

GEF Council Meeting
June 5-7, 2012
Washington, D.C.

Agenda Item 7

Annual Performance Report 2011

(Prepared by the GEF Evaluation Office)

Recommended Council Decision

The Council, having reviewed document GEF/ME/C.42/01, “*Annual Performance Report 2011,*” and document GEF/ME/C.42/02, “*Management Response to the Annual Performance Report 2011,*” notes that evidence emerges that the GEF Agencies are starting to involve GEF Operational Focal Points in a more systematic manner in monitoring and evaluation. The Council requests the GEF Agencies to continue to enhance their efforts to specify how Operational Focal Points will be engaged, when feasible and relevant, in project or program monitoring and evaluation.

Table of Contents

Executive Summary	1
Background	2
Findings and Conclusions	3
Project Completion Delays	7
Performance Matrix	9
Quality at Entry	10
Alignment with Focal Area Result Frameworks	12
Engagement of Operational Focal Points in project M&E	12
New GEF Policy Documents: Gender Mainstreaming, Environmental and Social Safeguards and Council Decision on Climate Adaptation and Resilience	12
Preliminary Findings – Assessment of Entry Arrangements for Impact Evaluation in GEF Projects	13
Overall Rating for Arrangements for Impact Measurement at CEO Endorsement	13
Quality of Indicators	14
Baseline information	15
Quality of planning for data collection on impact indicators	15
Management Action Record	15
Recommendation	16

Executive Summary

1. This document is the eighth annual performance report (APR) of the Global Environment Facility (GEF) Evaluation Office. The report presents a detailed account of some aspects of project results, of processes that may affect these results, of monitoring and evaluation (M&E) arrangements in completed GEF projects, and also of quality of terminal evaluation reports. This APR also focuses on 109 completed projects for which terminal evaluations were submitted during the fiscal year 2011 (FY11).
2. The APR primarily reviews the evidence presented in the terminal evaluation reports, with verification of performance ratings based primarily on desk reviews. The evaluation offices of different agencies have been conducting similar reviews and their ratings have been accepted for 93 projects. Sixteen projects were reviewed by the GEF Evaluation Office.
3. The APR 2011 contains the following seven conclusions:
 - (a) Outcome achievements of 82 percent of completed projects reviewed for FY11 were rated in the satisfactory range.
 - (b) The level of cofinancing materialized, as reported by the GEF Agencies, is on average higher than the level of expected cofinancing at the time of project approval.
 - (c) Quality of M&E at project closure is fluctuating with an average of 68 percent of projects being rated moderately satisfactory or above since 2006.
 - (d) The quality of 84 percent of the terminal evaluations submitted during FY11 was rated moderately satisfactory or above.
 - (e) Eighty percent of projects endorsed by the CEO in FY 2011 are compliant with minimum requirements for quality at entry as measured by GEF4 standards. In comparison, 76 percent of projects endorsed by the CEO during FY 2008 met the same minimum requirements.
 - (f) GEF projects at entry demonstrate a high level of alignment between project logical frameworks and focal area results frameworks per the new requirement in the 2010 M&E Policy.
 - (g) GEF Projects are beginning to specify how Operational Focal Points will be informed and where feasible, involved in M&E activities.
4. Based on the analysis presented in the APR the following recommendation is made: GEF Agencies should continue to enhance their efforts to specify how Operational Focal Points will be engaged, when feasible and relevant, in project or program monitoring and evaluation.
5. The full APR 2012 report, containing detailed evidence for the findings in this Working Document, has been uploaded to the GEF Evaluation Office website at www.gefio.org.

Background

6. This document is the eighth annual performance report (APR) that the Evaluation Office presents. It includes a detailed account of some aspects of project results, of processes that may affect these results, and of monitoring and evaluation (M&E) arrangements. For the fifth time, a performance matrix is presented, which summarizes the performance of the GEF Agencies on various parameters.

7. The APR 2011 continues the annual presentation of the assessment of project outcomes, project sustainability, project completion delays, materialization of cofinancing, and quality of monitoring in completed projects.

8. This year's cohort comprises 109 terminal evaluations received by the Evaluation Office. In total, 495 project terminal evaluations, representing \$1,817 million in actual GEF funding (compared to \$2,196 million of expected funding), have been submitted to the GEF Evaluation Office. These include terminal evaluations for 109 projects submitted to the Office in FY2011 that account for \$444 million in GEF funding¹.

9. For the **assessment of project outcomes, project sustainability, and delays in project completion**, 419 projects, for which terminal evaluation reports were submitted by the GEF Agencies to the Evaluation Office since FY 2005, have been considered.

10. For reporting on **materialization of cofinancing**, 413 projects for which terminal evaluation reports have been submitted since FY2005 were considered. Of these, information on materialization of cofinancing has been reported for 355 projects (86 percent). The GEF has invested a total of \$1,423 million in these 355 projects; the implementing agencies reported that a cofinancing of \$4,518 million materialized during implementation. The Office has reported actual cofinancing amounts for 102 projects of the FY2011 cohort, representing a total cofinancing amount of \$2,050 million (based on TE and Agency reports).

11. The **performance matrix** provides a summary of the performance of the GEF Agencies on relevant parameters. The matrix is limited to UNDP, UNEP and the World Bank because the completed projects from the other GEF agencies are still not significant in numbers. Most of the parameters included in the matrix are assessed on an annual basis by the Evaluation Office whereas others are tracked after two-to-three interval years. This year, information has been presented on ten parameters.

12. The **quality at entry** section presents findings from three reviews carried out by the Evaluation Office for APR2011. The first review is a re-assessment of the quality of M&E arrangements at the point of projects' entry into the GEF system. Similar reviews were carried out in 2005 and 2008. This study assesses quality at entry for a sample of 80 out of 137 full-sized projects that received CEO endorsement during FY2011. The 80 full-sized projects and programs account for US\$ 338 million in GEF investment grants. The second review tracks parameters aligned with new requirements in the GEF's 2010 M&E Policy, the GEF Policy on Gender Mainstreaming, Policy on Environmental and Social Safeguard Standards and the Council decision concerning the Mainstreaming and Targeting of Adaptation and Resilience. The reviews instrument was applied to

¹ All currency figures reported in this document are in US dollars.

the full FY2011 cohort of 137 FSPs. The third review pertains to in-depth technical assesses the project and program arrangements for impact evaluation at entry that the Evaluation Office carried out in collaboration with the GEF STAP. This review included a representative sample of 55 full size projects that were endorsed by the CEO of the GEF in FY 2011.

13. This year's **management action record (MAR)** tracks the level of adoption of 12 Council decisions based on 9 evaluation reports. The Evaluation Office was able to verify 10 of these decisions. Two have been graduated for having achieved a "high" adoption rating, and will not be tracked in the next MAR.

14. The APR primarily involves review of the evidence presented in the terminal evaluation reports, with verification of performance ratings based primarily on desk reviews. The evaluation offices of several agencies have been conducting similar reviews for past couple of years. Since FY2007 the UNEP's Evaluation Office has been providing performance ratings for all the completed GEF projects. Similarly, The World Bank's Independent Evaluation Group (IEG) conducts desk reviews of all the terminal evaluation reports produced by the management for the full size projects and conducts more intensive field verifications for a sample of these projects. In FY2009 UNDP's evaluation office began to review terminal evaluation reports and continued to do so in FY2010 & FY2011.

15. In FY2010 the GEF Evaluation Office accepted the ratings provided by the evaluation offices of UNDP, UNEP and the World Bank. Consistency between GEFEO ratings and ratings of the evaluation offices of the three Agencies had been established and therefore ratings from these agencies were accepted for 93 projects in the FY2011 cohort. To ensure continued compatibility, the Office will continue to review a sample of terminal evaluations from all Agencies as warranted in the coming years.

Findings and Conclusions

Conclusion 1: Outcome achievements of 82 percent of completed projects reviewed for FY2011 were rated in the satisfactory range.

16. The Evaluation Office rated the achievement of project outcomes on criteria of relevance, effectiveness, and efficiency. The key findings of this assessment are as follows:

- Considering the FY2005 to FY2011 cohorts, so far 414 projects have been rated on outcome achievements. Of these, outcome achievements of 83 percent have been rated moderately satisfactory or above. Of the total GEF investment in rated projects, 82 percent is in projects that were rated moderately satisfactory or above.
- Of the 109 projects of the FY2011 cohort, 82 percent (89 projects), were rated moderately satisfactory or above. This is lower than the FY2010 figure of 92 percent.
- The factors contributing to the decline in the average rating of outcomes this year compared to the past two years are not clear. Given that the number and type of TE reports submitted varies from year to year, this decline should not yet be interpreted as a trend.
- Of the total \$444 million GEF investment in the rated projects of FY2011, 80 percent, or \$355 million, was allocated to projects that were rated moderately satisfactory or above.

17. Within the FY2011 cohort, 86 percent of UNEP and 90 percent of UNDP implemented projects received outcome ratings of moderately satisfactory or above. Of the World Bank implemented projects, 68 percent were rated moderately satisfactory or above.
18. As seen in Table 1.1, within UNEP's portfolio of 7 projects in the FY2011 cohort, 14 percent (1) were rated highly satisfactory, 28 percent (2) rated satisfactory, and 43 percent (3) were rated moderately satisfactory. Of the 61 UNDP implemented projects, 44 percent were rated moderately satisfactory, 43 percent satisfactory and 3 percent highly satisfactory. For the World Bank implemented projects, 30 percent were rated moderately satisfactory, 35 percent satisfactory and 3 percent highly satisfactory. The complete disaggregated ratings for each agency for the period FY2008 to FY2011 are presented in Table 3.4. This performance does not indicate a trend.

Table 1.1: Trends in Outcome Ratings by Agency

<i>Fiscal Year</i>	<i>FY2008</i>	<i>FY2009</i>	<i>FY2010</i>	<i>FY2011</i>
World Bank				
<i>Number of Projects</i>	23	17	22	40
Highly Satisfactory	4%	6%	0%	3%
Satisfactory	65%	47%	27%	35%
Moderately Satisfactory	9%	41%	55%	30%
Moderately Satisfactory or Above	78%	94%	82%	68%
UNDP				
<i>Number of Projects</i>	29	22	17	61
Highly Satisfactory	7%	0%	18%	3%
Satisfactory	49%	55%	35%	43%
Moderately Satisfactory	24%	27%	47%	44%
Moderately Satisfactory or Above	80%	82%	100%	90%
UNEP				
<i>Number of Projects</i>	8	15	9	7
Highly Satisfactory	0%	7%	0%	14%
Satisfactory	38%	67%	44%	29%
Moderately Satisfactory	50%	26%	56%	43%
Moderately Satisfactory or Above	88%	100%	100%	86%

19. **Sustainability of project outcomes** is rated based on the level of risks to sustainability of outcomes on four dimensions: financial, sociopolitical, institutional and governance, and environmental. 104 projects from the FY2011 cohort were rated on overall sustainability of outcomes². (Table 1.2). The key findings of this assessment are:

- Of the 419 terminal evaluation reports submitted from FY2005 to FY2011, sustainability ratings have been provided for 392 (94 percent) of the projects representing \$1,696 million in GEF investment. Since FY2005, of the 392 projects rated on sustainability, 237 projects (60 percent) were rated moderately likely or above. In terms of GEF funding for the 392

² Twelve projects had no overall sustainability rating in the Agency TER. The overall rating was applied by GEFO using the Office's TE review guidelines

projects, an amount of \$1,050 million (62 percent) out of approximately \$1,696 million of funding was in the 237 projects with sustainability ratings of moderately likely or above.

- Of the four risks to sustainability dimensions, financial risks were the most frequently cited threats to outcome sustainability. Of the 65 projects which had ratings on all four dimensions, the outcomes of 30 percent of projects were unlikely or were moderately unlikely to be sustained due to financial risks. The outcomes achievements of 19 percent were unlikely or moderately unlikely to be sustained due to socio-political risks. Twenty (20) percent were unlikely or moderately unlikely to be sustained due to institutional and governance risks. And, 17 percent were unlikely or moderately unlikely to be sustained due to environmental risks.
- Out of the total GEF investment in FY2011, 59 percent (\$259 million out of a total US\$ 437 million invested in projects rated on sustainability) went to 61 projects rated moderately likely or above in terms of the sustainability of their outcomes. This is a change from the FY2010 level of 73 percent but only slightly lower than the long-term average (2005-2011) of 62 percent.

Table 1.2: Outcomes and Sustainability of Outcomes

<i>Fiscal Year</i>	<i>2005</i>	<i>2006</i>	<i>2007</i>	<i>2008</i>	<i>2009</i>	<i>2010</i>	<i>2011</i>
Number of TE submitted ³	40	66	41	60	55	48	109
Number of projects with outcomes rating	38	64	40	60	55	48	109
Percentage rated MS or above in outcomes rating	82%	84%	73%	81%	91%	92%	82%
Number of projects with sustainability of outcomes rating	38	54	39	58	51	48	104
Percentage rated ML or above in sustainability of outcomes rating	47%	65%	59%	59%	71%	63%	59%
Number of projects rated both on outcomes and sustainability of outcomes	38	54	39	58	51	48	104
Percentage of projects rated both MS/ML or above	42%	61%	51%	57%	67%	63%	56%

20. The Evaluation Office assesses the extent to which projects that were rated moderately satisfactory or above in achievement of outcomes were also rated moderately likely or above in sustainability of outcomes (Table 1.3). It found:

- Considering both outcome and sustainability ratings together, of the 392 projects rated to date, 224 projects (57 percent) are rated both moderately satisfactory or above in terms of outcomes *and* moderately likely or above in terms of sustainability. Regarding GEF investment in the period FY2005 to FY2011, \$979 million (58 percent) out of \$1,696 million was in 224 projects that were rated both MS or above in terms of outcome achievements *and* ML or above in terms of sustainability.
- In the FY2011 cohort, 58 out of 104 projects (56 percent) were rated both MS or above on outcomes *and* ML or above on sustainability. Fifty-four percent (\$236 million out of \$437

³ Due to resubmission of three TEs, the number of reports in years 2005 and 2008 has changed from the numbers published in APR 2010. The percentages for these years have been updated accordingly.

million) was invested in the 58 projects rated MS or above for outcomes and ML or above for sustainability.

Table 1.3: Outcomes and Sustainability of Outcomes

Fiscal Year	FY2005	FY2006	FY2007	FY2008	FY2009	FY2010	FY2011	Total
Number of Projects Rated ⁴	38	54	39	58	51	48	104	392
Category	by Number of Projects							
Sustainability of Outcomes rated ML or above	47%	65%	59%	59%	71%	63%	59%	60%
Outcomes rated MS or above & Sustainability of Outcomes rated ML or above	42%	61%	51%	57%	67%	63%	56%	57%
	by GEF Investment							
Sustainability of Outcomes rated ML or above	65%	60%	55%	60%	64%	73%	59%	62%
Outcomes rated MS or above & Sustainability of Outcomes rated ML or above	60%	56%	44%	58%	63%	73%	54%	58%

Conclusion 2: The level of cofinancing materialized, as reported by the GEF Agencies, is on average higher than the level of cofinancing expected at the time of project approval.

21. The GEF Council views cofinancing as an indicator of a project's sustainability⁵, country ownership⁶, and mainstreaming⁷ of GEF activities in activities of the partner institutions, and as a way to mobilize additional resources for the global environment⁸. As stated in last year's APR, the Council has shown continued preference for a higher overall cofinancing ratio for the GEF project portfolio. The Secretariat publications also portray cofinancing as an indicator of the effectiveness of GEF in mobilizing additional resources for generation of global environmental benefits. A few documents clearly state that the high cofinancing ratio achieved is an indicator of GEF's "multiplier" effect in generating additional resources for generation of global environmental benefits⁹.

22. Both promised- and actual cofinancing amounts per dollar of GEF grant have been increasing since FY 2005. For the FY 2011 cohort, the overall ratio of promised cofinancing to the GEF grant amount at approval was 2.7. The overall ratio of actual, or materialized, cofinancing to the GEF grant amount at approval is higher at 4.6 dollars per dollar of GEF grant. The GEF Evaluation Office has not verified these amounts.

⁴ Due to resubmission of three terminal evaluation reports, the number of reports in years 2005 and 2008 has changed from the numbers published in APR 2010. For FY2005: from 41 to 40 and for FY2008: from 62 to 60. The % for these years has been updated accordingly.

⁵ Highlights of Council Discussions, Agenda Item 9 – Work Program Submitted for Council Approval, Dec. 1999.

⁶ Highlights of Council Discussions, Agenda Item 11 – Strategic Approach to Enhance Capacity Building, Nov. 2003.

⁷ Discussions of the Council, Agenda Item 7 – Work Program, April/May 1997.

⁸ Joint Summary of Chairs: Highlights of Council Discussions – Agenda item 9, Cofinancing, October 2002.

⁹ GEF Contributions to Agenda 21, GEF, June 2000 (page 15); GEF-Dynamic Partnerships: Real Solutions, GEF, 2002 (page 6); GEF Annual Report 2005, GEF, 2005, (page 13).

23. The FY 2011 overall percentage of actual cofinancing to promised cofinancing is 171 percent, indicating that actual cofinancing was 71 percent higher than expected. This percentage is significantly higher than the FY 2009 - FY 2010 average of 131 percent and than the 98 percent calculated for the period FY 2005 to FY 2008.

24. While a detailed analysis of cofinancing is not presented in the 2011 APR, a detailed discussion on cofinancing in the Annual Performance Report 2009 indicated that reported cofinancing must be carefully interpreted. The analysis showed that although the quality of reporting on cofinancing both in project proposals and in terminal evaluations is improving, reporting on cofinancing by Agencies and in the TE reports is not always consistent with the agreed definition of cofinancing¹⁰.

25. Further, usage of a cofinancing ratio has its pitfalls as this ratio is prone to be skewed by outliers. Of the 419 projects for which terminal evaluations have been submitted since FY 2005, agencies have reported data on materialized cofinancing for 355 (85 percent). Furthermore, the top 5 percent of projects of these 355 accounted for close to 33 percent of the total reported materialized cofinancing. The jump in actual co-financing amounts for FY 2011 is due to eight UNDP projects.

Project Completion Delays

26. The Evaluation Office began tracking project completion delays (expected completion date and actual completion date) in FY 2005. Of the 389 projects for which this data is available, 20 percent were completed after a delay of at least two years and 11 percent after a delay of three years or more. Of the projects in the FY 2011 cohort, 21 percent (22 projects) were completed within the expected implementation period, 25 percent were completed after a delay of one to two years, 5 percent after a delay of two to three years, and 14 percent after a delay of more than three years.

Conclusion 3: Quality of M&E of closed projects is fluctuating with an average of 68 percent of projects being rated moderately satisfactory or above since 2006.
--

27. Out of 303 projects rated on quality of monitoring and evaluation since FY2006, 68 percent (208 projects) were rated moderately satisfactory or above. Among the projects of FY2011 cohort, 70 percent (64 out of 92 projects for which we have M&E assessments) were rated moderately satisfactory or above in terms of monitoring during implementation (Table 1.4). There is no difference among FSPs and MSPs. The result for FY2011 is not significantly different from the long-term average and presents no change in relation to FY2010. Fluctuations are present in previous years ratings as well.

Table 1.4 Quality of project monitoring and evaluation

¹⁰ The agreed definition, adopted in 2003 (GEF/C.20/6/Rev.1), points out several key conditions for a contribution to be considered cofinancing: (i) a contribution should be managed with the GEF allocation; (ii) should be part of the initial financial package; and, (iii) should be essential to achievement of GEF objectives. In the 2010 an information document was presented to Council (GEF/C.39/Inf.3, pgs. 36-37), which states that “Only those cofinancing resources that are committed as part of the initial financing package are termed “Cofinancing Resources.” Government cofinancing in this context is defined as commitments for baseline activities upon which the project would be built and without which the project could not be implemented successfully. The GEF Evaluation Office uses the definition in the 2003 Council working document.

Fiscal Year of TE submission	2004	2005	2006	2007	2008	2009	2010	2011
Terminal evaluation reports submitted	42	40	66	41	60	55	48	109
Projects rated on M&E Implementation	29	32	46	33	49	39	44	92
Percentage of projects rated moderately satisfactory or above	55%	66%	78%	61%	71%	62%	70%	70%

28. The percentage of projects rated moderately satisfactory or above on the quality of M&E at entry changed in relation to last year's cohort, from 73 percent to 66 percent. For M&E during implementation, there is also a change in the satisfactory range assessment, 75 percent in FY 2011 in relation to the last cohort (71 percent).

29. Consistent with previous years, there is a strong relationship between quality of M&E arrangements at entry and actual quality of monitoring during implementation. Of the completed projects from the FY2011 cohort that were rated both on quality of monitoring during implementation and quality of M&E at entry, only 48 percent of those rated in the unsatisfactory range on quality at entry were rated in the satisfactory range during implementation. In contrast, of those rated in the satisfactory range on quality at entry, 85 percent were also rated in the satisfactory range during implementation.

Conclusion 4: The quality of 84 percent of the terminal evaluations submitted during FY2011 was rated moderately satisfactory or above.

30. Since 2004, when the Evaluation Office first started rating the quality of terminal evaluation reports, 495 terminal evaluations have been submitted and 453 rated. Of these, the quality of 87 percent of reports has been rated moderately satisfactory or above (Table 1.5). Fifty-five (55) percent of the reports achieved a higher rating of satisfactory or above. The figures for the FY2011 cohort are 84 percent and 53 percent respectively. This is similar to last year's cohort, in which 87 percent of the terminal evaluation reports were rated moderately satisfactory or above and 58 percent were rated satisfactory or above.

Table 1.5 Percentage of Terminal Evaluation Reports Rated MS or Above, by Year of Submission

Number	42	40	66	41	60	55	45	104	453
Agency	FY2004	FY2005	FY2006	FY2007	FY2008	FY2009	FY2010	FY2011	All years
ADB	NA	NA	NA	NA	NA	NA	NA	100%	100%
UNDP	54%	91%	87%	94%	93%	95%	88%	80%	86%
UNEP	67%	50%	67%	100%	100%	100%	100%	100%	85%
WB	80%	100%	86%	94%	88%	94%	82%	88%	88%
All Agencies	69%	88%	83%	95%	92%	96%	87%	84%	87%

31. For the FY 2011 cohort, data was available on the time lags between project closure and terminal evaluation report submission for 108 projects. Figure 6.5 below shows that for the FY 2011 cohort, 42 percent of terminal evaluation reports were submitted within 12 months of operational project closure, and 31 percent were submitted within one to two years following project closure.

Performance Matrix

32. The performance matrix provides a summary of the performance of the GEF Agencies covering key areas such as results, processes affecting results and M&E. Although several parameters included in the matrix are assessed by the Evaluation Office on an annual basis, to mitigate fluctuations in performance ratings due to differences in project mix, the values presented in the matrix are, depending on the parameter, running averages of two to four years. Of the 10 parameters included in the performance matrix this year, information is provided on all 9 (see table 7.1).

33. The **project outcome** figures included in the matrix are based on the four-year running averages. The figures are based on the information provided in the terminal evaluation reports. Of the projects for which terminal evaluations have been submitted since FY 2008, the Evaluation Office rated outcome achievements (parameter 1) of 85 percent to be moderately satisfactory or above. The outcome achievements of 77 percent of the World Bank projects, 95 percent of UNEP projects and 88 percent of UNDP projects have been rated in the satisfactory range.

34. For **quality of supervision and adaptive management** (parameter 2), ratings are provided based on the findings of the follow-up assessments of project supervision last presented in the APR 2009, and figures remain valid until the next study is conducted, which is expected to take place in FY 2012. The projects considered for this assessment were under implementation during FY 2007 and FY 2008. The Office had rated overall quality of supervision and adaptive management of 85 percent to be moderately satisfactory or above. In the breakdown by Agency, 86 percent of the World Bank projects, 73 percent of UNEP projects and 92 percent of UNDP projects have been rated in the satisfactory range.

35. The reported materialization of **cofinancing** per dollar of approved GEF financing (parameter 3a) has achieved a ratio of 3.3 in the overall performance of GEF Agencies, varying from 5.3 for UNDP to 1.2 for UNEP and 2.1 for the World Bank. In percentage of promised cofinancing (parameter 3b), the overall performance of GEF Agencies reached 142 percent, being 202 for UNDP, 131 for UNEP and 87 for the World Bank. The analysis is based on the information provided by the Agencies in the terminal evaluation reports or through other communications. These figures have not been verified.

36. The **project preparation elapsed time** (parameter 4) consists of figures indicating the number of months required to prepare projects. The APR 2011 will not present figures yet, as they will be updated for OPS5 with data provided by the Agencies and the GEF Secretariat database.

37. Regarding **implementation completion elapsed time** (parameter 5), the Evaluation Office through the terminal evaluation review process tracks the average difference in months between expected and actual project completion for the cohort. Since FY 2008, the overall average difference for the three major Agencies is of 16 months, with UNDP averaging a difference of 19 months, UNEP 15 months, and the World Bank 12 months.

38. Regarding **independence of terminal evaluations or independent review of terminal evaluations** (parameter 6), the ratings are based on the extent to which Agencies have put in place

a system to independently recruit evaluators, and track and report on the quality of terminal evaluations. This allows the Office to assess the extent to which systems in the partner Agencies are conducive to unbiased and candid terminal evaluations. The independence of the terminal evaluations submitted by UNDP and UNEP is now rated highly satisfactory for both FSP and MSPs. The independence of World Bank terminal evaluations is rated highly satisfactory for FSPs and considered not applicable to MSPs, the latter which are not reviewed by the World Bank's Independent Evaluation Group (IEG) but by the GEFEO instead as per an agreement with the IEG.

39. For **realism of risk assessment (robustness of project-at-risk systems)**, ratings are **provided** based on the findings of the follow-up assessments of candor and realism last presented in the APR 2009 (parameter 7). The projects considered for both assessments were under implementation during FY 2007 and FY 2008 project supervision last presented in the APR 2009. The Office rated this overall assessment as 77 percent to be moderately satisfactory or above. In the breakdown by Agency, 80 percent of the World Bank projects, 73 percent of UNEP projects and 77 percent of UNDP projects have been rated in the satisfactory range. Figures remain valid until the next specific study is conducted, which is expected to take place for APR2012.

40. The review of **quality assurance of project M&E arrangements at entry** (parameter 8) is elaborated upon in section 8.1, and presents the results of a study on the quality of project M&E at entry for projects endorsed by the CEO in FY 2011. Overall 80 percent of the projects reviewed were compliant with critical parameters. The compliance rate on the same parameters in the previous review of quality of project M&E arrangements conducted for projects endorsed by the CEO in FY 2008 was 76 percent and in 58 percent in FY 2005.

41. The ratings for **quality of project M&E arrangements during implementation** (parameter 9) are also provided for terminal evaluations submitted since FY 2008. The figures are four-year averages and are based on the information provided in the terminal evaluation reports. The Office rated M&E during implementation as satisfactory or above at 70 percent. UNDP's cohort demonstrated the largest change, from 60 percent to 73 percent.

42. For **quality of project terminal evaluations** (parameter 10), the ratings cover terminal evaluations since FY 2010. The figures are exceptionally two-year averages, and are based on the terminal evaluation review process. The quality of all the terminal evaluation reports submitted by UNEP has been rated in the satisfactory range. The quality of the World Bank terminal evaluations is 85 percent. For UNDP the average for quality of terminal evaluations changed by 10 percent to 82 percent compared to the APR 2010 average.

Quality at Entry

Conclusion 5: Eighty percent of projects endorsed by the CEO in FY 2011 are compliant with minimum requirements for quality at entry as measured by GEF4 standards. In comparison, 76 percent of projects endorsed by the CEO during FY 2008 met the same minimum requirements.

43. The GEF Evaluation Office carried out an assessment of the quality of M&E arrangements of FSPs that were endorsed by the CEO during FY 2011 as a follow-up to assessment carried out during FY 2008 and FY 2005. An identical methodology was used in all three assessments. In all,

the Evaluation Office has assessed 236 FSPs for quality of M&E arrangements at entry: 74 from FY 2005, 82 from FY 2008 and 80 from FY2011.

44. The Evaluation Office rated the M&E arrangements of 80 percent of 80 FSPs that were CEO endorsed in FY 2011 to be in compliance with the minimum criteria of the 2006 M&E Policy. In FY 2008 76 percent of the projects were rated as compliant with the same requirements (Table 1.6). In comparison to the three original GEF Implementing Agencies, the Executing Agencies, as a group (ADB, AfDB, EBRD, IADB, IFAD FAO and UNIDO) have a significantly lower rate of compliance, 58 percent. Amongst the focal areas, the multi-focal area has the lowest rate of compliance at 33 percent. (Annex H).

Table 1.6 Trends in compliance with minimum requirements for M&E at Entry

		2005 (n = 74)	2008 (n = 82)	2011 (n = 80)
Minimum Criteria	Are the indicators in the logframe relevant to the chosen objectives and outcomes?	100%	100%	99%
	Are the indicators in the logframe sufficient to assess achievement of the objectives and outcomes?	76%	94%	96%
	Have the targets been specified for the indicators for project objectives and outcomes in the logframe?	89%	99%	100%
	Are the specified targets for indicators of project objective and outcomes based on initial conditions?	82%	94%	99%
	Has the complete and relevant baseline information been provided?	92%	87%	85%
	Has a separate budget been allocated to M&E activities?	92%	95%	97%
	Compliance Rate	58%	76%	80%

Conclusion 6: GEF projects at entry demonstrate a high level of alignment between project logical frameworks and focal area results frameworks per the new requirement in the 2010 M&E Policy.

45. In this review parameters were introduced which provide a baseline for the Evaluation Office regarding the state of affairs at the end of GEF4, and as it begins to track changes to requirements as set by new policies and decisions adopted by the GEF council for GEF5. To incorporate considerations within the GEF's 2010 M&E Policy the review instrument was updated to include minimum requirements from the revised M&E Policy¹¹ as well as from other new GEF policy documents, namely the Policy on Gender Mainstreaming, Policy on Environmental and Social Safeguard Standards¹² and the Council decision concerning the Mainstreaming and Targeting of Adaptation and Resilience¹³. The full FY2011 cohort of 137 FSPs was reviewed for these

¹¹ The GEF Monitoring and Evaluation Policy. November 2010. GEF/ME/C.39/6/Rev.1

¹² Both policies were approved at 40th meeting of GEF Council. May 2011. GEF/C.40/10/Rev.1.

¹³ This decision comes from the GEF Council at its 39th meeting stemming from "Evaluation of the GEF Strategic Priority for Adaptation (GEF/ME/C.39/4) requesting tools to ensure the mainstreaming and targeting of adaptation and resilience.

parameters. As these policies were not in effect at the time when the majority of these projects were initiated, the GEFO assessment includes no evaluative judgments.

Alignment with Focal Area Result Frameworks

46. In the context of the Results-based Management Framework of the GEF, focal area results frameworks were developed for GEF5 that outline each focal area's strategic objectives, expected outcomes and indicators, outcome targets, and core outputs. For this review, project results frameworks (i.e. logframes) were assessed to determine if indicators from the focal area results framework have been included in project M&E design, to facilitate portfolio level reporting. The Evaluation Office rated the linkage between project logical framework indicators and focal area results frameworks at 96 percent for compliance. The CEO Endorsement Request template includes a section asking for a description of the project's consistency with GEF strategies and strategic programs, which may account for the high level of compliance on this parameter. In addition, some focal area indicators have long been applied in the portfolio – for example, the Protected Area Management Effectiveness Tracking Tool (METT), which assesses the effectiveness of protected area management, has been widely utilized in the GEF's biodiversity portfolio for years, and thus was not a new development for GEF5. Other similar tools may have been informally applied in other focal areas.

Conclusion 7: GEF Projects are beginning to specify how Operational Focal Points will be informed and where feasible, involved in M&E activities.

Engagement of Operational Focal Points in Project M&E

47. The engagement of Operational Focal Points (OFPs) is a new requirement in the M&E policy and intended to reflect the efforts of countries to establish or improve national monitoring and evaluation, and include an emphasis on increased country ownership. As such it mandates that GEF Operational Focal Points should be fully consulted with and informed on the planning, conduct and results of any evaluation activity in the country, including project based activities.

48. Of the 137 FSPs endorsed by the CEO in the FY11, 24 (18 percent) already address the role of OFP by discussing how they would be updated and involved on the implementation and results management. Of those 24 projects, nine are from UNDP, seven are from UNIDO, two each from the WB and FAO and one each from UNEP, IFAD, ADB and IADB.

New GEF Policy Documents: Gender Mainstreaming, Environmental and Social Safeguards and Council Decision on Climate Adaptation and Resilience

49. The 40th GEF Council approved in May 2011 the GEF Policy on Gender Mainstreaming and the GEF Policy on Environmental and Social Safeguard Standards. At the 39th Meeting in November 2010, the Council issued a decision concerning the Mainstreaming and Targeting of Adaptation and Resilience. Parameters were introduced in the review of the full FY 2011 cohort of 137 FSPs to assess the *current* provision of information *only* in the CEO Endorsement document on these *future* requirements. These parameters were not reviewed for compliance but rather to give the office a view on the present status of information in the CEO Request for Endorsement.

50. There is currently no requirement for provision of information on gender mainstreaming and monitoring of environmental and social safeguards in the Request for CEO Endorsement. The

endorsement request does contain a section on climate risks. Nonetheless, the information contained therein is inconclusive and not necessarily reflective of the full body of material that may be available in different source documents. For example, additional information on gender mainstreaming may be in project approval documents. Concerning safeguards monitoring, Agencies have separate systems to track environmental and social risk mitigations which may not be reflected in descriptions about the project or its results monitoring.

51. In the context of the fifth overall performance study (OPS5), the GEF Evaluation Office will undertake further investigation of these parameters. This assessment will include a review of additional project documents and the systems that Agencies have in place to demonstrate gender mainstreaming and track environmental, social and climate risks.

Preliminary Findings – Assessment of Entry Arrangements for Impact Evaluation in GEF Projects

52. For this assessment, a representative sample of 55 projects that were endorsed by the CEO of the GEF in FY2011 was drawn using a stratified random sampling approach (Annex J). Each focal area was considered as a separate stratum. Of the 55 sampled proposals, 6 were identified to have little direct or proximate links with environmental stress reduction so were excluded from the review. The reporting in this section is based on the 49 project proposals that were reviewed. The findings presented in this report are preliminary as they only include desk reviews. Interviews with representatives of Agencies and the GEF Secretariat are yet to be carried out. The final report of the review will be presented in the Annual Impact Report 2012, which will be presented to Council in November 2012.

Overall Rating for Arrangements for Impact Measurement at CEO Endorsement

53. The review rated overall quality of impact measurement arrangements specified in project proposals to be moderately satisfactory or above for 69 percent of proposals. Forty nine percent of the proposals met a more stringent yardstick of “satisfactory” or above (Table 1.7).

Table 1.7: Overall quality at entry of arrangements for impact measurement

Quality of arrangements for impact measurement	Frequency	Percentage (Probability Adjusted)¹⁴
<i>Highly Satisfactory</i>	<i>7</i>	<i>15</i>
<i>Satisfactory</i>	<i>16</i>	<i>35</i>
Satisfactory or Above	23	49
<i>Moderately Satisfactory</i>	<i>9</i>	<i>20</i>
Moderately Satisfactory or Above	32	69
<i>Moderately Unsatisfactory</i>	<i>11</i>	<i>21</i>
<i>Unsatisfactory</i>	<i>4</i>	<i>8</i>
<i>Highly Unsatisfactory</i>	<i>2</i>	<i>2</i>
Total	49	100%

54. In general, project proposals that were endorsed as part of a programmatic approach tended to have lower quality-at-entry rating for impact measurement arrangements than those that were not part of such an approach. For example 58 percent of the proposals endorsed as part of a programmatic approach were rated moderately satisfactory or above and 38 percent were rated satisfactory or above. By way of comparison, of the proposals not endorsed under programmatic approach, 79 percent were rated moderately satisfactory or above and 59 percent were rated satisfactory or above.

55. For proposals endorsed under the programmatic approach, the time taken from the point of a project proposal's approval to its CEO endorsement appears to be linked with differences in ratings. However, this relationship is neither strong nor linear. The proposals that take too much time (> three years) or too little time (<18 months) from the point of approval to endorsement – tend to have lower ratings than others. This is a relationship that will be tracked in future cohorts to be able to draw any conclusions on this with greater certainty.

56. Projects for which preparation grants were not provided by the GEF tend to have lower ratings for quality at entry of M&E arrangements. Ratings for projects that received preparation grants tend to be higher, however the amount of the preparation grant does not seem to influence ratings. Given the small number of observations on projects with no project preparation grant support, it is difficult to draw strong conclusions.

Quality of Indicators

57. To assess the extent that the impact indicators specified in the proposals meet quality expectations, indicators for any given proposal were assessed using both a high and low – but reasonable – performance threshold. For example, when assessing the extent that specified

¹⁴ "To correct for any sampling probability-related biases due to the stratified random sampling approach, probability-adjusted percentages have been reported in the narrative and in the table of this section".

indicators cover major intended impacts – indicators covering all the major intended impacts of the project were considered to meet the high performance threshold. On the other hand, a lower threshold of expected performance would be when the indicators at least cover most of the major intended impacts (although they may also have some minor gaps). While 69 percent of the proposals indicators met the low quality threshold for dimensions such as scientific validity and congruence with project’s theory of change, only 16 percent of the proposals met the high threshold on these dimensions.

Baseline Information

58. Fourteen percent of the proposals met a high threshold of providing baseline information for all the specified indicators. Sixty three percent met a lower threshold of providing baseline information for most of the specified indicators and for the remaining 23 percent of indicators it was reported that baseline information will be gathered once project implementation starts.

59. The extent to which proposals provide baseline information needs to be assessed along with the quality of indicators for a more comprehensive account of the quality of baseline information. Fifty two percent of proposals met a low threshold on baseline information along with meeting the low thresholds for indicator quality dimensions such as coverage of expected impacts, congruence with project’s theory of change, and scientific validity. When a high threshold is applied, percentage of proposals that met all of them dropped to 6 percent.

Quality of Planning for Data Collection on Impact Indicators

60. For 81 percent (38) of proposals, the planned frequency for data collection was assessed to be appropriate for either all or most of the specified indicators. For 87 percent (41) of the proposals, institutional arrangements and responsibilities for collection of data on impact indicators had been specified. The capacities of the individuals and/or institutions that are vested with the responsibilities to gather information on impact indicators were assessed to be sufficient for 74 percent (35) of proposals. For 48 percent (21) of the projects the amount budgeted for impact measurement-related work was assessed to be sufficient. For additional 37 percent (20) it was ascertained that – although sufficient information is not available to assess whether the budget for impact measurement is adequate – the budgeted amount for M&E was sufficient.

61. In all, 62 percent of proposals were assessed to have met the low thresholds for acceptable performance on data collection, institutional arrangements and responsibilities, capacities of the institutions responsible for collection of data on impact, and sufficiency of budget.

Management Action Record

62. The Management Action Records (MAR) keep track of the level of adoption of Council’s decisions on the basis of evaluation findings and recommendations. This year’s MAR tracks the level of adoption of 12 Council decisions, which were based on 9 GEF Evaluation Office documents, by presenting ratings from GEF management and verification of these ratings by the Evaluation Office. The Evaluation Office was able to verify progress on adoption of 10 decisions.

63. Of the 12 Council Decisions for which adoption was rated, two (17 percent) were rated as 'high'. The decisions concerned the *Annual Country Portfolio Evaluation Report 2010* and the *Review of the Earth Fund*.

64. Of the 12 Council Decisions for which the level of adoption was rated, 42 percent were rated as "substantial" and 17 percent as 'medium'. For one decision the level of adoption was rated to be negligible. The adoption of the Council decision based on the *Country Portfolio Evaluation Report 2009* that the GEF Secretariat should conduct a survey to assess "exceptional situations concerning limited access to GEF partner International Financial Institutions" was rated as negligible because the GEF Secretariat has not yet conducted this survey.

65. Since its start the MAR has tracked the level of adoption for 100 GEF Council decisions based on 28 evaluations. Of these, 90 have graduated. Of the graduated decisions, 38 percent have been graduated because their adoption was rated 'high'. These include the two for which adoption was rated high in FY2011. Other decisions have been graduated because they were no longer relevant, for example because later decisions of Council replaced older ones.

Recommendation

66. There is early evidence of inclusion of Operational Focal Points in monitoring and evaluation plans at project entry, as required in the new GEF M&E policy. GEF Agencies should continue in this direction and enhance their efforts to specify how OFPs will be engaged, when feasible and relevant, in project or program monitoring and evaluation.