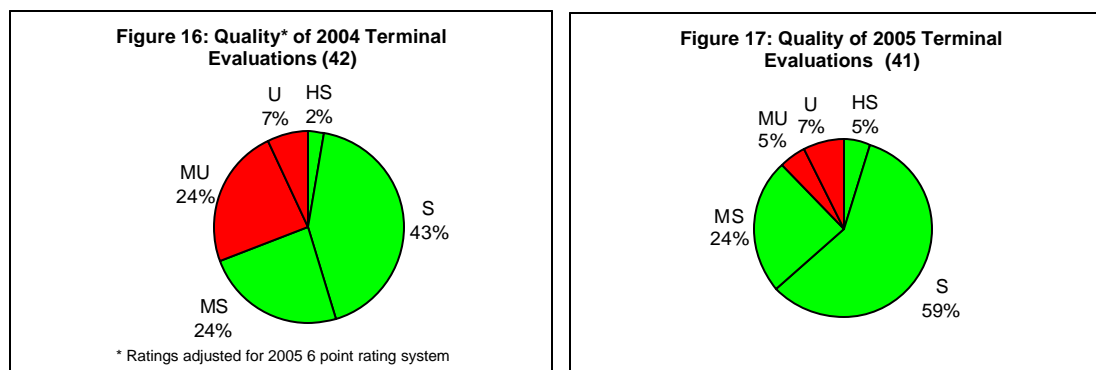
7.1 Approach

138. The forty one 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 were 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 co-financing 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

139. Terminal evaluations improved from 45% of marginally satisfactory and above in 2004 to 88 % in 2005 (figures 16 and 17).



140. The terminal evaluation quality ratings for the World Bank and UNDP were higher compared to 2004 (Table 17). Regarding UNEP, while numbers are still small to draw firm conclusions, the terminal evaluation quality ratings for 2005 dropped from those in 2004. Of the 41 terminal evaluations reviewed in 2005, two did not provide sufficient information to assess the project outcomes and five did not provide sufficient information to make an assessment of sustainability. Also, a higher proportion of CC terminal evaluations (4 out of 8 in 2004 and 6 out of 10 in 2005) seem to continue to provide insufficient information to make an assessment on the quality of the project M&E system.

Table 17: Ratings on the quality of Terminal Evaluations by Implementing Agency

Quality of terminal evaluations	WB		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
<i>Sub Total</i>	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
<i>Sub Total</i>	4	0	6	2	3	3
Total	21	12	13	23	9	6

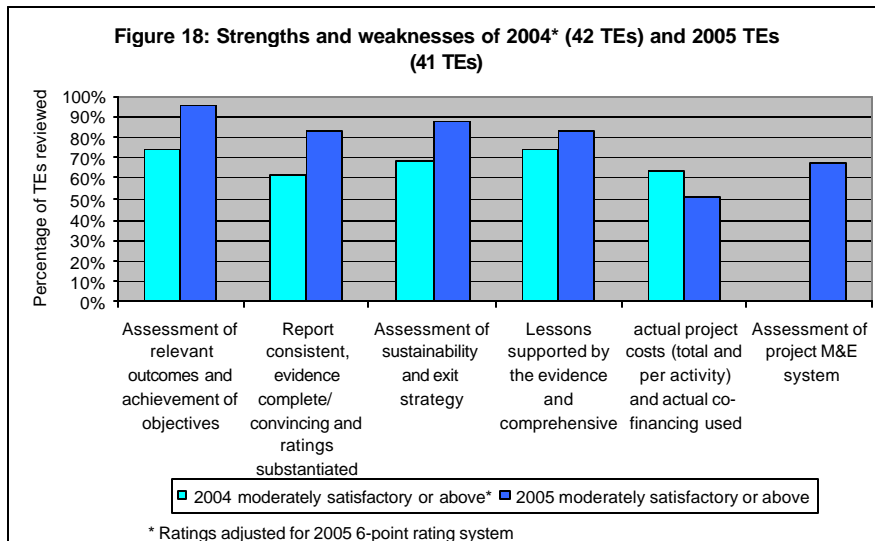
141. The quality of terminal evaluation reports is an area in which quick improvements can be expected, and it is likely that IAs actions undertaken to address issues raised in the 2004 Project Performance Report contributed to this significant improvement in the ratings of the quality of terminal evaluations (see Box 5). The issuance of clearer guidelines for the preparation of terminal evaluations in 2003 was also likely a contributing factor.

Box 5 Changes at UNEP and UNDP to Improve the Quality of Project Terminal Evaluations

UNDP and UNEP have recently undergone internal changes to improve their project evaluation processes, which also address better the needs of the GEF. For example, in FY 2005 UNEP took the important step to place its Evaluation and Oversight Unit directly under UNEP’s Executive Director, which provides the unit a greater independence from other operative units. Consistent to 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 of evaluations of GEF projects. UNEP has also decided that all GEF terminal evaluation reports will be subject to quality assessment reviews by UNEP’s Evaluation and Oversight Unit and that these reviews will be forwarded to the Office along with the evaluation reports.

In the case of UNDP, the study mentioned earlier in Box 2 “Status of Monitoring and Evaluation in UNDP-GEF Projects”, identified that an important factor underlying the large difference in the quality of terminal evaluations was the highly decentralized nature of the evaluative process in this agency. 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 these instruments through a series of country workshops and made them available through the internet. Also in 2005, the UNDP Evaluation Office assumed responsibility for evaluation for GEF-funded activities from Regional Coordination Units to properly harmonize GEF project evaluations with UNDP’s evaluation practices.

142. A more detailed assessment of the factors driving the quality of terminal evaluation reports using the Office criteria for this assessment shows that most of the key terminal evaluation quality issues that were identified last year are being addressed Figure 19²³. Only the reporting of actual project costs including co-financing got lower ratings for the 2005 terminal evaluations. Over 60 percent of 2004 terminal evaluations reported on actual project costs and this decreased to slightly over 50 percent of terminal evaluations providing moderately satisfactory or above information on actual project costs (total and per activity) and actual co-financing used.

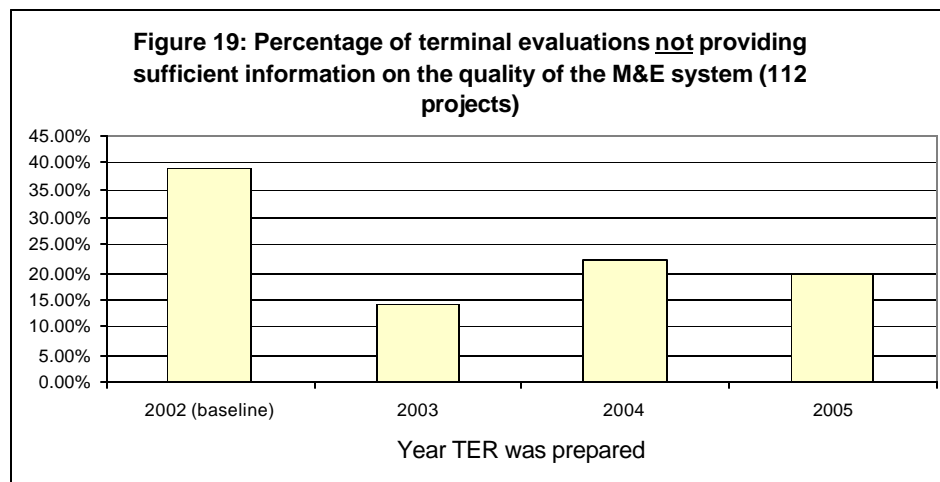


²³ The assessment of the quality of the project M&E system was carried out for the first time this year.

143. The ratings for the assessment of outcomes in terminal evaluations were higher in 2005 for all Implementing Agencies (See Appendix 7). For the World Bank and UNDP the ratings were higher in the other criteria, namely report consistency and evidence convincing, assessment of sustainability, and comprehensive lessons supported by evidence. For these three criteria, the 2005 ratings for UNEP dropped from those in 2004.

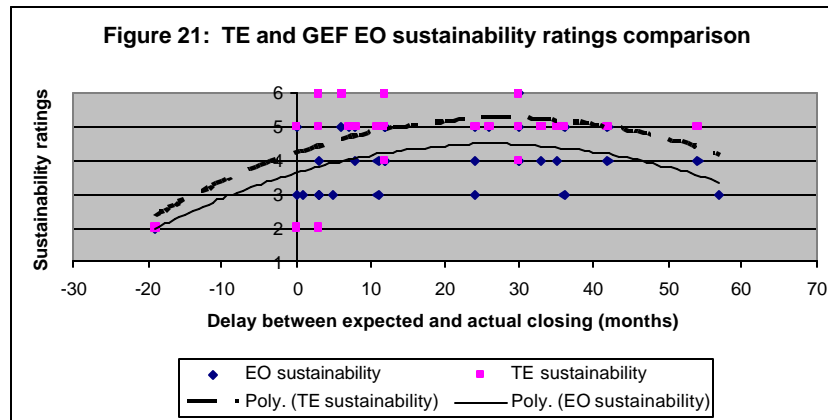
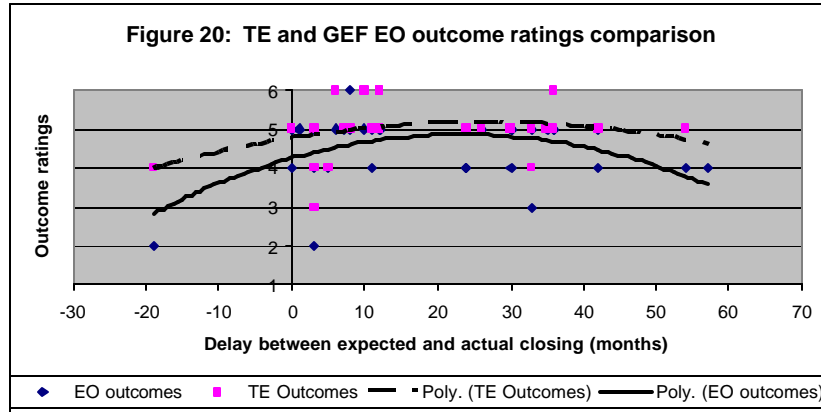
144. A more in depth analysis of reporting of actual project costs and co-financing for 2004 and 2005 indicates that most of the World Bank terminal evaluations were rated MS or above in their reporting of actual project costs and co-financing (10 of 12 in 2005), while UNDP and UNEP continue to have a large proportion of projects rated under MS on actual costs and co-financing (15 of 23 for UNDP and 2 of 6 for UNEP in 2005).

145. This year the Office has also begun tracking the reporting of project monitoring systems as part of the quality of terminal evaluations. A detailed assessment of trends regarding the proper reporting of monitoring in the terminal evaluations since 2002 (when the reviews of terminal evaluations began) reveals that terminal evaluation reporting on the quality of project monitoring seems to be improving since the Office issued the 2003 guidelines requesting that all terminal evaluations include an assessment of project monitoring systems as shown in figure 20.



7.3 Difference in Ratings by Implementing Agencies and GEF Evaluation Office

146. Figures 21 and 22 seem to suggest that on the whole, compared to the Office the Implementing Agencies tend to over rate project outcomes and sustainability. Given this difference, the Office conducted a more in depth analysis to assess the difference in the outcomes and sustainability ratings for individual projects by the Office and by the respective Implementing Agency.



Box 6: Binary Rating Scale

To calculate the rating difference for 2005 terminal evaluations using a binary system, the six point rating scale used for outcomes and for sustainability was converted to a binary system of 1 and 0. Thus, projects rated Moderately Satisfactory (MS), Satisfactory (S) or Highly Satisfactory (HS) were given a 1, and those rated Moderately Unsatisfactory (MU) or worse were given a 0. A similar process was followed for Moderately Likely (ML) sustainability or better, and Moderately Unlikely (MU) sustainability or worse. The next step was to calculate for each project, the difference between the rating provided by the GEF Evaluation Office and the rating provided in the terminal evaluation (or 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 was providing a rating higher than the one provided by the Office based on the evidence of the terminal evaluation. The average difference for each IA was calculated. The World Bank Independent Evaluation Group (IEG) also uses this binary system to assess the relevant difference in ratings.

147. Only 30 terminal evaluations reviewed in 2005 provided ratings on outcomes. When a binary rating scale is used of the 30 projects there is disconnect in the ratings by the Implementing Agencies and the Office in only two instances where Implementing Agencies had rated the project outcomes as MS whereas the Office rated them as MU or worse. For UNDP projects where ratings by both UNDP and the Office are available, there is no disconnect in the binary ratings. How representative is this of the UNDP's FY 2005 cohort is an altogether different question. 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 fact that all the UNDP projects for which both

UNDP and the Office provided ratings don't have any disconnect on a binary scale cannot be generalized to the total UNDP cohort for FY 2005.

Table 18: Average difference between IA and the GEF Evaluation Office outcome and sustainability ratings in 2005 terminal evaluations on binary scale

Agency	Total	Outcomes		Sustainability	
	Total number of terminal evaluations considered	Terminal evaluations that provided ratings on outcomes	Projects with Difference in Outcome Ratings for terminal evaluations and terminal evaluation reviews	Terminal evaluations that provided ratings on Sustainability	Projects with Difference in Sustainability Ratings for terminal evaluations terminal evaluation reviews
WB (IEG)	12	12	1	10	0
UNDP	23	13	0	10	0
UNEP	6	5	1	4	1
TOTAL	41	30	2	24	1

148. When binary rating scale is used there is disconnect in the World Bank's project outcome ratings and the Office ratings in only one out of 12 instances. The World Bank rated the outcomes of this project as Moderately Satisfactory whereas the Office rated it as Moderately Unsatisfactory²⁴. There was, however, no disconnect in the ratings by the World Bank and the Office on project sustainability (Table 18). For UNEP there was disconnect in outcome ratings on a binary scale in only one project out of five, and for sustainability in only one project out of four (Table 18). The '*Global - Barriers and Best practices in Integrated Management of Mountain Ecosystems*', project had been rated as Moderately Satisfactory by UNEP while the Office rated it as Unsatisfactory. The terminal evaluation indicates that this project failed to achieve many of its outcomes and sustainability is not ensured beyond project closure. The project was badly implemented due to poor project management which caused UNEP to terminate the project.

Table 19: Difference between IA and the GEF Evaluation Office Ratings on a Six Point Rating Scale

Agency	Total	Outcome		Sustainability	
	Total number of terminal evaluations considered	Terminal evaluations that provided ratings on outcomes	Projects with Difference in Outcome Ratings for terminal evaluations and terminal evaluation reviews	Terminal evaluations that provided ratings on Sustainability	Projects with Difference in Sustainability Ratings for terminal evaluations terminal evaluation reviews
WB (IEG)	12	12	2	10	3
UNDP	23	13	7	10	5
UNEP	6	5	4	4	3
Total	41	30	13	24	11

149. Even when the six point rating scale is used there is little difference in the World Bank's project outcomes and sustainability ratings and the Office ratings²⁵. Although the outcome and sustainability ratings by UNDP and UNEP do not differ with the ratings by the Office on a binary scale there were more apparent differences in the ratings on a six point scale. On a six point scale on average, UNDP over rated the outcomes of its projects by 0.4 points and sustainability by 0.3 points whereas UNEP over rated outcomes by 1.0 point and sustainability by 1.3 points. Despite the magnitude of the difference in average ratings – especially for UNEP – these differences in the outcome and sustainability ratings are not statistically significant. This is primarily because the

²⁴ This project was the *Ghana Natural Resources Management Project*.

²⁵ Ratings provided by the Independent Evaluation Group of the World Bank have been considered for the analysis.

numbers of observations where comparisons can be made are very small: 13 and 10 for UNDP and five and four for UNEP, for project outcomes and sustainability respectively (Table 19). In future, when more observations will be available better comparisons will be possible.

CHAPTER VIII: MANAGEMENT ACTION RECORDS

150. The Council approved the format and procedures for preparing the GEF Management Action Record (MAR) at its November 2005 meeting and requested the Secretariat and the GEF Evaluation Office to 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 the Implementing Agencies while at this juncture there is little involvement of the Executing Agencies.

151. Each MAR contains columns for recommendations, management responses, and Council decisions completed by the Office. Management is invited to provide a self-rating of the level of adoption of Council decisions on recommendations, and comments as necessary. Subsequently, the Office enters its own rating of adoption with comments in time for presentation to the Council. The ratings to assess the progress towards adoption the Council's decisions are the following:

- (a) **High** - Fully adopted;
- (b) **Substantial** - Largely adopted but not fully incorporated into policy, strategy or operations as yet;
- (c) **Medium** - Adopted in some operational and policy work, but not to a significant degree in key areas, and
- (d) **Negligible** - No evidence or plan for adoption, or plan and actions for adoption are in a very preliminary stage.

152. Management Action Records will be updated annually. After an item has been reported as fully adopted or not longer relevant, it will be deleted for the MAR and after all items have been adopted, the MAR will be archived.

153. In accordance with the procedures, the Office prepared draft MARs for reports that received a management response. These seven MARs were forwarded to the Secretariat on March 17, two months prior to the Council session. The 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 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 MARs has impaired the Office's ability to verify management's assessment of progress towards adoption of Council decisions.

154. The Office's assessment is in almost all cases indicative. In one case an exception must be made. GEF management assesses progress towards transparency in the GEF approval process as "medium", given the fact that the work towards establishing a new database for GEF projects has started. Our assessment of the adoption rate of the decision of Council on transparency in the system of June 2005 is based on corresponding evidence of the GEF Country Portfolio Evaluation and the on-going work in the Joint Evaluation of the GEF Activity Cycle and Modalities, as well as the consultative process. The Office concludes that the adoption rate has been negligible so far. The reality for project proponents on the country level has not changed. Information on where projects are in the process is still not available.

155. The Office believes that making information available in a transparent way is not rocket science, nor does it need to rely on new database software or systems. What is needed is discipline in gathering information and presenting it in a clear way on a website.

156. 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 to deal with the MAR. The result was that the Office did not have sufficient time to verify the ratings. Based on our knowledge through other sources and evaluations we have indicated the ratings that we believe would be justified. The Office will present the MARs to the GEF Management again in March 2007 and is confident that the second time around GEF Management will be able to deliver their own assessment of the adoption of Council decisions on evaluations in time to ensure that the Office can verify the ratings.

APPENDIX 1: RATINGS FOR THE ACHIEVEMENT OF OBJECTIVES, SUSTAINABILITY OF OUTCOMES AND IMPACTS, QUALITY OF TERMINAL EVALUATION REPORTS AND PROJECT M&E SYSTEMS

GEF Evaluation Office Ratings for the achievement of objectives, sustainability of outcomes and impacts, quality of terminal evaluation reports and project M&E systems

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”.

Criteria for the ratings on the outcomes²⁶

Based on the information in the report, the terminal evaluation review will make an assessment of the extent to which the project's major relevant objectives²⁷ 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 (i.e. original or modified project objectives²⁸)?
- C. **Efficiency:** Include an assessment of outcomes in relation to inputs, costs, and implementation times based on the following questions: Was the project cost –

²⁶ 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 which result from a development intervention; may also include changes resulting from the intervention which 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.

²⁷ 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.

²⁸ The GEF Office of Monitoring and Evaluation is currently working with the GEFSEC 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.

effective? How does the cost-time 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.

Criteria for the rating of impacts²⁹

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.

Criteria for the ratings on sustainability

Sustainability will be understood as the likelihood of continuation of project benefits after completion of project implementation³⁰. 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, socio-economic 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, socio-political, institutional frameworks and governance, and replication³¹. 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. Socio-political: 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 in their interest that the project benefits continue to flow? Is there

²⁹ 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.

³⁰ GEF Project Cycle. GEF/C.16/Inf.7. October 5, 2000.

³¹ 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, etc. GEF Project Cycle, GEF/C.16/Inf.7, October 5, 2000

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 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?
- 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:

$$\text{Sustainability rating}^{32} = (A+B+C+D)/4$$

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.

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 on-going or completed project, its design, implementation and results. Evaluation may involve the definition of appropriate standards, the examination of performance against

³² Note: For terminal evaluations reviewed in FY05 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.

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.³³

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 towards projects objectives. The tools used might include a base line, clear and practical indicators and data analysis systems, or studies to assess results were planned and carried out at specific times in the project.
- b. Whether the M&E system was used effectively for project management.

Rating system for the quality of project M&E systems

A	B
Effective M&E system in place (Indicators, baselines, etc.)	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 evaluations 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.

Criteria for the assessment of the quality of terminal evaluation reports

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

1. 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?
2. Was the report consistent and the evidence complete and convincing and were the ratings substantiated when used?
3. Did the report present a sound assessment of sustainability of outcomes?
4. Were the lessons and recommendations supported by the evidence presented?
5. Did the report include the actual project costs (total and per activity) and actual co-financing used?

³³ Glossary of Key Terms in Evaluation and Results Based Management. OECD – DAC. pp. 21 and 27.

6. Did the report include an assessment of the quality of the project M&E system and its use for project management?

Rating system for quality of terminal evaluation reports

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.

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

Then the quality of the terminal evaluations reports will be:

$$\text{Quality of the terminal evaluation report} = 0.3*(A + B) + 0.1*(C+D+E+F)$$

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

APPENDIX 2 METHODOLOGICAL BRIEF USED FOR THE REVIEW OF MONITORING ARRANGEMENTS AT ENTRY

Steps followed for the Review

- The Annual Performance Review Approach Paper was shared with the GEF Secretariat and the Implementing Agencies to receive their feedback on the draft framework for ‘assessment of the quality assurance system for M&E arrangements at entry in the GEF projects.’
- The Council Policies and the GEF guidance for project M&E systems at entry were reviewed to identify the present expectations for M&E arrangements at entry.
- The guidance provided to the GEF Secretariat project reviewers and the TOR’s for the STAP roster reviewers was reviewed to determine the expectations from them in addressing the M&E issues in their reviews.
- The ‘F’ test and Chi Square test were used to assess whether there is a significant difference in the performance among the focal areas and Implementing and Executing Agencies. Where ever differences in the performance have been cited, they are at a 90% confidence level.
- To assess the extent to which the expectations for M&E arrangements entry are being complied with all the 74 full size projects that received CEO Endorsement during FY 2005 were examined³⁴.
- The project reviews by the GEF Secretariat, the STAP roster reviewers and the GEF Council were examined to determine the extent to which the project review process is able to address the M&E issues.
- Interviews were conducted with the GEF Secretariat Focal Area Leads and M&E Coordinators from the Implementing Agencies to probe further the issues coming out of the reviews and to identify concerns.

³⁴ Three projects that were part of the studied cohort – the “Coral Reef Targeted Research and Capacity Building for Management” (International Waters), the “Support Program for National Capacity Self-Assessments” (Multi Focal Areas), and the “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 may also not be an effective management tool for such projects. However, a main reason to include these projects for assessments is to be consistent with the Terminal Evaluation Reviews (terminal evaluation reviews). Further, since there are only three such projects their inclusion will not substantially change the overall conclusions of the study.

- The preliminary findings of the review were disclosed to the GEF Secretariat, Focal Area Task Forces, and Implementing Agencies to verify factual accuracy of data and to identify possible methodological concerns.

Assessment of the quality of M&E Plans at Entry

- An instrument was developed to assess the quality of the M&E plans. This instrument measures 13 specific aspects (parameters) of M&E quality, which are based on the Review Criteria of the GEF Secretariat (2000) and the guidelines contained in the M&E Policies and Procedures (2002). In some cases, parameters outlined in these documents were refined to facilitate consistency and objectivity in the application of the assessment instrument³⁵. Certain technical or operational elements were not included in this instrument as this would have required specialized technical expertise on individual projects and also would have introduced greater subjectivity into the review process³⁶.
- The 13 parameters of M&E used in this review are:

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 or information been provided?
- Has a separate budget been allocated to 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 objective 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?
- Has the Baseline data collection methodology been explained?
- Have the responsibilities for the M&E activities been clearly specified?
- Have the time frames been specified for the M&E activities?
- Have the performance standards (targets) been specified in the logframe for the project outputs?

³⁵ 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.

³⁶ The parameters that were mentioned in the two documents but not used in the instrument include discussion on key assumptions of the project, sufficiency of M&E budget; and adaptive management.

- Do the project documents mention having made a provision for mid term and terminal evaluation?
- The information needed for rating on the 13 assessment parameters was gathered by examining the logframe (or results framework), M&E section, Appendix, budget tables, and other sections of the project documents that mention M&E. Each project’s performance on the 13 M&E parameters was then recorded and scored using the assessment instrument that is presented as Appendix 2A.
- The score on each individual parameter could range between one and three, where one was the minimum possible score and three was the maximum. The score of two corresponded with the bare minimum level of expected performance required for compliance in any given parameter. Since compliance on any parameter requires only a bare minimum level of performance, one would expect the M&E plans to score better than or at least equal to the required minimum standard.
- For a project to be considered acceptable on quality of M&E it should score two or more on each of the five parameters classified as critical. Although “other” parameters are also important, none of them is important enough that non-compliance on it alone would justify overall rejection of the M&E plan. However, if performance on many of such parameters is deficient then this is taken to indicate inadequate preparation of the M&E plan. Thus, for “other parameters” the emphasis is on the cumulative score rather than an individual pass/fail rating.
- The total score (summation of the scores of the project on all the parameters) for a project to be consider acceptable is 26 or more.

APPENDIX 2A: INSTRUMENT FOR ASSESSMENT OF M&E PLANS

S. No	Parameters	Response and Raw Score
1	Is there at least one specific indicator in the logframe for each of the project objectives and outcomes?	Yes.....3 No.....1
2	Are the indicators in the logframe relevant to the chosen objectives and outcomes?	Yes.....3 Yes, almost all are relevant.....2 No, most are irrelevant.....1
3	Are the indicators in the logframe sufficient to assess achievement of the objectives and outcomes?	Sufficient.....3 Largely Sufficient.....2 Some important indicators are missing.....1
4	Are the indicators for project objectives and outcomes quantifiable?	Yes.....3 Only some of them are.....2 No, or else it has not been shown how the indicators could be quantified.....1
5	Has the complete and relevant baseline information been provided?	Yes, complete baseline info provided.....3 Partial info but baseline survey in 1 st year.....2.5 No info but baseline survey in 1 st year.....2 Only partial baseline information.....1.5 No info provided.....1
6	Has the methodology for determining the Baseline	Yes.....3

	been explained?	No.....1
7	Has a separate budget been allocated to M&E activities?	Yes.....3 No.....1
8	Have the responsibilities been clearly specified for the M&E activities?	Yes, and clearly specified.....3 Yes, broadly specified.....2 No.....1
9	Have the time frames been specified for the M&E activities?	Yes, for all the activities.....3 Yes, but only for major activities2 No.....1
10	Have the performance standards (targets) been specified in the logframe for the project outputs?	Yes, for all the outputs.....3 Yes, but only for major outputs.....2 No.....1
11	Have the targets been specified for the indicators for project objectives and outcomes in the logframe?	Yes, for most.....3 Yes, but only for some indicators2 No1
12	Are the specified targets for indicators of project objective and outcomes based on initial conditions?	Yes, for most.....3 Yes, but only for some of the indicators.....2 No.....1
13	Do the project documents mention having made a provision for mid term and terminal evaluation?	Yes, both mid term and terminal evaluation.....3 Only terminal evaluation.....2.5 Only mid term evaluation.....1.5 No information provided.....1

APPENDIX 3: PERFORMANCE OF THE PORTFOLIO ON M&E ARRANGEMENTS AT ENTRY ON 13 PARAMETERS³⁷

- **M&E Criterion #1:** It is important to know whether there is at least one specific indicator for each objective or outcome listed in a given project logframe. 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% of the projects had a corresponding specific indicator for each of the objectives and outcomes, 43% lacked such indicators in one or more instances.
- **M&E Criterion #2:** 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. Presence of irrelevant indicators also indicates lack of clarity on how various project components will help in achieving the overall objectives of the project. In 78% of the projects all indicators were relevant and in the remaining 22% almost all of the listed indicators were relevant. Thus, all the projects complied on this parameter.
- **M&E Criterion #3:** It is essential that the specified indicators together are sufficient to help us know the extent to which project's overall objectives and outcomes have been achieved. For example, there may be instances where even though a project is missing a few specific indicators, or may have some indicators that are irrelevant, the set of indicators when taken together is sufficient. This is a critical parameter and for compliance on it the listed indicators should be sufficient or largely sufficient. For 47% of the projects the indicators were sufficient without any qualifications. For 28% – allowing some minor omissions – the indicators were largely sufficient. Thus, 76% of the projects were in compliance on this parameter³⁸.
- **M&E Criterion #4:** Specifying the indicators in a form that is quantifiable facilitates in establishing objective targets. For compliance on this parameter all or at least some of the indicators should be presented in quantifiable form. For 57% of the projects all the indicators and for 41% some of the indicators were quantifiable implying compliance by 97% of projects. The remaining 3% were in non compliance as none of their listed indicators was in a quantifiable form.
- **M&E Criterion #5:** Unless we know where we started, it is difficult to determine the progress. Baseline information, thus, forms a basis for determining progress. Therefore, this parameter has been identified as a critical parameter. Although there is a strong case for requiring the baseline information to be provided upfront, keeping in mind the difficulties and costs involved in establishing baseline conditions for very

³⁷ For details on each parameter see Annex II and Annex III. Also, Annex V lists performance of individual projects on each parameter.

³⁸ The numbers do not add up due to rounding off.

complex projects, the present Project Review Criteria 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 the first year of project implementation. Nineteen percent of projects provide the complete baseline information upfront; 20% provide baseline information on some indicators while promising to provide the remaining information within the first year; and, 53% projects just promise that they would provide the baseline information within the first year. Thus, 92% of the projects are in compliance on this parameter. The remaining 8% were in non compliance. This finding also needs to be seen in the light of the new M&E policy of GEF (2006). The policy calls for providing baseline information upfront except in “rare” situations wherein baseline information could be provided within the first year. Clearly the current situation, where 53% projects just promise to baseline information within first year without providing any baseline information, the exception is being made in more instances than what can be called “rare.”

- **M&E Criterion #6:** An explanation on how the baseline for an indicator will be determined helps in ascertaining whether or not the chosen baseline methodology is feasible. Since it requires a high degree of project specific technical expertise, the feasibility and technical merits of the given methodology were not assessed. However, the assessment instrument notes those cases in which a description of the baseline collection methodology has been provided. Eighty-four percent of the projects that provided some explanation of how baselines for indicators will be determined were in compliance whereas the remaining 16% of the projects that didn't provide any information were in non compliance.
- **M&E Criterion #7:** Allocation of sufficient budget to M&E activities is essential to ensure that M&E activities are not stalled for want of funds. How much budget will be sufficient for carrying out M&E activities satisfactorily is, however, dependent on factors such as size of the project, focal area, and institutional, local and technological context. Due to these differences, a great degree of variation may be expected across the projects. While it is difficult to determine whether or not the budget allocated to M&E is sufficient, in cases where no budget has been allocated to M&E it can be safely inferred that the financial support to M&E must be insufficient. This has been identified as a critical parameter and for compliance on it a project should make explicit provision for M&E activities in the budget. Ninety two percent of the projects explicitly allocate some budget to M&E activities, whereas 8% don't.
- **M&E Criteria #8 and #9:** For sound M&E planning and implementation, it is important to specify the responsibilities and time frames for each of the M&E activities. For compliance on these parameters responsibilities and timeframes for at least some of the M&E activities should be specified. Fifty-seven percent of projects clearly specify M&E responsibilities, 42% broadly specify them, and one project (1%) didn't specify them. A similar pattern was observed in terms of specifying the time frames for M&E: 57% of projects specify time frames for all M&E activities, 42% for some, and 1% for none. Thus, in both these parameters 99% of the projects complied with the Council expectations.

- **M&E Criterion #10:** Specification of targets for the outputs of the projects 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 projects provide targets for all the project outputs; 35% provide it for some, whereas 5% provide it for none. Thus, 95% projects were in compliance on this parameter.
- **M&E Criterion #11:** Whether or not a project achieves its desired results is dependent upon the *ex ante* expectations on 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 specify targets for all the indicators, 43% specify targets for some of the indicators, while 12% specified no targets. Thus, 89% of the projects were in compliance on this parameter.
- **M&E Criterion #12:** Realistic targets for indicators 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. In order to be realistic, the specified targets should be based on some assessment of the initial conditions and on the level of change that could be reasonably expected by the end of the project. 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 has been identified as a critical parameter and for compliance on it the specified targets for at least some indicators should be based on an assessment of the initial conditions. For 23% of the projects the targets for all the indicators and for 59% some of the indicators are based on an assessment of the initial conditions. Thus, 82% of the projects performed satisfactorily on this parameter. Nineteen percent of the projects, for whom none of the targets were based on an assessment of initial conditions, had unsatisfactory performance.
- **M&E Criterion #13:** The Review Criteria (2000) and the M&E 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 the GEF projects, and most of them also provide for mid-term reviews, whether or not these are mentioned in the project documents is more an indication of how well evaluation and review activities have been integrated into the M&E plans rather 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 mention that they will conduct both mid-term review and terminal evaluation; another 5% mention that they will conduct terminal evaluation. Thus, 78% of the projects are in compliance on this parameter.

APPENDIX 4: AGENCY PROJECT-AT-RISK MONITORING INVENTORY CARD

Agency Name: _____

Agency has project monitoring & 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 ___

Report includes judgment of performance of project components: Yes ___ No ___

Report includes judgment of project risk: Yes ___ No ___

Report assesses project management performance: Yes ___ No ___

Report assesses project financial management: Yes ___ No ___

Report assesses project M&E performance: Yes ___ No ___

Report tracks project disbursement history: Yes ___ No ___

Report documents any delays in project effectiveness, key events: Yes ___ No ___

Projects with performance problems or risks are Identified as at-risk or non-performing: Yes ___ No ___

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 ___

Follow-up on at-risk projects includes time-bound action plan: Yes ___ No ___

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.

Appendix 5: Project-at-Risk Inventory

Agency Project-At-Risk System Inventory

	ADB	AfDB	EBRD	FAO	IADB	IFAD	UNEP	UNDP	UNIDO	WB	TOTAL "YES"
Agency has project monitoring & reporting system	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	10
--If yes, system is electronic/MIS	X	X	X	X	X**	X	X	X	X	X	10
--If yes, system is paper-based											0
Reports required to be updated at least twice/yr	YES	NO	YES	YES	NO	NO	YES	YES	YES	NO	6
Report includes judgment of overall project performance	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	10
Report includes judgment of performance of project components	YES	NO	YES	YES	YES	NO***	YES	YES	YES	YES	8
Report includes judgment of project risk	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES****	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
Report assesses project M&E performance	YES	NO	YES	NO	YES	YES	NO	YES	NO	YES	6
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
Projects with performance problems or risks are identified as at-risk or non-performing	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	10
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
Follow-up on at-risk projects includes time-bound action plan	YES	NO*	YES	YES	NO	NO	YES	NO	NO	NO*****	4
At-risk projects are tracked by Agency management	YES	YES	YES	NO	YES	YES	NO	YES	YES	YES	8

	ADB	AfDB	EBRD	FAO	IADB	IFAD	UNEP	UNDP	UNIDO	WB	TOTAL "YES"
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	

Note: items shown in bold are Critical Elements for risk monitoring

* Only for countries with portfolio improvement plan

** Not all modules of MIS were operational at time of review

*** Reports include narrative/qualitative information on performance

**** As of FY05, risk rating is calculated by MIS, not by project supervision staff

***** This is considered good practice but not required

APPENDIX 6: LIST OF TERMINAL EVALUATION REPORTS REVIEWED AND GEF EO RATINGS

Project name	Implementing Agency	Outcomes	Sustainability	Quality of the Terminal Evaluation Report							Quality of project M&E system			Year effective or Prodoc signature	Year IA prepared the Terminal evaluation
				Overall rating	outcomes / objectives	Report consistency	Sustainability	Lessons	Actual costs and co-financing	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
Cote 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 Project (COREMAP I)	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	UNDP	MS	ML	MU	S	U	U	S	HU	S	HU	HU	UA	1995	2004

Biodiversity Protection															
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 Project (REIMP)	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
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 (PARC) 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 (Pilot Project in Rivne) Stage I	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 IW of the Pacific Small Islands and Development States (SIDS)	UNDP	S	MU	MS	MS	MU	S	U	S	S	U	U	HU	2000	2004
Regional (Bulgaria, Croatia, Hungary, Romania, Slovak Republic) - Transfer of Environmentally	UNDP	S	L	S	HS	HS	HS	MU	MU	HS	MS	MS	MS	2001	2005

Sound Technologies (TEST) to Reduce Transboundary Pollution in the Danube River Basin.															
Regional Africa - 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 (Mexico) - 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

APPENDIX 7: QUALITY OF TERMINAL EVALUATION REPORTS BY IMPLEMENTING AGENCY AND ASSESSMENT CRITERIA

Appendix 6A: Quality of Assessment of Outcomes in terminal evaluations

Quality of terminal evaluations	WB		UNDP		UNEP	
	2004	2005	2004	2005	2004	2005
HS	2	1	1	4	2	1
S	8	9	4	11	5	1
MS	7	2	3	7	0	3
<i>Sub Total</i>	<i>17</i>	<i>12</i>	<i>8</i>	<i>22</i>	<i>7</i>	<i>5</i>
MU	1	0	3	0	0	1
U	3	0	2	1	2	0
HU	0	0	0	0	0	0
<i>Sub Total</i>	<i>4</i>	<i>0</i>	<i>5</i>	<i>1</i>	<i>2</i>	<i>1</i>
Total	21	12	13	23	9	6

Appendix 6B: Quality of terminal evaluation Report in terms of being Consistent and Convincing

Quality of terminal evaluations	WB		UNDP		UNEP	
	2004	2005	2004	2005	2004	2005
HS	2	0	0	3	1	0
S	6	12	3	8	4	1
MS	7	0	4	8	0	2
<i>Sub Total</i>	<i>15</i>	<i>12</i>	<i>7</i>	<i>19</i>	<i>5</i>	<i>3</i>
MU	3	0	2	2	3	0
U	2	0	4	1	1	1
HU	1	0	0	1	0	2
<i>Sub Total</i>	<i>6</i>	<i>0</i>	<i>6</i>	<i>4</i>	<i>4</i>	<i>3</i>
Total	21	12	13	23	9	6

Appendix 6C: Quality of terminal evaluation Report in terms of Assessment of Sustainability

Quality of terminal evaluations	WB		UNDP		UNEP	
	2004	2005	2004	2005	2004	2005
HS	2	1	2	3	1	0
S	6	8	4	17	3	1
MS	6	3	2	1	4	2
<i>Sub Total</i>	<i>14</i>	<i>12</i>	<i>8</i>	<i>21</i>	<i>8</i>	<i>3</i>
MU	6	0	3	0	0	1
U	0	0	2	2	1	0
HU	1	0	0	0	0	2
<i>Sub Total</i>	<i>7</i>	<i>0</i>	<i>5</i>	<i>2</i>	<i>1</i>	<i>3</i>
Total	21	12	13	23	9	6

Appendix 6D: Quality of terminal evaluation in terms of Comprehensive Lessons Being Well Supported

Quality of terminal evaluations	WB	UNDP	UNEP

	2004	2005	2004	2005	2004	2005
HS	2	1	1	2	0	0
S	11	10	6	9	5	2
MS	4	1	1	8	2	1
<i>Sub Total</i>	<i>17</i>	<i>12</i>	<i>8</i>	<i>19</i>	<i>7</i>	<i>3</i>
MU	2	0	3	2	1	2
U	2	0	1	2	1	0
HU	0	0	1	0	0	1
<i>Sub Total</i>	<i>4</i>	<i>0</i>	<i>5</i>	<i>4</i>	<i>2</i>	<i>3</i>
Total	21	12	13	23	9	6

Appendix 6E: Quality of terminal evaluation Report in terms of providing Information on Actual Project Costs and Co-Financing

Quality of terminal evaluations	WB		UNDP		UNEP	
	2004	2005	2004	2005	2004	2005
HS	4	0	0	3	0	0
S	9	5	1	4	3	1
MS	6	5	4	1	1	3
<i>Sub Total</i>	<i>19</i>	<i>10</i>	<i>5</i>	<i>8</i>	<i>4</i>	<i>4</i>
MU	2	2	1	3	0	1
U	0	0	5	8	3	0
HU	0	0	2	4	2	1
<i>Sub Total</i>	<i>2</i>	<i>2</i>	<i>8</i>	<i>15</i>	<i>5</i>	<i>2</i>
Total	21	12	13	23	9	6