SECOND BENCHMARKING REVIEW
OF ECG MEMBERS' EVALUATION PRACTICES FOR THEIR PRIVATE SECTOR INVESTMENT OPERATIONS AGAINST THEIR AGREED GOOD PRACTICE STANDARDS

January 25, 2005

Walter I. Cohn & Associates, LLC
6212 Goodview Street, Bethesda, Maryland, 20817, USA
Telephone: (301) 320-2686 Fax: (301) 320-2529 E-mail: widacohn@comcast.net

CONTENTS

ABBREVIATIONS & DEFINITIONS

AfDB	African Development Bank
AsDB	Asian Development Bank
Best practice standards	Practices going beyond good practice standards that are desirable but not essential
EBRD	European Bank for Reconstruction & Development
ECG	Evaluation Cooperation Group (of the MDBs)
EIB	European Investment Bank
Good practice standards	The key principles and practices that any development institution that finances the private sector should follow if it is to have a satisfactory evaluation system
GPS or GPS-IO	Good Practice Standards for Private Sector Investment Operations
Harmonization standards	Minimum standards necessary to permit comparability of reported operational results among the MDBs, as called for by the Development Committee
IADB	Inter-American Development Bank
IIC	Inter-American Investment Corporation
IFC	International Finance Corporation
IMF	International Monetary Fund
MC	Materially consistent with substance of GPS
MDB	Multilateral Development Bank
MIGA	Multilateral Investment Guarantee Agency
Other standards	Standards (other than *harmonization* standards) specified in ECG-GPS 2nd edition, May 2003, not essential for comparability of reported results but desirable for enhanced evaluation relevance, accountability and learning within each institution
NR	Not relevant, i.e., the nature of the MDB's operations (not its policies or practices) *or* the MDB's total reliance on evaluations carried out by the central evaluation department (rather than evaluations by the MDB's operational staff) makes it impossible for the MDB to follow the standard, *or* the MDB meets a best practice standard specified in the GPS
NC	Not materially consistent with GPS standard
WBG	World Bank Group
WGPSE or the Working Group	The ECG's Working Group on Private Sector Evaluation

SECOND BENCHMARKING REVIEW
OF ECG MEMBERS' EVALUATION PRACTICES FOR PRIVATE SECTOR INVESTMEMENT OPERATIONS AGAINST THEIR AGREED GOOD PRACTICE STANDARDS

EXECUTIVE SUMMARY

1. Almost nine years ago, a Development Committee Task Force called for harmonization of evaluation methodologies, performance indicators and rating criteria by the Multilateral Development Banks (MDBs). In February 2001, a working group of the MDB's Evaluation Cooperation Group (WGPSE) issued good practice standards for evaluation of private sector investment operations (GPS), and in October 2002, I reported on an assessment of the extent of harmonization achieved around those GPS. In May 2003, following a review of the GPS based on my findings, the WGPSE agreed on a revised set of standards. The present paper reports on my second assessment of the extent of harmonization as of the third quarter of 2004. It covers all but one of the WGPSE member agencies.[1]

2. The member institutions have made significant progress towards harmonization over the past two years. All but one have taken steps to this end and, overall, the members' policies and practices are now materially consistent with 59% of the good practice harmonization standards (76% weighted by the value of the commitments of the individual institutions). Although changes in the standards limit the validity of comparisons, the 59% compares favorably with the 39% average score in 2002. The improvements can be seen not only in the overall results but also in the changes that virtually all members have undertaken in adopting individual standards.

3. The members, however, still fall short of the Development Committee Task Force's call for comparability of development effectiveness assessments. Some comparability is now possible between IFC and IIC, though caveats would still be needed. Greater caveats would be needed for comparisons with EBRD, which considers itself different because of pursuing transition, rather than development. Meaningful comparisons will not be possible with MIGA, which has made great strides but, because of significant cancellations before projects reach early operating maturity, may not be able to evaluate a random, representative sample of the projects for which it provides guarantees. And no comparisons can be made with the three institutions (AfDB, AsDB and IADB) that operate primarily in the public sector, which do not yet have GPS-based evaluation systems for their private sector investments.

4. Many of the remaining shortfalls are attributable to factors outside the control of the members' evaluation departments. For example, several institutions have not established at-approval benchmarks for satisfactory expected returns and, thus, provide no basis for assessing outcomes for two performance indicators. More broadly, institutional mandates, policies or practices may limit some members' ability to follow the GPS. And several institutions—notably, the three operating primarily in the public sector—have not allocated the resources needed for evaluating their private sector operations in line with the GPS or may have faced resistance from their private sector operational departments.

5. But some shortfalls may be attributable to the evaluation departments themselves. Virtually all the members have made significant progress, but a few—perhaps understandably—may have devoted inadequate attention to harmonization with the GPS. They have been slow to develop and carry out standards and procedures fully consistent with the GPS for private sector investments. Until the lagging institutions and their evaluation departments give higher priority to harmonizing their practices with the agreed GPS, they will continue to fall short of the Development Committee's 1996 mandate.

---

[1] The members of the working group comprise the AfDB, AsDB, EBRD, EIB, IADB, IIC, IFC and MIGA. EIB does not fall within the purview of the 1996 harmonization mandate covered by this report.

SECOND BENCHMARKING REVIEW
OF ECG MEMBERS' EVALUATION PRACTICES
FOR PRIVATE SECTOR INVESTMEMENT OPERATIONS
AGAINST THEIR AGREED GOOD PRACTICE STANDARDS

1. BACKGROUND

Context

1. Almost nine years ago, a Development Committee Task Force called for harmonization of evaluation methodologies, performance indicators and criteria by the Multilateral Development Banks (MDBs):

> The development of objective indicators of performance is…essential for the public accountability of the MDBs and their ability to justify their use of public resources to shareholder governments, parliaments, and the public. Currently, it is not possible to compare their operational results, or even to describe them in a common language. Major public sector institutions like the MDBs must be able to account for their efforts in readily understood terms. A common methodology for evaluating their portfolios should be developed.... A determined effort should be made to harmonize performance indicators and evaluation criteria, taking into account the differing circumstances of each institution. The lessons learned from these evaluations should be shared among the MDBs with a view to applying them quickly in new operations.[1]

The MDB presidents endorsed these recommendations and called for "further intensification of collaboration among MDB evaluation units in harmonizing evaluation standards…."[2]

2. In February 2001, in response to this mandate, the Working Group on Private Sector Evaluation (WGPSE, or the Working Group) unanimously agreed on a set of good-practice standards for evaluation of private sector investment operations (the GPS-IO or, more briefly, the GPS).[3] As part of these standards, the members of the WGPSE agreed that they would "arrange for independent periodic crosscutting assessments of the extent to which these…good-practice standards are being applied in each member agency's evaluations and annual reporting, and report the findings to the MDB Presidents."[4]

3. My first assessment of the application of the GPS, completed in late 2002, not surprisingly found that harmonization was at an early stage.[5] The members' policies and practices were, on

[1] Development Committee, Task Force on Multilateral Development Banks, "Serving a Changing World—Report of the Task Force on Multilateral Development Banks," March 15, 1996, p. 18.
[2] Development Committee, "Report from the Multilateral Development Banks on Implementation of the Major Recommendations of the MDB Task Force Report," March 26, 1998, p. 4.
[3] MDB-ECG, WGPSE, "MDB-ECG Good-Practice Standards for Evaluation of Private Sector Investment Operations," April 23, 2001.
[4] *Op. cit.,* p. 17. The members of the WGPSE comprise AfDB, AsDB, EBRD, EIB, IFC, IADB, IIC and MIGA. The members of the ECG comprise AfDB, AsDB, EBRD, EIB, IADB Group, IMF, and WBG. The GPS are intended to apply to those investment or guarantee operations of the ECG members in developing and transition countries where there is no sovereign recourse for the MDB. Thus, the GPS apply to all the operations of IFC, IIC and MIGA, the bulk of the operations of EBRD, and smaller shares of the operations of the remaining members. To facilitate the exposition, "members" refers to the members of the WGPSE, rather than the ECG. EIB, which devotes roughly 85% of its resources to the European Union and does not fall within the purview of the 1996 harmonization mandate (despite accounting for 22% of the MDB's 2003 private sector commitments in developing countries), is not covered by this report.
[5] Walter I. Cohn & Associates, LLC, "Benchmarking of ECG Members' Evaluation Practices for Their Private Sector Investment Operations Against Their Agreed Good-Practice Standards, October 18, 2002.

average, materially consistent with only 39% of the standards. The ratings for individual members ranged from 8% to 93%. Three factors accounted for the shortfalls. First, some members believed they had insufficient leverage to mobilize the management and operational staff cooperation and budgetary allocations needed for adoption and application of the standards. Second, some had begun to address the issue of adopting the standards only recently. Third, a few, on further consideration, disagreed with some previously agreed standards or considered them to be inappropriate to their own institution's circumstances.

4. Based on this assessment, the Working Group agreed on a set of revised standards in May, 2003.[6] The revised GPS reduced the number of standards from 93 to 72, by dropping some and combining others. In addition, the revised GPS created a distinction between *harmonization standards*—the minimum standards necessary to permit comparability of reported operational results among the MDBs, as prescribed by the Development Committee—and *other standards*—standards not essential for comparability of reported results but nonetheless desirable for enhanced evaluation relevance, accountability and learning within each institution. After characterizing 22 standards as not required for comparability, 50 harmonization standards remained. The revised GPS also distinguished between *good practice standards*—the key principles and practices that any development institution that finances the private sector should follow if it is to have a satisfactory evaluation system—and *best practice standards*—practices going beyond good practice standards that are desirable but not essential. The revised edition of the GPS contains 50 good practice harmonization standards, 21 other good practice standards, 11 best practice harmonization standards, and 11 other best practice standards.[7]

5. At the request of the WGPSE, I have carried out a second assessment of the member institutions' application of the revised standards, and this paper reports on my findings with respect to the good practice harmonization standards, i.e., the minimum standards necessary to permit comparability of reported operational results among the MDBs.[8] The remainder of the present chapter summarizes the methodology used for assessing the extent of harmonization. Chapter 2 reports on the progress made. Chapter 3 discusses the issues encountered in assessing performance and applying the scoring system. And Chapter 4 provides some concluding remarks.

Methodology

6. This assessment used the following rating system in assessing the members' evaluation policies and practices as of the third quarter of 2004:[9]

---

[6] ECG, WGPSE, "MDB-ECG Good-Practice Standards for Evaluation of Private Sector Investment Operations," Second Edition, May 16, 2003. Annex 3 of the current report sets down the criteria I used in interpreting the standards.

[7] The good practice standards add up to 71, rather than 72, because one standard (GPS 23) has only a best practice option. One standard (GPS 13) has two alternatives for good practice.

[8] ECG, WGPSE, Consultant Terms of Reference for Second Assessment of WGPSE Member Practices against MDB-ECG Good-Practice Standards for Evaluation of Private Sector Investment Operations, May 2004.

[9] This approach can be criticized on at least two grounds, both valid. First, all standards receive equal weight even though some are clearly more important than others and even though a Not Materially Consistent (NC) rating on some standards requires NC ratings for others. Second, the results give an unjustified impression of precision. Alternative approaches, however, would also be subject to criticism. Seeking to agree weights to be assigned to each standard would have added a major, unnecessary complication. Provision of overall judgments not backed by systematic analysis could have been criticized as too subjective. Although the quantitative indicators in Chapters 2 should be taken as broadly indicative of the extent of harmonization, rather than as precise measures, the approach used provides a useful summary of the extent of harmonization achieved.

Table 1: Rating System

Rating	Description
Materially Consistent (MC)	Member's practices are materially consistent with substance of GPS.
Not Consistent (NC)	Member's practices are not materially consistent with substance of GPS.
Not Relevant (NR)	The nature of the MDB's operations (not its policies or practices) *or* the MDB's total reliance on evaluations carried out by the central evaluation department (rather than evaluations by the MDB's operational staff) makes it impossible for the MDB to follow the standard *or* the MDB meets an equivalent or higher standard specified in the GPS.

7. The rating system differs from the system used for the 2002 benchmarking exercise in three respects. First, it eliminates the earlier distinction between materially consistent and partly consistent. This distinction would have added little and would have been inconsistent with GPS 41, which calls for rating scales with a balance between positive and negative characterizations. Second, the new system drops the assignment of numerical scores used for the first assessment. These scores would have added little to the findings. Third, the rating system has broadened the definition of "not relevant." In addition to assigning "not relevant" ratings where the nature of an MDB's operations made it impossible for the MDB to follow the standard, I have assigned "not relevant" ratings for (a) standards that could not be met by an MDB that does not call for self-evaluations prepared by the operational departments and is totally reliant on evaluations by its central evaluation department and (b) transitional standards that were irrelevant for MDBs meeting an alternative prescribed good practice standard or a related best practice standard. As in the first assessment, I have calculated the consistency percentages excluding the standards rated as not relevant.

8. Based on a review of relevant documents and communications with each member during June-August 2004, I completed a harmonization matrix for each institution.[10] These matrices contain the detailed ratings and the basis for the ratings. Each MDB's evaluation department had opportunities to comment on the completed matrix for its institution. Annex 2 records the disagreements expressed by the MDBs following the discussion of the first draft of this paper (in October 2004) as well as my comments on the issues involved.[11]

9. Because of the 2003 revisions in the 2001 GPS and the changes in the rating system, the findings of this harmonization review are not strictly comparable with the 2002 review (para. 12).[12] Nevertheless, because readers will undoubtedly wish to make comparisons, the report provides information on the 2002 results.

[10] The ratings reflect developments through early September 2004.
[11] Some institutions asked for the statistics to distinguish between standards where the evaluation department can ensure consistency with the GPS and standards where the evaluation department cannot ensure consistency with the GPS. I have not made this distinction in the statistics because some standards would not fit neatly into one of these two categories, and in any case, doing so would not change the assessed extent of the members' alignment around the agreed GPS.
[12] In addition, my own judgments changed in a few cases. These changes, however, did not materially affect the findings.

2. PROGRESS ON GOOD PRACTICE HARMONIZATION STANDARDS

Overall Findings

10. Virtually all the surveyed member institutions have made significant progress towards harmonization. On average, their practices were materially consistent with 59% of the good practice harmonization standards, well above the 39% for the good practice standards in 2002. Weighted by the relative volume of commitments made by each institution during 1999-2003, the overall score would be 77%.[13]

11. (Four members disagree with 27 ratings, i.e., 8% of the total number of individual ratings. As explained in Annex 2, EBRD, IFC and MIGA disagree with my judgments on thirteen ratings, and AsDB, EBRD and MIGA disagree with fourteen ratings for standards they consider irrelevant. These members argue that certain standards are irrelevant because of conflicts with institutional practices (5 standards), conflicts with their mandates (3), a lack of upstream operational standards (2), conflicts with institutional policies (1), or the small number of their private sector projects (5). Had I accepted their views, the average score would have increased from 59% to 66 %.)

12. The 2004 results, of course, are not entirely comparable with the 2002 findings. First, the 2004 scores relate only to the 50 good practice harmonization standards (excluding the "other" good practice standards) in the second edition GPS, rather than the 86 undifferentiated good practice standards in the first edition. Second, the requirements of several of the second edition standards differ materially from the comparable first edition standards. Third, the revised scoring system has broadened the definition of "not relevant" and, thus, bolstered the percentage scores for some MDBs independently of any real progress towards harmonization.

13. Still, initiatives undertaken by most of the MDBs over the past two years corroborate the significant progress reflected in the statistics. Some of these initiatives have yet to bear fruit but others have already resulted in material improvements.

Findings, by Member

14. As in 2002, the ratings vary greatly by member. Individual members earned MC scores ranging from 8% to 92% of the good practice harmonization standards. Leaving aside IADB, which has not moved towards harmonization and is beginning to develop an evaluation system inconsistent with the GPS-IO, the overall average would be 68%, rather than 59%. More generally, the four institutions operating solely or primarily in the private sector earned significantly higher ratings—73% to 92%— than the three operating primarily in the public sector—8% to 48%. (Table 2 and Annex 1.)

[13] Weights derived from data provided by IFC: IFC 41%, EBRD 30%, MIGA 18%, IADB 5%, IIC 2%, AfDB 2%, and AsDB 2%. If EIB were to be included, it would have a weight of 18%, and the weights of the other members would decline proportionately.

Table 2: Percent of MC Ratings, By Member

Member	2004 Scores	2002 Scores	What 2004 Scores Would Be If Members' Views on Contested Ratings Were Accepted[a]
IFC	92%	93%	98%
IIC	82%	55%	82%
EBRD	76%	48%	98%
MIGA	73%	23%	91%
AsDB	48%	33%	52%
AfDB	38%	8%	38%
IADB	8%	9%	8%
Overall	59%	39%	66%

[a] For purposes of transparency, I have provided information on what the scores would be if the members were to rate themselves. These scores are not comparable to the scores in the two previous columns, which reflect independent judgments. As indicated in Annex 2, I disagree with the member self ratings on the standards in question.

15. As in 2002, IFC received the highest score, essentially because the GPS largely reflect its practices. Its 92% rating for the good practice harmonization standards reflects a few mandate changes enhancing the independence of its evaluation department and a few decreases attributable to changes in the standards that it has not yet adopted. (If its views on three disputed standards were to be accepted, IFC would have a 98% score.) IFC plans to make further improvements next year.

16. IIC increased its score sharply, from 55% to 82%. It now requires its operational staff to prepare self-evaluation reports on 100% of its investments reaching early operating maturity and has brought most of its evaluation procedures in line with the GPS. It is considering further changes that should increase its score in the next benchmarking exercise.

17. EBRD also increased its score sharply, from 48% to 76%. (Were EBRD's views on eleven disputed ratings to be accepted, its score would be 98%.) EBRD has taken steps to enhance the independence of its evaluation department and has changed its procedures to adopt several additional good practice harmonization standards. It envisages further improvements next year, particularly with respect to reporting. But it does not plan on adopting other standards, notably certain standards relating to performance dimensions, indicators and methods for establishing benchmarks, because it considers that some of these standards conflict with its mandate, policies, practices and institutional culture. In addition, EBRD considers that it cannot adopt certain standards because management has not established hurdle rates for *ex ante* financial and economic returns and returns on equity investments.

18. MIGA achieved the highest percentage point increase, from 23% to 73%. (Were MIGA's views on eight standards to be accepted, its score would be 91%.) In 2002, it established an independent unit that could devote its attention entirely to evaluation. It has now adopted procedures that are broadly consistent with the GPS and has begun issuing annual reviews on its evaluation findings. MIGA should be able to meet several other standards in the next year or so, e.g., by developing indicators and benchmarks for assessing the contribution of individual operations to its own profitability. But it will find it more difficult or even impossible to meet some other standards. (See, e.g., para. 26.)

19. AsDB achieved a modest increase, from 33% to 48%. (If its views on the relevance of four standards were to be accepted, its score would be 52%.) A new mandate, effective 2004, has enhanced the independence of AsDB's evaluation department. A pilot completion report has been prepared based partly on the GPS. And AsDB's evaluation department has recently engaged a consultant to draft new guidelines for a GPS-based system for the evaluation of private sector operations. The results of his work should permit substantial further improvements.

20. AfDB increased its score sharply from a low base but still met only 38% of the good practice harmonization standards. The up-graded ratings arise mainly from new guidelines issued in early 2004. AfDB expects to put the new guidelines into effect in 2005 and to begin reporting on findings after that. Although much remains to be done, AfDB's performance is creditable in light of its having had to abandon its Abidjan headquarters in late 2002.

21. IADB met the requirements of only 8% of the good practice harmonization standards. The only standards it met, all relating to the independence of its central evaluation unit (GPS 1-7), are broadly similar to the ones applicable to public sector evaluation. Although IADB recently put into effect a system for monitoring development impacts and expects to use these monitoring reports as the basis for a self-evaluation system, the system mirrors IADB's practices for public sector operations. IADB has made no progress towards harmonization with the GPS-IO.

Findings, by GPS Section

22. The ratings vary significantly by GPS section (Table 3 and Annex 1). The average scores for individual sections range from 46% to 76% of the good practice harmonization standards, up from 22% to 66% in 2002.

Table 3: Percent of MC Ratings, By Section

GPS Section	% of Good Practice Harmonization Standards	2004 Scores	2002 Scores
Roles of independent & self evaluation (GPS 1-8)	12%	78%	66%
Evaluation timing, population, coverage & sampling (GPS 9-15)	14%	52%	22%
Instructions, execution and validation (GPS 16-25)	10%	70%	39%
Evaluative scope (GPS 26-50)	50%	58%	37%
Annual reporting & process transparency (GPS 51-62)	14%	46%	32%
Identification of lessons, dissemination, and ensuring application of lessons (GPS 63-72)	0%	n.a.	35%
Overall	100%	59%	39%

23. *Roles of independent and self-evaluation—ensuring the evaluation department's independence.* As in 2002, this section received the highest percentage of MC ratings—78%, compared with 66% two years ago. (None of the members questioned the relevance of these standards or my judgments on them.) The evaluation departments for all seven members now have Board-approved mandates providing specific assurances of independence. Improvements in mandate provisions allowed higher ratings for nearly all members.

24. But some shortfalls remain. The most important arise from a lack of explicit mandate provisions (a) to ensure unrestricted access to clients, projects, and co-financiers (3 members) and (b) to ensure that evaluation departments can transmit their reports to the institutions' boards of directors without management clearance or any management-imposed restrictions on their scope or contents (4 members). The members' boards can remedy these shortfalls by revising the evaluation mandates to reflect these provisions.

25. *Evaluation timing, population coverage and sampling—ensuring that each institution evaluates a random, representative sample drawn from the entire population of projects reaching early operating maturity.* The percentage of MC ratings improved more than for any other section—from 22% to 52%. (Three members disagreed with my judgments on a total of six ratings, and one considered one standard to be not relevant. Were their views to be accepted, the average score would increase to 94%.) All members made some improvements allowing higher scores, except for IFC, which already scored 100% in the 2002 assessment, and IADB, which is not following the GPS-IO for its private sector activities. Absent IADB, the average would be 61%.

26. But some shortfalls remain. Four members still do not limit the population subject to evaluation to investments deemed to have reached early operating maturity *during* the year, and three have not fully adopted the GPS definition of early operating maturity for non-financial market operations. Accepting these standards should not pose major issues for them. Only two members have fully adopted the GPS definitions of early operating maturity for financial market operations. These definitions need to be reviewed for easier application in practice. And only two members select a random, representative sample of sufficient size to meet the GPS requirements. Evaluating enough investments to permit valid generalizations on performance has budgetary implications, and the institutions not currently covering a sufficient number would need to allocate additional resources. Moreover, one member (MIGA) cannot meet the standards on sampling, since roughly a third of its clients cancel their guarantees before projects reach early operating maturity. MIGA, thus, would not have access to the project sites and the information needed to evaluate these operations (unless it could modify its guarantee documentation to permit access for evaluation purposes). Its evaluation results will consequently be subject to bias and, hence, cannot be compared with the evaluation results of other members.

27. *Instructions, execution and validation—ensuring the quality of the self-evaluation reports that provide the base for the evaluation system.* The percentage of MC ratings for this section (70%) was well above the 2002 results for the roughly comparable standards (39%). (None of the members questioned the relevance of these standards or my judgments on them.) All members made some improvements that permitted higher scores, except for IIC and IFC, which already scored 100% in 2002, and IADB. Absent IADB, the average would rise to 77%.

28. But shortfalls remain, largely because three members have not yet (i) issued detailed instructions for the preparation of evaluation reports and (ii) prepared self-evaluation reports that can be reviewed and validated by the central evaluation department. No issues of principle are involved here, but budgetary constraints have undoubtedly contributed to the delays in requiring self-evaluation reports.

29. *Evaluative scope—ensuring that the members evaluate the same performance dimensions, use the same indicators to assess performance and adopt appropriate benchmarks.* The members' practices are now materially consistent with 58% of these standards, which represent half the good practice harmonization standards, up from 37% for the roughly comparable good practice standards in 2002. (Two members disagreed with the relevance of the standards or with my judgments on a total of nine ratings. Were their views to be accepted, the average score would increase to 62%.) All the members (other than IADB) made improvements that permitted higher scores for this section. Absent IADB, the average would be 68%.

30. Still, the members' practices continue to fall short of the good practice harmonization standards and, because of the critical importance of these standards for meeting the Development Committee's mandate, the shortfalls undermine the prospects for comparability. Some shortfalls arise because standards call for comparing performance with at-approval thresholds used in approving or rejecting projects. Without thresholds, which are normally set by an institution's management, the evaluation department cannot meet the GPS requirements. This problem can be overcome by (a) management's establishing the necessary *ex ante* hurdle rates, (b) the evaluation department's identifying *de facto* hurdle rates based on a review of a random sample of investment approvals or else on a discussion with management on hurdle rates actually applied, or (c) the Working Group members' agreeing on common benchmarks for use in evaluations by all the MDBs (para. 41).

31. Other shortfalls can be attributed, depending on one's viewpoint, to (a) insuperable conflicts with members' mandates, policies, or practices, (b) disagreements about what certain standards should require, or (c) resistance to changes in performance dimensions and indicators adopted before agreement on the GPS. Two approaches can be pursued to overcome these problems. First, the Working Group may wish to discuss again the key issues involved in a few standards. For example, they may wish to discuss whether MDB additionality should be treated as one of the indicators considered in rating the MDB's work quality (as now prescribed by GPS 39), treated as one of the indicators in rating the results on the ground (as currently done by one of the members), or elevated to transparent, stand-alone status as a new, fourth dimension, alongside the results on the ground, the MDB's investment profitability, and the MDB's work quality. Similarly, they may wish to discuss whether contributions to improved living standards should be retained as a separate indicator in assessing the results on the ground (GPS 34) or else blended into the criteria for the indicator dealing with economic costs and benefits (GPS 33). Second, members affected by these problems may wish to report on performance (at least on a transition basis) using two separate frameworks—one consistent with their historical practices and their interpretation of their MDB's mandate and the other consistent with the harmonization GPS.[14]

32. *Annual reporting and process transparency—ensuring adequate reporting on results and on the evaluation process.* The review assigned MC ratings for only 46% of the good practice harmonization standards in this section, a modest increase from the 32% for the roughly comparable standards in 2002. (Four members disagreed with the relevance of the standards or my judgments on a total of twelve ratings. Were their views to be accepted, the average score would increase to 61%.) Only three members improved their ratings for comparable harmonization standards. Absent IADB, the average would be 63%.

33. Important shortfalls remain. Three members have not met most of the standards relating to reporting on evaluation findings because of not having put into effect GPS-based evaluation systems for their private sector operations. Five members have not met the standard calling for annual reports on the quality and efficacy of their evaluation systems. And no member has fully met the requirements of the five-part standard (GPS 52) calling for provision of sufficient information to make readers aware of possible biases in the sample of projects covered by the annual review (though one is likely to meet the requirements next year). No issues of principle are involved here.

34. *Identification of lessons, disseminating, and ensuring application of lessons.* This section includes no harmonization standards.

[14] Doing so would be comparable to a company's issuing one set of financial statements based on local accounting standards and another based on international accounting standards.

Conclusions on good practice harmonization standards

35. All but one of the members has made significant progress but, overall, the members still fall short of the Development Committee's mandate. Some comparability is now possible between IFC and IIC, though caveats would still be needed. Additional caveats would be needed for comparisons with EBRD. Meaningful comparisons are not possible with MIGA, which has made great strides but, because of its high percentage of cancellations before early operating maturity, cannot evaluate a random, representative sample of the operationally mature projects for which it provides guarantees. And no comparisons can be made with the three institutions operating primarily in the public sector, which do not yet have GPS-based evaluations on their private sector operations.

36. Many, though not all, of the shortfalls may be attributable to factors outside the control of the members' evaluation departments.

- Several institutions have not established at-approval benchmarks for satisfactory expected equity returns or economic rates of return. Absent these operational benchmarks, *ex post* evaluations cannot assign ratings based on comparisons between re-estimated returns and *ex ante* hurdle rate returns. Para. 30 suggests possible solutions, calling for action by management, the evaluation departments or the Working Group.
- Several members attribute shortfalls to institutional mandates, policies, or practices and argue that evaluation departments have no power on their own to bring about the necessary changes. Para. 31 suggests possible solutions, calling for action by the evaluation departments or the Working Group.
- The nature of MIGA's business prevents it from being able to overcome the sampling problem cited in paras. 26 and 35. This constraint appears to be insuperable, though it might possibly be overcome for new approvals by modifying related guarantee documentation.
- Several institutions—particularly the institutions operating primarily in the public sector—have not allocated the resources needed for evaluating their private sector operations. These institutions may possibly consider that their evaluation efforts should be focused on the public sector where, in contrast to the private sector, operations are not subject to a market test that may provide a proxy for development effectiveness.[15] They may also limit the resources devoted to their private sector operations because these operations account for only a small share of their operations.[16] Or the private sector operational departments in these institutions may possibly be using the resource constraint argument to avoid the responsibility of reporting on development outcomes. Until these institutions allocate the resources needed, progress towards harmonization will be limited.

37. But some shortfalls may be attributable to the evaluation departments themselves. Virtually all the members have made significant progress, but a few—perhaps understandably—may have devoted inadequate attention to harmonization with the GPS. They have been slow to develop and apply standards and procedures fully consistent with the GPS for private sector investments. Until the lagging institutions and their evaluation departments give higher priority to the GPS, they will continue to fall short of the 1996 Development Committee mandate.

[15] The *Annual Reviews* issued by IFC's Operations Evaluation Group have consistently found that investments that yield satisfactory results to IFC typically achieve good development outcomes. *Annual Review of IFC's Evaluation Findings: FY2003,* March 2004, pp. 1-2. These findings, of course, may not apply to the operations of other institutions.

[16] In 2003, roughly 5% for AsDB and IADB and 11% for AfDB.

3. ISSUES ENCOUNTERED IN ASSESSING PERFORMANCE

38. The 2003 revisions dealt with the problems encountered in assessing member performance in the 2002 review. Still, a few standards remain unclear, and a few further ideas have emerged for possible revisions. Annex 3 comments on the standards that may need revision. Some just need clarification. At least five, however, involve substantive issues:

- GPS 12 defines "early operating maturity" for financial markets projects with identifiable sub-projects but (a) does not cover financial markets projects *without* identifiable sub-projects and (b) assumes—probably incorrectly—that evaluation departments can devote the time and effort needed to determine (at the time of defining the population for sampling) if investment fund sub-projects have all generated revenues for 12 months, one of the tests for early operating maturity.
- GPS 31 calls for assigning ratings based on the project portfolio's profit contribution to the financial intermediary or investment fund. This standard is unrealistic. Financial institutions in developing or transition economies are unlikely to be able to provide information on the contribution to their own profitability of their sub-loans or sub-investments financed by an MDB loan or investment. Moreover, the standard, as written, would not apply to financial markets projects that do not finance a specific set of projects.
- GPS 26 covers both (i) additionality, role and contribution and (ii) work quality, i.e., at-entry screening, appraisal and structuring; monitoring and supervision quality. These indicators address quite different issues. An MDB may do an excellent job in screening, appraising, structuring, supervising and administering its investments but still may be doing things that the private sector could have done on its own or may have failed to deliver any special value-added. The Working Group may wish to discuss whether additionality and work quality should be treated as part of the same dimension (as is now the case), treated as one of the indicators in rating the results on the ground (as currently done by EBRD), or elevated to stand-alone status as a new, fourth dimension, alongside the results on the ground, the MDB's investment profitability, and the MDB's work quality.
- GPS 36, 43 and 46 call for assessing financial returns on investment projects and the financial returns to the MDB against corporate-wide quantitative benchmarks. EBRD would like the Working Group to reconsider these standards.
- GPS 62 covers reporting on "the quality and efficacy of the MDB's evaluation system, including the self-evaluation system, any gaps in coverage of the MDB's operations, the work of [the central evaluation department], the generation and application of lessons learned in new operations, and any differences between the MDB's practices and the GPS." EBRD has pointed out that the last element—reporting on differences between the member's practices and the harmonization GPS—differs from the other elements and has suggested that consideration be given to moving this element to a new good practice harmonization standard.

4. CONCLUSIONS

39. All but one of the members have made significant progress towards harmonization over the past two years, but the members still have a way to go to meet the concerns of the Development Committee Task Force and the presidents. It is still not possible to compare the MDB's development effectiveness except, to a limited extent, among IFC, IIC and EBRD.

40. Further progress will depend on the institutions and their evaluation departments giving greater priority to the changes needed to embrace the harmonization standards. The evaluation departments, especially in the lagging institutions, can make some progress by devoting greater efforts to harmonization, but progress also depends very heavily on the support of top management and management's willingness to make changes in institutional practices. Several impediments to adoption of the harmonization standards arise from upstream practices under the

control of management or even from interpretations of institutional mandates (para. 31). These impediments cannot be overcome by the evaluation units on their own.

41. Achieving comparability of evaluation findings will also call for dealing with the issues raised in para. 38 and for agreeing on common benchmarks. When the members report on their investments' development or transition outcomes, the profitability of the investment to the MDB itself, and how well the MDB has carried out its work, they should—if they are to meet the Development Committee's and the presidents' expectations—be reporting on substantially the same things, based on the same indicators. When they assign ratings, they should use equivalent categories, language, and benchmarks. Thus, if one member rates a project as, e.g., satisfactory for its development or transition outcome, the basis for the rating should be substantially the same as it would be for any other member, and the rating should be the same as any other member would have assigned. And it goes without saying that, if two MDBs co-finance a project, they should arrive at the same development or transition outcome ratings. I have prepared a separate working paper comparing the performance dimensions, indicators, and benchmarks used by each member to serve as a point of departure in seeking to achieve this objective.

January 25, 2005

ANNEX 1: FREQUENCY DISTRIBUTION OF MATERIALLY-CONSISTENT RATINGS FOR GOOD PRACTICE HARMONIZATION STANDARDS, BY MEMBER
(Contested Ratings Highlighted in Yellow. See Annex 2.)

Std #	Harmoni-zation or Other Standards	Standard	AfDB	AsDB	EBRD	IDB	IIC	IFC	MIGA	Totals	Proportion MC
		Roles of Independent and Self-Evaluation • **Good Practice Harmonization standards (out of 6; out of 5 for MIGA)**	**4**	**5**	**5**	**4**	**4**	**5**	**4**	**32**	**78%**
		Structure and role of independent evaluation, i.e., of Central Evaluation Department (CED): • ***Good Practice Harmonization standards (out of 5)***	3	5	4	4	4	4	4	29	83%
1	H	CED has a Board-approved mandate statement, designed to ensure independence and relevance.	MC	MC	MC	MC	MC	MC	MC	7	7/7
2	H	*Good Practice*. The mandate provides that the Board of Directors oversees CED's work and that the CED's reporting line, staff, budget and functions are organizationally independent from the MDB's operational, policy and strategy departments and related decision-making.	MC	MC	MC	MC -	MC	MC	MC	7	7/7
4	H	Under its mandate, CED's scope of responsibility extends, without restriction, to all determinants of the MDB's operational results.	MC	MC	MC	MC	MC	MC	MC	7	7/7
5	H	The mandate states that CED has unrestricted access to MDB's staff, records, co-financiers, clients and projects. The mandate may, however, allow for restrictions on access to clients and projects in jeopardy cases, where an evaluator's visit could prejudice the MDB's financial interests or materially increase the risk of litigation. Should client access be restricted in jeopardy cases, the number of such cases should be reported in the MDB's annual report or annual review.	NC	MC	MC	NC+	NC+	MC	MC	4	4/7
6	H	The mandate provides that CED transmits its reports to MDB's Board after review and comment by management but without management clearance or any management-imposed restrictions on their scope and contents.	NC+	MC	MC	MC	MC	NC+	NC+	4	4/7

ANNEX 1: FREQUENCY DISTRIBUTION OF MATERIALLY-CONSISTENT RATINGS FOR GOOD PRACTICE HARMONIZATION STANDARDS, BY MEMBER
(Contested Ratings Highlighted in Yellow. See Annex 2.)

Std #	*Harmoni-zation or Other Standards*	*Standard*	*AfDB*	*AsDB*	*EBRD*	*IDB*	*IIC*	*IFC*	*MIGA*	*Totals*	*Proportion MC*
		Responsibilities of operations departments in self-evaluation and related reporting: • ***Good practice Harmonization standards (out of 1; out of 0 for MIGA)***	1	0	1	0	0	1	0	3	50%
8	H	*Good practice.* Execute XASRs on investments[1] selected pursuant to GPS 14-15 in accordance with CED's sample selection and evaluation guidelines.	MC-	NC	MC	NC	NC+	MC	NR	3	3/6
		Evaluation Timing, Population, Coverage and Sampling • **Good Practice Harmonization standards (out of 7; out of 6 for MIGA)**	**2**	**4**	**6**	**0**	**5**	**6**	**2**	**25**	**52%**
		Identification of population from which sample for evaluation is to be drawn; timing of consideration for evaluation: • ***Good Practice Harmonization standards (out of 4; out of 3 for MIGA)***	2	2	4	0	3	3	1	15	56%
9	H	Taking into consideration information on project maturity status provided by other departments, CED determines the *population* from which the investments to be evaluated each year are to be drawn.	MC-	MC	MC	NC	MC	MC	MC	6	6/7
10	H	The population from which the investments to be evaluated each year are to be drawn consists of the investments that will have reached *early operating maturity* (as defined in GPS 11 and 12) during the year. • Subject to certain exclusions, specified below, the population includes all disbursed (including partially cancelled) investments [2]--whether still active or	NC+	NC+	MC	NC	MC	MC	NC	3	3/7

[1] For guarantee operations, references to "investments," here *et passim*, should be replaced by "guarantees."

ANNEX 1: FREQUENCY DISTRIBUTION OF MATERIALLY-CONSISTENT RATINGS FOR GOOD PRACTICE HARMONIZATION STANDARDS, BY MEMBER (Contested Ratings Highlighted in Yellow. See Annex 2.)

Std #	*Harmoni-zation or Other Standards*	*Standard*	*AfDB*	*AsDB*	*EBRD*	*IDB*	*IIC*	*IFC*	*MIGA*	*Totals*	*Proportion MC*
		already closed (paid-off, sold or written off)--that have reached early operating maturity. The population also includes investments already closed, even if they never reached early operating maturity. • Excluded from the population are dropped and cancelled investments, very small investments made under special promotional programs (e.g., direct investments in SMEs that are evaluated on a program basis through a CED special study), subscribed rights offerings and investments undertaken to help finance cost overruns on projects previously financed by the MDB. • Projects that have not yet reached early operating maturity are excluded from the current evaluation year's population and rolled forward for inclusion in the population in a future year when they will have reached early operating maturity. • Investments are included in the population from which the sample for evaluation is drawn only once, i.e., only for the year in which they will have reached early operating maturity									
11	H	All operations other than the financial markets operations specifically covered by GPS 12 are deemed to have reached *early operating maturity* when (a) the project financed will have been substantially completed, (b) the project financed will have generated at least 18 months of operating revenues for the company and (c) the MDB will have received at least one set of audited annual financial statements covering at least 12 months of operating revenues generated by the project.	MC+	NC+	MC	NC	MC	MC-	NC+	4	4/7
12	H	Financial markets projects with identifiable sub-projects financed by the MDB's investment are deemed to have reached *early operating maturity* when: • For lending operations: at least 18 months shall have elapsed after the MDB's final disbursement of its loan. • For investment funds: substantially all of the projects financed will have generated at least 12 months of operating revenues.	NC+	MC-	MC	NC	NC	NC+	NR	2	2/6

[2] For guarantee operations, references to "disbursed investments," here *et passim*, should be replaced by "committed guarantees."

ANNEX 1: FREQUENCY DISTRIBUTION OF MATERIALLY-CONSISTENT RATINGS FOR GOOD PRACTICE HARMONIZATION STANDARDS, BY MEMBER
(Contested Ratings Highlighted in Yellow. See Annex 2.)

Std #	*Harmoni-zation or Other Standards*	*Standard*	*AfDB*	*AsDB*	*EBRD*	*IDB*	*IIC*	*IFC*	*MIGA*	*Totals*	*Proportion MC*
		Evaluation coverage: • ***Good Practice Harmonization standards (out of 1—only 1 counted)***	0	0	0	0	1	1	0	2	29%
13	H	*Good practice:* Preparation of XASRs (with XASR-As), PERs, or a combination of the two on a random, representative sample of sufficient size to establish, *for a combined three-year rolling sample,* success rates at the 95% confidence level, with sampling error not exceeding ±5%, for the population's development (transition) outcome, MDB investment outcome and MDB work quality.	NC	NC	NC+	NC	MC[3]	MC	NC	2	2/7
		Sampling: ***Good Practice Harmonization standards (out of 2)***	0	2	2	0	1	2	1	8	57%
14	H	The CED selects the operations for XASRs and PERs from the evaluation year's population (as defined above), subject to the following standard.	NC	MC	MC	NC	NC+	MC	MC-	4	4/7
15	H	If coverage is less than 100%, the sample should be both random and representative. Notwithstanding this principle, a CED may wish to select projects to be covered by PERs based, e.g., on the potential for learning, the high profile of an operation, credit and other risks, whether the sector is a new one for the MDB, the likelihood of replication, or the desirability of balanced country and sector coverage. If so and if the CED wishes to combine the PER with the XASR-A findings in reporting annual success rates, it uses stratified sampling methodology, as follows: • It splits the population into two strata. The first consists of the projects CED selects for PERs. The second consists of the remaining projects, i.e., the population other than the projects selected for the first stratum. • CED evaluates 100% of the first stratum. The success rates from the sample are, thus, identical to the success rates for this portion of the population.	NC	MC	MC	NC	MC	MC	NC+	4	4/7

[3] MC rating assigned because IIC met requirements of Best Practice Alternative 2, i.e., preparation of XASRs (with XASR-As), PERs, or a combination of the two on 100% of the investments in the population.

ANNEX 1: FREQUENCY DISTRIBUTION OF MATERIALLY-CONSISTENT RATINGS FOR GOOD PRACTICE HARMONIZATION STANDARDS, BY MEMBER
(Contested Ratings Highlighted in Yellow. See Annex 2.)

Std #	Harmoni-zation or Other Standards	Standard	AfDB	AsDB	EBRD	IDB	IIC	IFC	MIGA	Totals	Proportion MC
		• CED selects a random sample from the second stratum. The operational staff prepares XASRs on the projects selected, and CED prepares XASR-As on these projects (or a random sample of them). The sample is sufficiently large to give reliable estimates of the success rates for that stratum. • Based on the weight of each stratum in the overall population, CED then calculates the weighted average success rates and sampling errors, following the normal procedures for stratified sampling. If the CED wishes to select projects to be covered by PERs as above *but does not wish to combine the PER with the XASR-A findings in reporting annual success rates,* it draws the sample to be covered by XASRs from the full population for the year, without previously eliminating the projects to be covered by PERs. To the extent that specific projects may be selected for XASR-As and PERs, CED would use the PER ratings, rather than the XASR-A ratings in reporting on success rates, since CED will have carried out a more rigorous review in these cases.									
		Instructions, Execution and Validation • **Good Practice Harmonization standards (out of 5; out of 3 for MIGA)**	**1**	**4**	**5**	**0**	**5**	**5**	**3**	**23**	**70%**
		Instructions & familiarisation: • ***Good Practice Harmonization standards (out of 2)***	0	2	2	0	2	2	2	10	71%
16	H	In consultation with operations departments, CED prepares, refines and disseminates instructions for the preparation of XASRs and PERs in sufficient detail to promote consistency and objectivity in execution scope, analysis and ratings. As part of dissemination efforts, some CEDs may wish to conduct workshops to familiarize the XASR teams with requirements and supporting documentation for achieving good-practice execution.	NC+	MC	MC	NC	MC	MC	MC-	5	5/7

ANNEX 1: FREQUENCY DISTRIBUTION OF MATERIALLY-CONSISTENT RATINGS FOR GOOD PRACTICE HARMONIZATION STANDARDS, BY MEMBER
(Contested Ratings Highlighted in Yellow. See Annex 2.)

Std #	***Harmoni-zation or Other Standards***	***Standard***	***AfDB***	***AsDB***	***EBRD***	***IDB***	***IIC***	***IFC***	***MIGA***	***Totals***	***Proportion MC***
17	H	*Good practice*: The instructions include ratings guidelines with benchmarks and standard reporting templates that include the performance ratings matrix.	NC+	MC-	MC	NC	MC	MC	MC-	5	5/7
		Execution: o ***Good Practice Harmonization standards (out of 1)***	1	1	1	0	1	1	1	6	86%
19	H	The research for XASRs and PERs draws from a file review; discussions with available staff involved with the operation since its inception; independent research (e.g. on market prospects); a field visit to obtain company managers' insights and to the project site to observe and assess outcomes; and discussions with parties who are knowledgeable about the country, company and project (e.g. MDB specialists, company employees and auditors, suppliers, customers, competitors, bankers, any relevant government officials, industry associations, and local NGOs).	MC-	MC	MC	NC	MC	MC	MC-	6	6/7
		Review and independent validation: • ***Good Practice Harmonization standards (out of 2; out of 0 for MIGA)***	0	1	2	0	2	2	0	7	58%
22	H	CED conducts an independent review of each XASR to verify scope responsiveness, evident reliability of the analysis, impartiality and consistency in ratings judgments, and appropriateness and completeness of the identified lessons, and then, *for each randomly selected XASR to be used in the annual synthesis report on evaluation results*, prepares an XASR-A on the final-edition XASR that records its independent judgments on the report's quality in relation to the guidelines, assigned ratings and lessons.	NC+	NC+	MC	NC	MC-	MC	NR	3	3/6
24	H	Following preparation of each draft XASR-A, CED reviews with the XASR team and its manager the basis for its judgments where its ratings differ from those in the final edition XASR.	NC+	MC	MC	NC	MC	MC	NR	4	4/6
		Evaluative Scope **Good Practice Harmonization standards (out of 25; out of 24 for IFC & MIGA)**	**12**	**7**	**17**	**0**	**21**	**24**	**20**	**101**	**58%**

ANNEX 1: FREQUENCY DISTRIBUTION OF MATERIALLY-CONSISTENT RATINGS FOR GOOD PRACTICE HARMONIZATION STANDARDS, BY MEMBER
(Contested Ratings Highlighted in Yellow. See Annex 2.)

Std #	*Harmoni-zation or Other Standards*	*Standard*	*AfDB*	*AsDB*	*EBRD*	*IDB*	*IIC*	*IFC*	*MIGA*	*Totals*	*Proportion MC*
		Performance dimensions evaluated: • ***Good Practice Harmonization standards (out of 5)***	1	2	3	0	4	5	4	19	54%
26	H	*Good practice*: The scope of the XASR (and XASR-A) or PER includes, at a minimum, • The *project's development or transition outcome*, i.e., the project's "results on the ground" relative to the MDB's mission. • The *MDB investment's profitability* (contribution to its corporate profitability objective), and • The *MDB's additionality and work quality* (additionality is defined in GPS 39; work quality is also referred to as bank handling, operational effectiveness, or execution quality).	NC	NC+	NC+	NC	MC-	MC	NC+	2	2/7
27	H	The operation's performance under each of these dimensions is analyzed according to standard indicators, and the operation's performance for each indicator is rated according to criteria and benchmarks specified in the guidelines.	MC-	NC+	MC	NC	MC	MC	MC-	5	5/7
28	H	The performance reflected in the relevant indicator ratings is synthesized into ratings for each of the three performance dimensions, specified above.	NC	NC+	NC+	NC	MC	MC	MC-	3	3/7
29	H	Project outcomes for each of the indicators are assessed on a "with v. without project" basis..	NC+	MC	MC	NC	NC+	MC	MC	4	4/7
30	H	Assessments of development or transition outcomes for each of the development or transition outcome indicators take into consideration the sustainability of the results.	NC	MC-	MC-	NC	MC	MC	MC	5	5/7

ANNEX 1: FREQUENCY DISTRIBUTION OF MATERIALLY-CONSISTENT RATINGS FOR GOOD PRACTICE HARMONIZATION STANDARDS, BY MEMBER
(Contested Ratings Highlighted in Yellow. See Annex 2.)

Std #	Harmoni-zation or Other Standards	Standard	AfDB	AsDB	EBRD	IDB	IIC	IFC	MIGA	Totals	Proportion MC
		Indicators for the development or transition outcome: • *Good Practice Harmonization standards (out of 5)*	3	1	4	0	4	5	5	22	63%
31	H	The project's development or transition outcome is based partly on the *project's contribution to the company's business success,* measured mainly: • For capital expenditure projects: by the project's after-tax financial rate of return (FRR); • For financial markets projects: by the project portfolio's profit contribution to the financial intermediary or investment fund; • For other projects: by the project's profit contribution and the achievements of the company's at-approval business objectives.	NC	NC+	MC	NC	MC	MC	MC	4	4/7
32	H	The project's development or transition outcome is based partly on the *project's contribution to the country's private sector development and/or its development of efficient capital markets and/or its transition to a market economy.*	MC	MC	MC	NC	MC	MC	MC	6	6/7
33	H	The project's development or transition outcome is based partly on its *economic viability.* For non-financial markets operations: Based mainly on the project's net quantifiable economic benefits and costs, as measured by the project's real economic rate of return (ERR). For financial market operations: Rated on whether the sub-projects financed are economically viable (as reflected, e.g., in sub-project ERRs or the portfolio credit or equity IRR performance combined with the absence of portfolio concentrations in protected industries); whether the project has led to use of economic viability criteria in the intermediary's or investment company's investment decisions; and benefits to the economy. In most cases, quantitative information on the economic viability of sub-projects is not available to the MDB. The judgment, therefore, relies on assessing portfolio financial performance and an assessment of the extent	NC	NC+	MC-	NC	MC-	MC	MC	4	4/7

ANNEX 1: FREQUENCY DISTRIBUTION OF MATERIALLY-CONSISTENT RATINGS FOR GOOD PRACTICE HARMONIZATION STANDARDS, BY MEMBER
(Contested Ratings Highlighted in Yellow. See Annex 2.)

Std #	*Harmoni-zation or Other Standards*	*Standard*	*AfDB*	*AsDB*	*EBRD*	*IDB*	*IIC*	*IFC*	*MIGA*	*Totals*	*Proportion MC*
		to which the intermediary or investment company invests in protected industries.									
34	H	The project's development or transition outcome is based partly on its *contribution to the country's living standards*. This rating reflects the project's economic benefits and costs to those who are neither its owners nor its financiers, i.e., customers, employees, government, suppliers, competitors, local residents, etc.	MC	NC	NC	NC	NC+	MC	MC-	3	3/7
35	H	The project's development or transition outcome is based partly on its *environmental sustainability* (benchmarked against compliance with the MDB's specified standards in effect (i) at investment approval and (ii) at the time of the evaluation). "Environment" includes the physical environment and, to the extent covered by the MDB's policies, also includes social, cultural, and health and safety impacts.	MC	NC+	MC	NC	MC-	MC	MC	5	5/7
		Indicators for MDB's investment profitability: • ***Good Practice Harmonization standards (out of 1)***	0	0	0	0	0	1	0	1	14%
36	H	*Good practice.* MDB investment's profitability is based upon the investment's *gross contribution* in relation to corresponding at-approval standards for minimally satisfactory expected performance.	NC	NC+	NC+	NC	NC	MC	NC	1	1/7
		Indicators for MDB's work quality: • ***Good Practice Harmonization standards (out of 4)***	3	1	3	0	4	4	4	19	68%
37	H	The rating for the MDB's work quality is based partly on *at-entry screening, appraisal and structuring work*; i.e., how effectively the MDB carries out its work prior to approval of the investment.	MC	NC+	MC	NC	MC	MC	MC-	5/7	5/7
38	H	The rating for the MDB's work quality is based partly on its *monitoring and supervision quality*, i.e., how effectively the MDB carries out its work after approval of the investment.	MC	NC+	MC	NC	MC	MC	MC-	5/7	5/7
39	H	The rating for the MDB's *additionality and work quality* is based partly on its role, and contribution, i.e., the need for the MDB's participation relative to other available financing and the quality of the MDB's additionality from inception to	MC	NC+	NC	NC	MC	MC	MC-	4/7	4/7

ANNEX 1: FREQUENCY DISTRIBUTION OF MATERIALLY-CONSISTENT RATINGS FOR GOOD PRACTICE HARMONIZATION STANDARDS, BY MEMBER
(Contested Ratings Highlighted in Yellow. See Annex 2.)

Std #	*Harmoni-zation or Other Standards*	*Standard*	*AfDB*	*AsDB*	*EBRD*	*IDB*	*IIC*	*IFC*	*MIGA*	*Totals*	*Proportion MC*
		evaluation. The rating judgment considers compliance with basic operating principles, the MDB's contribution to client capacity building objectives (as relevant), it's the operation's consistency with furtherance of the MDB's corporate, country and sector strategies, and its clients' satisfaction with the MDB's service quality.									
40	H	Assessments of the MDB's work quality should be made independently of the ratings assigned for development or transition outcomes and MDB's investment profitability. These assessments, which are benchmarked against corporate good practice, reflect the quality of the MDB's contributions to good or bad outcomes, not the good or bad outcomes themselves.	NC+	MC-	MC-	NC	MC-	MC	MC-	5/7	5/7
		Performance ratings—principles and benchmarks: • ***Good Practice Harmonization standards (out of 10; out of 9 for IFC & MIGA)***	5 0	3 0	7 0	0 0	9 0	9 2	7 0	40 2	59% 8%
41	H	Within the rating scales (e.g., ranging from *unsatisfactory* to *excellent*), there should be balance between positive and negative characterizations (i.e., if there are four ratings, two are less than good and two are good or better). The words used to describe these ratings should accurately reflect whether the judgments are less than good or else good or better.	MC	NC+	MC	NC	MC	MC	MC	5/7	5/7
42	H	Each of the evaluated performance attributes is assigned a rating using a 4- to 6-point scale for each indicator ratings and a 4- to 6-point scale for each synthesis rating.	MC	NC+	MC	NC	MC	MC	MC	5/7	5/7

ANNEX 1: FREQUENCY DISTRIBUTION OF MATERIALLY-CONSISTENT RATINGS FOR GOOD PRACTICE HARMONIZATION STANDARDS, BY MEMBER (Contested Ratings Highlighted in Yellow. See Annex 2.)

Std #	*Harmoni-zation or Other Standards*	*Standard*	*AfDB*	*AsDB*	*EBRD*	*IDB*	*IIC*	*IFC*	*MIGA*	*Totals*	*Proportion MC*
43	H	The ratings benchmarks for the project's financial rate of return after-tax (FRR) can be determined by *either* of the following methods:									
		Good practice (transitional). Using an arbitrary scale, e.g., 20% or more after taxes in real terms for excellent, 10-19.9% for satisfactory, 5-9.9% for partly unsatisfactory, and <5% for unsatisfactory.	NC	NC	NC	NC	MC	NR	MC-	2	2/6
		Best practice. Comparing the re-estimated after-tax FRR in real terms with each project company's weighted average cost of capital at the time the project financing is committed. Although this approach is also subject to problems, it is clearly better than the alternatives. See Attachment 2.	NC	NC	NC	NC	NC	MC	NC	1	1/7
44	H	The benchmarks for the project's real economic rate of return (ERR) are set in relation to the MDB's ERR benchmarks used in approving or rejecting projects. The ERR benchmarks may be universal or may vary by country or business sector. They are not, however, the same as the ERRs projected at appraisal for specific projects.	NC	MC-	NC	NC	MC	MC	NC	3	3/7
45	H	*Good practice.* Loan performance benchmarks are set in relation to the MDB's expectations at approval.	MC	NC+	MC-	NC	MC	MC	NC	4	4/7
46	H	*Good practice.* Equity investment rates of return are benchmarked (or discounted if an NPV measure is used) against standards for minimally satisfactory expected performance at approval.	NC	NC	NC	NC	NC	MC	NR	1	1/6
47	H	Ratings of non-quantitative indicators require that relative qualitative judgments be made. The criteria should reflect the extent to which performance has been consistent with the MDB's policies, prescribed standards for corporate sustainability and recognized good-practice standards. The criteria for the judgments should be clearly specified in the instructions for the preparation of XASRs and in the CED's annual review. .	MC	NC+	MC	NC	MC	MC	MC-	5	5/7
48	H	The synthesis ratings for the three dimension (development or transition outcomes, profitability to the MDB, and the MDB's work quality) reflect summary qualitative performance judgments based on the underlying indicator ratings. They are not simple averages of the indicator ratings.	MC	NC	MC-	NC	MC	MC	MC-	5	5/7

ANNEX 1: FREQUENCY DISTRIBUTION OF MATERIALLY-CONSISTENT RATINGS FOR GOOD PRACTICE HARMONIZATION STANDARDS, BY MEMBER
(Contested Ratings Highlighted in Yellow. See Annex 2.)

Std #	*Harmoni-zation or Other Standards*	*Standard*	*AfDB*	*AsDB*	*EBRD*	*IDB*	*IIC*	*IFC*	*MIGA*	*Totals*	*Proportion MC*
		Standard XASR attachments: These attachments provide the basis for review and independent verification of the XASR's judgments and conclusions. They include:									
49	H	Details of the financial and economic rate of return derivations (with transparent assumptions and cash flow statements).	NC	MC	MC	NC	MC	MC	MC	5	5/7
50	H	For each safeguard dimension addressed in the MDB’s environmental and social guidelines, a comprehensive summary of environmental, worker health and safety, and social outcome compliance information with sufficient evidence from a field visit and/or client reporting to support the assigned outcome and related MDB work quality ratings.	NC	MC	MC	NC	MC	MC	MC-	5	5/7
		Annual Reporting and Process Transparency o **Good Practice Harmonization standards (out of 7; out of 6 for MIGA)**	**0**	**4**	**4**	**0**	**6**	**5**	**3**	**22**	**46%**
		Annual synthesis reporting: Annual Review o ***Good Practice Harmonization standards (out of 6; out of 5 for MIGA)***	0	4	4	0	5	5	4	22	**54%**
51	H	CED prepares an annual review addressed to the MDB’s management, staff and Board of Directors. The scope of the annual review includes, *inter alia*, a synthesis of the CED's validated findings from all XASRs and PERs generated and reviewed during the period covered.	NC+	MC	MC	NC	MC	MC	MC	5	5/7

ANNEX 1: FREQUENCY DISTRIBUTION OF MATERIALLY-CONSISTENT RATINGS FOR GOOD PRACTICE HARMONIZATION STANDARDS, BY MEMBER
(Contested Ratings Highlighted in Yellow. See Annex 2.)

Std #	*Harmoni-zation or Other Standards*	*Standard*	*AfDB*	*AsDB*	*EBRD*	*IDB*	*IIC*	*IFC*	*MIGA*	*Totals*	*Proportion MC*
52	H	The annual review should provide sufficient information to make the reader aware of possible biases in the sample of projects covered by the annual review. Consequently, the annual review: • Describes how the population was identified, how the sample was selected and, if stratification was applied or part of the sampling was non-random, states the rationale. • Reports on the number of XASRs and PERs for the year. • Includes an annex profiling the important characteristics of the evaluated sample against the population (cf. GPS 15). • Reports on the mean number of months between the dates projects reached early operating maturity, as defined in GPS 11 and 12) and the dates the corresponding XASRs or PERs were issued. • If less than 100% of the population has been covered, provides information on statistical confidence levels and states explicitly whether reported success rates can be attributed to population.	NC+	NC	NC+	NC	NC+	NC+	NC+	0	0/7
53	H	The annual review either (i) describes the ratings criteria and benchmarks in an annex or else (ii) refers to a website providing this information.	NC+	MC	MC	NC	MC	MC	MC	5	5/7
54	H	The ratings reported should be those of CED.	NC+	MC	MC	NC	MC	MC	MC	5	5/7
55	H	*Good practice.* CEDs should disclose the differences between CED and operating staff ratings and the materiality of the differences. Where CED ratings are reflected partly in XASR-As and partly in PERs, the CED should disclose the differences between CED and operating staff ratings separately for the XASRs and the PERs. The disclosure is made in global terms, not on a project-by-project basis and is limited to differences in binary outcome and work quality success ratings.	NC	MC	NC	NC	MC	MC	NR	3	3/6
56	H	*Good practice:* For each rating dimension and indicator, the annual review shows the *proportion of the evaluated sample* in each performance-rating category.	NC	NC	MC	NC	MC	MC	MC-	4	4/7
		Process transparency: Annual Report. (The annual report can be included in the annual review if an MDB wishes to do so.) ***Good Practice Harmonization standards (out of 1)***	0	0	0	0	1	0	0	1	14%

ANNEX 1: FREQUENCY DISTRIBUTION OF MATERIALLY-CONSISTENT RATINGS FOR GOOD PRACTICE HARMONIZATION STANDARDS, BY MEMBER
(Contested Ratings Highlighted in Yellow. See Annex 2.)

Std #	*Harmoni-zation or Other Standards*	*Standard*	*AfDB*	*AsDB*	*EBRD*	*IDB*	*IIC*	*IFC*	*MIGA*	*Totals*	*Proportion MC*
62	H	CED reports annually to the MDB's management and Board on the quality and efficacy of the MDB's evaluation system, including the self-evaluation system, any gaps in coverage of the MDB's operations, the work of CED, the generation and application of lessons learned in new operations, and any differences between the MDB's practices and the GPS.	NC	NC	NC+	NC	MC-	NC+	NC+	1	1/7

Grand Totals (Number of MC Ratings)	***19***	***24***	***38***	***4***	***41***	***45***	***32***	***203***	***59%***
Percentage of MC Ratings	***38%***	***48%***	***76%***	***8%***	***82%***	***92%***	***73%***		***59%***

ANNEX 2: DISAGREEMENTS WITH RATINGS ASSIGNED FOR GOOD PRACTICE HARMONIZATION STANDARDS

Std #	***Standard***	***Disagreement***	***Response***
	Roles of Independent and Self-Evaluation		
	Structure and role of independent evaluation, i.e., of Central Evaluation Department (CED):		
1	CED has a Board-approved mandate statement, designed to ensure independence and relevance.		
2	*Good Practice*. The mandate provides that the Board of Directors oversees CED's work and that the CED's reporting line, staff, budget and functions are organizationally independent from the MDB's operational, policy and strategy departments and related decision-making.		
4	Under its mandate, CED's scope of responsibility extends, without restriction, to all determinants of the MDB's operational results.		
5	The mandate states that CED has unrestricted access to MDB's staff, records, co-financiers, clients and projects. The mandate may, however, allow for restrictions on access to clients and projects in jeopardy cases, where an evaluator's visit could prejudice the MDB's financial interests or materially increase the risk of litigation. Should client access be restricted in jeopardy cases, the number of such cases should be reported in the MDB's annual report or annual review.		
6	The mandate provides that CED transmits its reports to MDB's Board after review and comment by management but without management clearance or any management-imposed restrictions on their scope and contents.		
	Responsibilities of operations departments in self-evaluation:		
8	*Good practice.* Execute XASRs on investments[1] selected pursuant to GPS 14-15 in accordance with CED's sample selection and evaluation guidelines.		
	Evaluation Timing, Population, Coverage and Sampling		
	Identification of population from which sample for evaluation is to be drawn; timing of consideration for evaluation:		
9	Taking into consideration information on project maturity status provided by other departments, CED determines the *population* from which the investments to be evaluated each year are to be drawn.		
10	The population from which the investments to be evaluated each year are to be drawn consists of the investments that will have reached *early operating maturity* (as defined in GPS 11 and 12) during the year. • Subject to certain exclusions, specified below, the population includes all disbursed	**MIGA disagrees with judgment:** MIGA considers that it meets the conditions for this GPS to the extent feasible. It cannot visit or get information on	Because of excluding closed projects, MIGA inevitably evaluates a biased sample of the projects it assists. This issue is material, since ca. 1/3 of MIGA's clients cancel

[1] For guarantee operations, references to "investments," here *et passim*, should be replaced by "guarantees."

ANNEX 2: DISAGREEMENTS WITH RATINGS ASSIGNED FOR GOOD PRACTICE HARMONIZATION STANDARDS

Std #	Standard	Disagreement	Response
	(including partially cancelled) investments [2]--whether still active or already closed (paid-off, sold or written off)--that have reached early operating maturity. The population also includes investments already closed, even if they never reached early operating maturity. • Excluded from the population are dropped and cancelled investments, very small investments made under special promotional programs (e.g., direct investments in SMEs that are evaluated on a program basis through a CED special study), subscribed rights offerings and investments undertaken to help finance cost overruns on projects previously financed by the MDB. • Projects that have not yet reached early operating maturity are excluded from the current evaluation year's population and rolled forward for inclusion in the population in a future year when they will have reached early operating maturity. • Investments are included in the population from which the sample for evaluation is drawn only once, i.e., only for the year in which they will have reached early operating maturity	closed projects and, hence, excludes them from the population.	their guarantees before the projects reach early operating maturity.
11	All operations other than the financial markets operations specifically covered by GPS 12 are deemed to have reached *early operating maturity* when (a) the project financed will have been substantially completed, (b) the project financed will have generated at least 18 months of operating revenues for the company and (c) the MDB will have received at least one set of audited annual financial statements covering at least 12 months of operating revenues generated by the project.	**MIGA disagrees with judgment:** MIGA defines early operating maturity for operations other than financial markets operations based on 2 criteria: the project has been on MIGA's books for at least 3 years and has been in operation for at least 18 months.	MIGA's definition (a) does not call for the project to have been substantially completed and (b) does not call for MIGA to have received at least one set of audited annual financial statements covering at least 12 months of operating revenues generated by the project. The first of these provisions is necessary to allow for projects that may start operating well before completion, e.g., plantation projects with extended implementation periods. The second of these provisions is necessary to ensure that the evaluators have an adequate basis for assessing performance.
12	Financial markets projects with identifiable sub-projects financed by the MDB's investment are deemed to have reached *early operating maturity* when:	**IFC disagrees with judgment:** IFC's test is that at least 18 months shall have elapsed since	IFC argument assumes that the time from investment funds' final disbursement of sub-investments to

[2] For guarantee operations, references to "disbursed investments," here *et passim,* should be replaced by "committed guarantees."

ANNEX 2: DISAGREEMENTS WITH RATINGS ASSIGNED FOR GOOD PRACTICE HARMONIZATION STANDARDS

Std #	*Standard*	*Disagreement*	*Response*
	• For lending operations: at least 18 months shall have elapsed after the MDB's final disbursement of its loan. • For investment funds: substantially all of the projects financed will have generated at least 12 months of operating revenues.	the financial intermediary made the final disbursements of its sub-investments. IFC argues that this test results in IFC's fund investments entering IFC's deemed early maturity population (with only few exceptions) when substantially all of their sub-projects have generated in fact more than 12 months of operating revenues.	time when projects start generating revenue is no greater than 6 months. Unlikely to be the case for equity investments, which typically are disbursed before loans.
	Evaluation coverage:		
13	*Good practice:* Preparation of XASRs (with XASR-As), PERs, or a combination of the two on a random, representative sample of sufficient size to establish, *for a combined three-year rolling sample,* success rates at the 95% confidence level, with sampling error not exceeding ±5%, for the population's development (transition) outcome, MDB investment outcome and MDB work quality.	**EBRD disagrees with judgment:** It notes that it has adopted the stratified sampling methodology outlined in GPS 15. In addition, EBRD believes that the GPS should be changed to make preparation of a combination of XASRs (with XASR-As) and PERs a best practice. **MIGA disagrees with judgment:** It considers that it cannot meet the requirements of the good practice standard or the transitional good practice standard because it cannot visit or get information on closed projects and, hence, must exclude them from the population.	EBRD has not yet determined that its sample size is large enough to meet the test of establishing success rates at the 95% confidence level, with sampling error not exceeding ±5%. It plans to report on whether it has met this test in its report on EBRD's 2004 evaluations. Since ca. 1/3 of MIGA's clients cancel their guarantees before the projects reach early operating maturity, MIGA is limited to evaluating a small, biased sample of the projects it assists. Its evaluation coverage is consequently inadequate and does not meet the requirements of this standard.
	Transitional good practice [i.e.., alternative standard]: Preparation of XASRs (with XASR-As), PERs, or a combination of the two on a random, representative sample equivalent to 60% or more of the investments in the population. In using this standard, an MDB reports on the confidence level and sampling error applicable to the success rates for the population's development or transition outcome, MDB investment outcome and MDB work quality. An MDB can use this standard only until its combined three-year rolling population of projects	**EBRD disagrees with judgment:** It points out that its procedures call for it to evaluate 60% of the projects reaching early operating maturity and that, in	The transitional good practice standard is not available to EBRD, since its combined 3-year rolling population of projects reaching early operating maturity exceeds 50.

Std #	Standard	Disagreement	Response
	reaching early operating maturity reaches 50.	fact, it has covered 76% on a cumulative basis. **AsDB considers that this standard is not applicable because it has too few private sector projects to make application of this standard practical.**	The transitional good practice standard was designed specifically for institutions with few projects.
	Sampling:		
14	The CED selects the operations for XASRs and PERs from the evaluation year's population (as defined above), subject to the following standard.		
15	If coverage is less than 100%, the sample should be both random and representative. Notwithstanding this principle, a CED may wish to select projects to be covered by PERs based, e.g., on the potential for learning, the high profile of an operation, credit and other risks, whether the sector is a new one for the MDB, the likelihood of replication, or the desirability of balanced country and sector coverage. If so and if the CED wishes to combine the PER with the XASR-A findings in reporting annual success rates, it uses stratified sampling methodology, as follows: • It splits the population into two strata. The first consists of the projects CED selects for PERs. The second consists of the remaining projects, i.e., the population other than the projects selected for the first stratum. • CED evaluates 100% of the first stratum. The success rates from the sample are, thus, identical to the success rates for this portion of the population. • CED selects a random sample from the second stratum. The operational staff prepares XASRs on the projects selected, and CED prepares XASR-As on these projects (or a random sample of them). The sample is sufficiently large to give reliable estimates of the success rates for that stratum. • Based on the weight of each stratum in the overall population, CED then calculates the weighted average success rates and sampling errors, following the normal procedures for stratified sampling. If the CED wishes to select projects to be covered by PERs as above *but does not wish to combine the PER with the XASR-A findings in reporting annual success rates,* it draws the sample to be covered by XASRs from the full population for the year, without previously	**MIGA disagrees with judgment:** MIGA contends that its sample is both random and representative.	Because of excluding closed projects and projects that have resulted or are expected to result in claims, MIGA's sample is not representative of the projects it assists.

ANNEX 2: DISAGREEMENTS WITH RATINGS ASSIGNED FOR GOOD PRACTICE HARMONIZATION STANDARDS

Std #	Standard	Disagreement	Response
	eliminating the projects to be covered by PERs. To the extent that specific projects may be selected for XASR-As and PERs, CED would use the PER ratings, rather than the XASR-A ratings in reporting on success rates, since CED will have carried out a more rigorous review in these cases.		
	Instructions, Execution and Validation		
	Instructions & familiarization:		
16	In consultation with operations departments, CED prepares, refines and disseminates instructions for the preparation of XASRs and PERs in sufficient detail to promote consistency and objectivity in execution scope, analysis and ratings. As part of dissemination efforts, some CEDs may wish to conduct workshops to familiarize the XASR teams with requirements and supporting documentation for achieving good-practice execution.		
17	*Good practice*: The instructions include ratings guidelines with benchmarks and standard reporting templates that include the performance ratings matrix.		
	Execution:		
19	The research for XASRs and PERs draws from a file review; discussions with available staff involved with the operation since its inception; independent research (e.g. on market prospects); a field visit to obtain company managers' insights and to the project site to observe and assess outcomes; and discussions with parties who are knowledgeable about the country, company and project (e.g. MDB specialists, company employees and auditors, suppliers, customers, competitors, bankers, any relevant government officials, industry associations, and local NGOs).		
	Review and independent validation:		
22	CED conducts an independent review of each XASR to verify scope responsiveness, evident reliability of the analysis, impartiality and consistency in ratings judgments, and appropriateness and completeness of the identified lessons, and then, *for each randomly selected XASR to be used in the annual synthesis report on evaluation results*, prepares an XASR-A on the final-edition XASR that records its independent judgments on the report's quality in relation to the guidelines, assigned ratings and lessons.		
24	Following preparation of each draft XASR-A, CED reviews with the XASR team and its manager the basis for its judgments where its ratings differ from those in the final edition XASR.		
	Evaluative Scope		

ANNEX 2: DISAGREEMENTS WITH RATINGS ASSIGNED FOR GOOD PRACTICE HARMONIZATION STANDARDS

Std #	*Standard*	*Disagreement*	*Response*
	Performance dimensions evaluated:		
26	*Good practice*: The scope of the XASR (and XASR-A) or PER includes, at a minimum, • The *project's development or transition outcome*, i.e., the project's "results on the ground" relative to the MDB's mission. • The *MDB investment's profitability* (contribution to its corporate profitability objective), and • The *MDB's additionality and work quality* (additionality is defined in GPS 39; work quality is also referred to as bank handling, operational effectiveness, or execution quality).	**EBRD considers that this standard is not applicable because it conflicts with EBRD's mandate:** It argues that EBRD's mandate implies that its operations should be evaluated in relation to (a) its mandate-related indicators (transition impact, environmental performance & change, and additionality), (b) sound banking indicators (project & company financial performance & fulfillment of objectives other than transition objectives), and (c) Bank effectiveness-related indicators. EBRD considers that additionality, which is mandated by the Agreement Establishing the EBRD (and probably by all the MDBs' mandates), should be an indicator considered in assessing transition impact, rather than work quality. In effect, without additionality, the outcomes or impacts of a project cannot be attributed to an MDB's assistance.	Although EBRD's evaluation categories differ somewhat from the GPS and although EBRD doesn't allocate indicators to the performance dimensions in the way specified by the GPS, it has agreed to use the available indicators to rate transition outcome in a way that would be broadly—though not completely —comparable with the ratings of the other MDBs. It plans to report on its "transition outcome" ratings (as well as its "transition impact" ratings) in its annual report on its 2004 evaluations. The standard calls for assessing the project's "results on the ground." Including judgments on additionality here may blur these judgments. Nevertheless, by treating additionality as part of work quality, the GPS may not give adequate importance to this factor. Indeed, by treating additionality purely as an element of work quality, an evaluation could rate the development or transition outcome of a project as highly satisfactory even if the institution was not needed to permit the project to go forward in the way it did. The WGPSE may wish to discuss this issue. Here, "the MDB's investment profitability"—which MIGA does not rate—is part of the core of the

ANNEX 2: DISAGREEMENTS WITH RATINGS ASSIGNED FOR GOOD PRACTICE HARMONIZATION STANDARDS

Std #	*Standard*	*Disagreement*	*Response*
		MIGA disagrees with judgment: It considers the NC+ rating to be inappropriately low and inconsistent with other ratings where investment profitability was missing.	standard.
27	The operation's performance under each of these dimensions is analyzed according to standard indicators, and the operation's performance for each indicator is rated according to criteria and benchmarks specified in the guidelines.		
28	The performance reflected in the relevant indicator ratings is synthesized into ratings for each of the three performance dimensions, specified above.	**EBRD considers that this standard is not applicable because it conflicts with EBRD's policy.** EBRD's Evaluation Policy Review, 2004, calls for assignment of an Overall Performance rating for projects evaluated, combining mandate-related indicators, sound banking-related indicators and Bank-effectiveness related indicators.	EBRD's policies do not prevent it from assigning a "transition outcome" rating, broadly—though not entirely—comparable to the GPS category, reflecting results on the ground. Indeed, to make the outcomes of EBRD's rating system more comparable with the outcomes of the performance evaluation systems of other MDBs, EBRD's Evaluation Policy Review, 2004, calls for assignment of such a rating. EBRD plans to add an annex reporting on these ratings in reporting on its 2004 evaluation results.
29	Project outcomes for each of the indicators are assessed on a "with v. without project" basis.		
30	Assessments of development or transition outcomes for each of the development or transition outcome indicators take into consideration the sustainability of the results.		
	Indicators for the development or transition outcome:		
31	The project's development or transition outcome is based partly on the *project's contribution to the company's business success*, measured mainly: • For capital expenditure projects: by the project's after-tax financial rate of return (FRR); • For financial markets projects: by the project portfolio's profit contribution to the		

ANNEX 2: DISAGREEMENTS WITH RATINGS ASSIGNED FOR GOOD PRACTICE HARMONIZATION STANDARDS

Std #	***Standard***	***Disagreement***	***Response***
	financial intermediary or investment fund; • For other projects: by the project's profit contribution and the achievements of the company's at-approval business objectives.		
32	The project's development or transition outcome is based partly on the *project's contribution to the country's private sector development and/or its development of efficient capital markets and/or its transition to a market economy.*		
33	The project's development or transition outcome is based partly on its *economic viability.* For non-financial markets operations: Based mainly on the project's net quantifiable economic benefits and costs, as measured by the project's real economic rate of return (ERR). For financial market operations: Rated on whether the sub-projects financed are economically viable (as reflected, e.g., in sub-project ERRs or the portfolio credit or equity IRR performance combined with the absence of portfolio concentrations in protected industries); whether the project has led to use of economic viability criteria in the intermediary's or investment company's investment decisions; and benefits to the economy. In most cases, quantitative information on the economic viability of sub-projects is not available to the MDB. The judgment, therefore, relies on assessing portfolio financial performance and an assessment of the extent to which the intermediary or investment company invests in protected industries.		
34	The project's development or transition outcome is based partly on its *contribution to the country's living standards.* This rating reflects the project's economic benefits and costs to those who are neither its owners nor its financiers, i.e., customers, employees, government, suppliers, competitors, local residents, etc.	**EBRD considers that this standard is not applicable because it conflicts with EBRD's institutional mandate:** It argues that, as reflected in all of EBRD's policy documents, EBRD's goal is transition and that improving the standard of living and conditions of labor is a by-product, not a goal or objective.	Article 2(1)(iii) of EBRD's Establishing Agreement calls for EBRD "to foster productive investment … thereby assisting in…raising productivity, the standard of living and conditions of labour…."
35	The project's development or transition outcome is based partly on its *environmental sustainability* (benchmarked against compliance with the MDB's specified standards in effect (i) at investment approval and (ii) at the time of the evaluation). "Environment" includes the physical environment and, to the extent covered by the MDB's policies, also includes social, cultural, and health and safety impacts.		
	Indicators for MDB's investment profitability:		
36	*Good practice.* MDB investment's profitability is based upon the investment's *gross contribution* in relation to corresponding at-approval standards for minimally satisfactory	**EBRD considers that this standard is not applicable**	The standard would make no sense if the actual equity return were

Std #	Standard	Disagreement	Response
	expected performance.	**because it conflicts with EBRD's institutional practices:** EBRD's evaluation department estimates *ex post* returns on EBRD's investments using the same model as used *ex ante.* EBRD, however, has not established at-approval standards for minimally satisfactory expected performance. Instead, it looks at each proposal as a whole and, in some cases, is prepared to make an investment yielding a marginally positive or even a negative projected financial contribution. Given this institutional practice, EBRD considers that performance should be compared with the expectations for the specific investments evaluated.	compared with a "marginal positive or even a negative projected financial contribution" that EBRD decided was acceptable in a particular case. EBRD may decide to make an investment despite a negative projected financial contribution on its equity because of the very strong expected transition impact. Indeed, doing so may make excellent sense. If everything turns out to be as expected, rating the *transition impact* as satisfactory or even excellent would be fine. But saying that a loss on EBRD's equity investment was a satisfactory investment outcome for EBRD would make "satisfactory" a meaningless term. Establishing some at-approval standards may be necessary if EBRD is to meet this standard.
	Indicators for MDB's work quality:		
37	The rating for the MDB's work quality is based partly on *at-entry screening, appraisal and structuring work*; i.e., how effectively the MDB carries out its work prior to approval of the investment.		
38	The rating for the MDB's work quality is based partly on its *monitoring and supervision quality*, i.e., how effectively the MDB carries out its work after approval of the investment.		
39	The rating for the MDB's *additionality and work quality* is based partly on its role, and contribution, i.e., the need for the MDB's participation relative to other available financing and the quality of the MDB's additionality from inception to evaluation. The rating judgment considers compliance with basic operating principles, the MDB's contribution to client capacity building objectives (as relevant), it's the operation's consistency with furtherance of the MDB's corporate, country and sector strategies, and its clients' satisfaction with the MDB's service quality.	**EBRD considers that this standard is not applicable because it conflicts with EBRD's institutional mandate and practices:** EBRD considers that additionality, which is mandated by the Agreement Establishing the EBRD, should be an indicator considered in	See comments on GPS 26, above.

ANNEX 2: DISAGREEMENTS WITH RATINGS ASSIGNED FOR GOOD PRACTICE HARMONIZATION STANDARDS

Std #	*Standard*	*Disagreement*	*Response*
		assessing transition impact, rather than work quality. In effect, without additionality, no impact can be attributed to an MDB's financing of a project.	
40	Assessments of the MDB's work quality should be made independently of the ratings assigned for development or transition outcomes and MDB's investment profitability. These assessments, which are benchmarked against corporate good practice, reflect the quality of the MDB's contributions to good or bad outcomes, not the good or bad outcomes themselves.		
	Performance ratings—principles and benchmarks:		
41	Within the rating scales (e.g., ranging from *unsatisfactory* to *excellent*), there should be balance between positive and negative characterizations (i.e., if there are four ratings, two are less than good and two are good or better). The words used to describe these ratings should accurately reflect whether the judgments are less than good or else good or better.		
42	Each of the evaluated performance attributes is assigned a rating using a 4- to 6-point scale for each indicator ratings and a 4- to 6-point scale for each synthesis rating.		
43	The ratings benchmarks for the project's financial rate of return after-tax (FRR) can be determined by *either* of the following methods: ***Good practice (transitional).*** Using an arbitrary scale, e.g., 20% or more after taxes in real terms for excellent, 10-19.9% for satisfactory, 5-9.9% for partly unsatisfactory, and <5% for unsatisfactory.	**EBRD considers that this standard is not applicable because it conflicts with EBRD's practices and because EBRD lacks established upstream standards:** EBRD does not have corporate wide *ex ante* standards for determining whether the FRR is satisfactory. EBRD's evaluations compare re-estimated FRRs with the FRRs projected for the specific projects. EBRD would like to see its practice adopted as an alternative standard.	The absence of upstream standards does not prevent EBRD from adopting an arbitrary scale for purposes of *ex post* evaluation. Moreover, by rating a project's financial performance in relation to the *ex ante* FRR for the specific project, EBRD could end up with some highly anomalous findings. For example, a project with an *ex ante* return of 40% (presumably well above any MDB's hurdle rate) and an *ex post* return of, say, 29%, i.e., more than 25% below the expected return, would be rated as unsatisfactory. But a project with an *ex ante* return of 10% and an *ex post* return of, say, 11%, i.e., 10% above the projected return, would be rated as excellent.
44	The benchmarks for the project's real economic rate of return (ERR) are set in relation to the MDB's ERR benchmarks used in approving or rejecting projects. The ERR benchmarks may	**MIGA considers this GPS as not applicable because the**	MIGA's management needs to address this issue if MIGA is to be

Std #	Standard	Disagreement	Response
	be universal or may vary by country or business sector. They are not, however, the same as the ERRs projected at appraisal for specific projects.	**institution has not established upstream benchmarks.**	able to move towards harmonization.
45	*Good practice.* Loan performance benchmarks are set in relation to the MDB's expectations at approval.		
46	*Good practice.* Equity investment rates of return are benchmarked (or discounted if an NPV measure is used) against standards for minimally satisfactory expected performance at approval.	**EBRD considers that this standard is not applicable because it conflicts with EBRD's practices:** EBRD looks for a market rate of return commensurate with the business risk it takes and, thus, considers each investment on its own merits.	EBRD's management needs to address this issue if EBRD is to be able to move towards harmonization.
47	Ratings of non-quantitative indicators require that relative qualitative judgments be made. The criteria should reflect the extent to which performance has been consistent with the MDB's policies, prescribed standards for corporate sustainability and recognized good-practice standards. The criteria for the judgments should be clearly specified in the instructions for the preparation of XASRs and in the CED's annual review. .		
48	The synthesis ratings for the three dimension**s** (development or transition outcomes, profitability to the MDB, and the MDB's work quality) reflect summary qualitative performance judgments based on the underlying indicator ratings. They are not simple averages of the indicator ratings.		
	Standard XASR attachments: These attachments provide the basis for review and independent verification of the XASR's judgments and conclusions. They include:		
49	Details of the financial and economic rate of return derivations (with transparent assumptions and cash flow statements).		
50	For each safeguard dimension addressed in the MDB's environmental and social guidelines, a comprehensive summary of environmental, worker health and safety, and social outcome compliance information with sufficient evidence from a field visit and/or client reporting to support the assigned outcome and related MDB work quality ratings.		
	Annual Reporting and Process Transparency		
	Annual synthesis reporting: Annual Review		
51	CED prepares an annual review addressed to the MDB's management, staff and Board of Directors. The scope of the annual review includes, *inter alia*, a synthesis of the CED's validated findings from all XASRs and PERs generated and reviewed during the period covered.		

ANNEX 2: DISAGREEMENTS WITH RATINGS ASSIGNED FOR GOOD PRACTICE HARMONIZATION STANDARDS

Std #	Standard	Disagreement	Response
52	The annual review should provide sufficient information to make the reader aware of possible biases in the sample of projects covered by the annual review. Consequently, the annual review: • Describes how the population was identified, how the sample was selected and, if stratification was applied or part of the sampling was non-random, states the rationale. • Reports on the number of XASRs and PERs for the year. • Includes an annex profiling the important characteristics of the evaluated sample against the population (cf. GPS 15). • Reports on the mean number of months between the dates projects reached early operating maturity, as defined in GPS 11 and 12) and the dates the corresponding XASRs or PERs were issued. • If less than 100% of the population has been covered, provides information on statistical confidence levels and states explicitly whether reported success rates can be attributed to population.	**EBRD disagrees with judgment:** Its latest annual review was consistent with 3 of the 5 requirements of this standard, and it has indicated that its next annual review will meet the 2 remaining standards.	EBRD reports on the number of months between last disbursement and the time PED undertakes to prepare an OPER, i.e., not between the time projects reach early operating maturity and the dates XASRs or PERs are issued. Moreover, EBRD's annual review does not provide information on statistical confidence levels. Although EBRD has indicated that its next annual review will meet all 5 requirements, the present assessment is based on what the MDB's have already done, rather than on intentions.
		IFC disagrees with judgment: It meets all tests other than the 4th bullet paragraph and considers this test to be of lesser importance and inappropriate, since it doesn't show that all projects that have reached maturity per GPS 11 and 12 have been included in the sampled population, nor that all projects included in it have in fact reached early operating maturity.	WGPSE should reconsider including this bullet paragraph in the harmonization standard. But, consistent with the approach taken throughout this assessment, all substantive requirements of a standard must be met for a materially consistent rating is to be assigned.
		MIGA disagrees with judgment: MIGA notes that it meets most of the requirements of this standard.	MIGA does not report on the mean number of months between early operating maturity and the dates of PERs and does not state that reported success rates cannot be attributed to the population of projects assisted.
		AsDB considers that this	AsDB should be able to meet the requirements of this standard despite

ANNEX 2: DISAGREEMENTS WITH RATINGS ASSIGNED FOR GOOD PRACTICE HARMONIZATION STANDARDS

Std #	*Standard*	*Disagreement*	*Response*
		standard is not applicable because it has too few private sector projects to make application of this standard practical.	its small number of projects. A small number of projects does not obviate the need for transparency.
53	The annual review either (i) describes the ratings criteria and benchmarks in an annex or else (ii) refers to a website providing this information.		
54	The ratings reported should be those of CED.		
55	*Good practice.* CEDs should disclose the differences between CED and operating staff ratings and the materiality of the differences. Where CED ratings are reflected partly in XASR-As and partly in PERs, the CED should disclose the differences between CED and operating staff ratings separately for the XASRs and the PERs. The disclosure is made in global terms, not on a project-by-project basis and is limited to differences in binary outcome and work quality success ratings.	**EBRD considers that this standard is not applicable because it conflicts with EBRD's institutional culture:** PED prefers a less confrontational approach (though it does note differences between PED and staff ratings in the individual evaluation papers). EBRD questions whether this standard should be considered as a harmonization standard.	The WGPSE may wish to discuss this issue.
56	*Good practice:* For each rating dimension and indicator, the annual review shows the *proportion of the evaluated sample* in each performance-rating category.	**AsDB considers that this standard is not applicable because its small sample makes it impractical to follow the standard.**	For institutions with a small number of projects, I assigned MC rating even if raw numbers, rather than proportions given. AsDB has not provided raw numbers.
57	*Good practice:* The annual review analyzes the evaluation results and highlights the findings. In doing so, it notes whether findings are statistically significant.	**AsDB considers that this standard is not applicable because its small sample makes it impractical to follow the standard.**	The number of projects may influence the extent of reporting in the annual review, not the need for reporting.
58	There is an annex containing a representative sample of XASR or PER abstracts prepared on operations selected from approximately the middle of each development or transition outcome-rating group, illustrating the application of the ratings. There is non-disclosure of any company specifics in the content of this annex to protect the confidentiality of client information. This GPS applies only to MDBs with at least 30 projects evaluated on a rolling three-year basis.	**AsDB considers that this standard is not applicable because its small sample makes it impractical to follow the standard.**	For institutions with a small number of private sector projects, providing this information would compromise client confidentiality. To some extent, this problem can be overcome by allowing abstracts to be selected on a cumulative basis. Still, Annex 3 suggests that some

Std #	*Standard*	*Disagreement*	*Response*
			exemption should be provided for smaller institutions. Meanwhile, I had to rate performance based on the GPS reflected in the 2d edition.
	Process transparency: Annual Report. (The annual report can be included in the annual review if an MDB wishes to do so.)		
62	CED reports annually to the MDB's management and Board on the quality and efficacy of the MDB's evaluation system, including the self-evaluation system, any gaps in coverage of the MDB's operations, the work of CED, the generation and application of lessons learned in new operations, and any differences between the MDB's practices and the GPS.	**EBRD disagrees with judgment:** It argues that it already reports on these issues. **IFC disagrees with judgment:** IFC indicates that it did not report on differences between its practices and the GPS because of the new exercise underway using the updated standards and that it will provide a comprehensive report in the future. **MIGA disagrees with judgment:** MIGA notes that it meets most of the requirements of this standard.	EBRD reports on most of these issues but not on differences between the MDB's practices and the GPS. It will add this information in next annual report. Ratings based on actions taken, rather than intentions. MIGA does not meet all material requirements of standard. In particular, it does not report on differences between its practices and the GPS.
	Summary		
	A. Considers standard not applicable • ***Conflicts with mandate*** o ***EBRD (incl. 1 where inst. practices also cited)*** • ***Conflicts with policies***	***3*** ***1***	

ANNEX 2: DISAGREEMENTS WITH RATINGS ASSIGNED FOR GOOD PRACTICE HARMONIZATION STANDARDS

Std #	Standard	Disagreement	Response
	o *EBRD*		
	• *Conflicts with institutional practices or culture (incl. 2 where other factors cited)*	*5*	
	o *EBRD*		
	• *Lack of upstream standards*	*1*	
	o *EBRD (incl. 1 where inst. practices also cited)*	*1*	
	o *MIGA*		
	• *Small numbers*	*5*	
	o *AsDB*		
	• *Total number of <u>standards</u> considered not applicable*	*5*	
	o *AsDB*	*8*	
	o *EBRD*	*1*	
	o *MIGA*	*14*	
	o *Total*		
	B. Disagrees with judgment		
	• *EBRD*	*3*	
	• *IFC*	*3*	
	• *MIGA*	*7*	
	• *Total*	*13*	
	C. Total		
	• *Considers standard not applicable*	14	
	• *Disagrees with judgment*	13	
	• *Grand total*	27	

ANNEX 3: CRITERIA USED FOR APPLYING GOOD PRACTICE HARMONIZATION STANDARDS AND SUGGESTIONS FOR POSSIBLE CHANGES IN THESE STANDARDS

Std #	***Harmon-ization or Other Standards***	***Standard***	***Criteria Used for Applying Standards & (in bold font) Suggestions for Possible Changes in Standards***
		Roles of Independent and Self-Evaluation	
		Structure and role of independent evaluation, i.e., of Central Evaluation Department (CED):	
1	H	CED has a Board-approved mandate statement, designed to ensure independence and relevance.	Must be approved by Board of Directors to be rated as MC.
2	H	*Good Practice*. The mandate provides that the Board of Directors oversees CED's work and that the CED's reporting line, staff, budget and functions are organizationally independent from the MDB's operational, policy and strategy departments and related decision-making.	MC requires either (i) a statement that the head of evaluation reports to the Board or (ii) a statement that the evaluation department is completely independent of operational management.
4	H	Under its mandate, CED's scope of responsibility extends, without restriction, to all determinants of the MDB's operational results.	
5	H	The mandate states that CED has unrestricted access to MDB's staff, records, co-financiers, clients and projects. The mandate may, however, allow for restrictions on access to clients and projects in jeopardy cases, where an evaluator's visit could prejudice the MDB's financial interests or materially increase the risk of litigation. Should client access be restricted in jeopardy cases, the number of such cases should be reported in the MDB's annual report or annual review.	MC requires (i) explicit statement calling for access to clients and projects and (ii) no restrictions on access to staff, records, co-financiers, clients & projects, other than in jeopardy cases.
6	H	The mandate provides that CED transmits its reports to MDB's Board after review and comment by management but without management clearance or any management-imposed restrictions on their scope and contents.	Essence of standard is freedom from management clearance or any management-imposed restrictions on scope and contents of CED reports. MC rating requires explicit statement in mandate or equivalent document designed to protect CED against management-imposed restrictions. Only 3 MDBs received MC ratings (emphasis supplied): • IADB Group. "The [OVE] Director ***is solely responsible*** for the content of [OVE] reports submitted to the Board • AsDB. Board-approved document provides that draft reports will be circulated for review and comment but "OED will *determine* which comments will be incorporated in the final version." • EBRD. A Report by the Chairman of the Audit

ANNEX 3: CRITERIA USED FOR APPLYING GOOD PRACTICE HARMONIZATION STANDARDS AND SUGGESTIONS FOR POSSIBLE CHANGES IN THESE STANDARDS

Std #	*Harmon-ization or Other Standards*	*Standard*	*Criteria Used for Applying Standards & (in bold font) Suggestions for Possible Changes in Standards*
			Committee sent to EBRD's Board states "The Corporate Director, PED, noted that…he carried *sole responsibility* for the contents of the reports."
		Responsibilities of operations departments in self-evaluation:	**Deleted words not relevant to GPS 8.**
8	H	*Good practice.* Execute XASRs on investments[1] selected pursuant to GPS 14-15 in accordance with CED's sample selection and evaluation guidelines.	
		Evaluation Timing, Population, Coverage and Sampling	
		Identification of population from which sample for evaluation is to be drawn; timing of consideration for evaluation:	
9	H	Taking into consideration information on project maturity status provided by other departments, CED determines the *population* from which the investments to be evaluated each year are to be drawn.	
10	H	The population from which the investments to be evaluated each year are to be drawn consists of the investments that will have reached *early operating maturity* (as defined in GPS 11 and 12) during the year. • Subject to certain exclusions, specified below, the population includes all disbursed (including partially cancelled) investments [2]--whether still active or already closed (paid-off, sold or written off)--that have reached early operating maturity. The population also includes investments already closed, even if they never reached early operating maturity. • Excluded from the population are dropped and cancelled investments, very small investments made under special promotional programs (e.g., direct investments in SMEs that are evaluated on a program basis through a CED special study), subscribed rights offerings and investments undertaken to help finance cost overruns on projects previously financed by the MDB. • Projects that have not yet reached early operating maturity are excluded from the current evaluation year's population and rolled forward for inclusion in the population in a future year when they will have reached early operating maturity. • Investments are included in the population from which the sample for evaluation is	

[1] For guarantee operations, references to "investments," here *et passim,* should be replaced by "guarantees."
[2] For guarantee operations, references to "disbursed investments," here *et passim,* should be replaced by "committed guarantees."

ANNEX 3: CRITERIA USED FOR APPLYING GOOD PRACTICE HARMONIZATION STANDARDS AND SUGGESTIONS FOR POSSIBLE CHANGES IN THESE STANDARDS

Std #	*Harmon-ization or Other Standards*	*Standard*	*Criteria Used for Applying Standards & (in bold font) Suggestions for Possible Changes in Standards*
		drawn only once, i.e., only for the year in which they will have reached early operating maturity	
11	H	All operations other than the financial markets operations specifically covered by GPS 12 are deemed to have reached *early operating maturity* when (a) the project financed will have been substantially completed, (b) the project financed will have generated at least 18 months of operating revenues for the company and (c) the MDB will have received at least one set of audited annual financial statements covering at least 12 months of operating revenues generated by the project.	All tests must be met for MC rating.
12	H	Financial markets projects with identifiable sub-projects financed by the MDB's investment are deemed to have reached *early operating maturity* when: • For lending operations: at least 18 months shall have elapsed after the MDB's final disbursement of its loan. • For investment funds: substantially all of the projects financed will have generated at least 12 months of operating revenues.	**When determining population, CED may be unable to determine if sub-projects have generated at least 12 months of revenues. This information may arise only in evaluation itself. The standard should be revised.**
		Evaluation coverage:	
13	H	*Good practice:* Preparation of XASRs (with XASR-As), PERs, or a combination of the two on a random, representative sample of sufficient size to establish, *for a combined three-year rolling sample*, success rates at the 95% confidence level, with sampling error not exceeding ±5%, for the population's development (transition) outcome, MDB investment outcome and MDB work quality.	For purposes of scoring, good practice & transitional good practice have been combined, and MC rating on Best practice Alternative 2 (Preparation of XASRs (with XASR-As), PERs, or a combination of the two on 100% of the investments in the population) have been considered as equivalent to MC for good practice standard. **Standard should be revised to clarify that ±5% means ±5 percentage points, rather than ±5% of the estimated success rate.**
		Sampling:	
14	H	The CED selects the operations for XASRs and PERs from the evaluation year's population (as defined above), subject to the following standard.	

ANNEX 3: CRITERIA USED FOR APPLYING GOOD PRACTICE HARMONIZATION STANDARDS AND SUGGESTIONS FOR POSSIBLE CHANGES IN THESE STANDARDS

Std #	*Harmon-ization or Other Standards*	*Standard*	*Criteria Used for Applying Standards & (in bold font) Suggestions for Possible Changes in Standards*
15	H	If coverage is less than 100%, the sample should be both random and representative. Notwithstanding this principle, a CED may wish to select projects to be covered by PERs based, e.g., on the potential for learning, the high profile of an operation, credit and other risks, whether the sector is a new one for the MDB, the likelihood of replication, or the desirability of balanced country and sector coverage. If so and if the CED wishes to combine the PER with the XASR-A findings in reporting annual success rates, it uses stratified sampling methodology, as follows: • It splits the population into two strata. The first consists of the projects CED selects for PERs. The second consists of the remaining projects, i.e., the population other than the projects selected for the first stratum. • CED evaluates 100% of the first stratum. The success rates from the sample are, thus, identical to the success rates for this portion of the population. • CED selects a random sample from the second stratum. The operational staff prepares XASRs on the projects selected, and CED prepares XASR-As on these projects (or a random sample of them). The sample is sufficiently large to give reliable estimates of the success rates for that stratum. • Based on the weight of each stratum in the overall population, CED then calculates the weighted average success rates and sampling errors, following the normal procedures for stratified sampling. If the CED wishes to select projects to be covered by PERs as above *but does not wish to combine the PER with the XASR-A findings in reporting annual success rates*, it draws the sample to be covered by XASRs from the full population for the year, without previously eliminating the projects to be covered by PERs. To the extent that specific projects may be selected for XASR-As and PERs, CED would use the PER ratings, rather than the XASR-A ratings in reporting on success rates, since CED will have carried out a more rigorous review in these cases.	

ANNEX 3: CRITERIA USED FOR APPLYING GOOD PRACTICE HARMONIZATION STANDARDS AND SUGGESTIONS FOR POSSIBLE CHANGES IN THESE STANDARDS

Std #	***Harmon-ization or Other Standards***	***Standard***	***Criteria Used for Applying Standards & (in bold font) Suggestions for Possible Changes in Standards***
		Instructions, Execution and Validation	**References to "instructions" in this section should be replaced by "guidelines," since CEDs do not—and should not—have authority to issue instructions to operational staff.**
		Instructions & familiarization:	
16	H	In consultation with operations departments, CED prepares, refines and disseminates instructions for the preparation of XASRs and PERs in sufficient detail to promote consistency and objectivity in execution scope, analysis and ratings. As part of dissemination efforts, some CEDs may wish to conduct workshops to familiarize the XASR teams with requirements and supporting documentation for achieving good-practice execution.	
17	H	*Good practice*: The instructions include ratings guidelines with benchmarks and standard reporting templates that include the performance ratings matrix.	
		Execution:	
19	H	The research for XASRs and PERs draws from a file review; discussions with available staff involved with the operation since its inception; independent research (e.g. on market prospects); a field visit to obtain company managers' insights and to the project site to observe and assess outcomes; and discussions with parties who are knowledgeable about the country, company and project (e.g. MDB specialists, company employees and auditors, suppliers, customers, competitors, bankers, any relevant government officials, industry associations, and local NGOs).	

ANNEX 3: CRITERIA USED FOR APPLYING GOOD PRACTICE HARMONIZATION STANDARDS AND SUGGESTIONS FOR POSSIBLE CHANGES IN THESE STANDARDS

Std #	*Harmon-ization or Other Standards*	*Standard*	*Criteria Used for Applying Standards & (in bold font) Suggestions for Possible Changes in Standards*
		Review and independent validation:	
22	H	CED conducts an independent review of each XASR to verify scope responsiveness, evident reliability of the analysis, impartiality and consistency in ratings judgments, and appropriateness and completeness of the identified lessons, and then, *for each randomly selected XASR to be used in the annual synthesis report on evaluation results*, prepares an XASR-A on the final-edition XASR that records its independent judgments on the report's quality in relation to the guidelines, assigned ratings and lessons.	
24	H	Following preparation of each draft XASR-A, CED reviews with the XASR team and its manager the basis for its judgments where its ratings differ from those in the final edition XASR.	
		Evaluative Scope	
		Performance dimensions evaluated:	
26	H	*Good practice*: The scope of the XASR (and XASR-A) or PER includes, at a minimum, • The *project's development or transition outcome*, i.e., the project's "results on the ground" relative to the MDB's mission. • The *MDB investment's profitability* (contribution to its corporate profitability objective), and • The *MDB's additionality and work quality* (additionality is defined in GPS 39; work quality is also referred to as bank handling, operational effectiveness, or execution quality).	**1. By covering additionality and work quality together, the GPS may not give adequate importance to these factors. Indeed, an evaluation could rate the development or transition outcome of a project as highly satisfactory even if the institution was not needed to permit the project to go forward in the way it did. The WGPSE may wish to discuss this issue.** **2. WGPSE should consider splitting this standard into (i) work quality, *per se*, and (ii) additionality, role and contribution. First, these two components address different issues. An MDB may do an excellent job in screening, appraising, structuring, supervising and administering its investments, but still it may be doing things that the private sector could have done on its own. Second, splitting these two components would facilitate greater comparability between EBRD and the other**

ANNEX 3: CRITERIA USED FOR APPLYING GOOD PRACTICE HARMONIZATION STANDARDS AND SUGGESTIONS FOR POSSIBLE CHANGES IN THESE STANDARDS

Std #	*Harmon-ization or Other Standards*	*Standard*	*Criteria Used for Applying Standards & (in bold font) Suggestions for Possible Changes in Standards*
			members. **3. EBRD proposes that an alternative good practice be added, based on EBRD's evaluation framework, i.e., using as performance indicators (a) its mandate-related indicators (transition impact, environmental performance & change, and additionality), (b) sound banking indicators (project & company financial performance & fulfillment of objectives other than transition objectives), and (c) Bank effectiveness-related indicators (excluding additionality, which would be included as part of the first performance indicator). This proposal would undermine the prospects of meaningful comparability.**
27	H	The operation's performance under each of these dimensions is analyzed according to standard indicators, and the operation's performance for each indicator is rated according to criteria and benchmarks specified in the guidelines.	
28	H	The performance reflected in the relevant indicator ratings is synthesized into ratings for each of the three performance dimensions, specified above.	
29	H	Project outcomes for each of the indicators are assessed on a "with v. without project" basis.	**Needs to be stated only with respect to development or transition outcome and related indicators, since implicit for investment profitability and work quality. Would be desirable to clarify.**
30	H	Assessments of development or transition outcomes for each of the development or transition outcome indicators take into consideration the sustainability of the results.	
		Indicators for the development or transition outcome:	
31	H	The project's development or transition outcome is based partly on the *project's contribution to the company's business success*, measured mainly: • For capital expenditure projects: by the project's after-tax financial rate of return (FRR); • For financial markets projects: by the project portfolio's profit contribution to the	For financial markets projects, MC ratings have been assigned even when contribution to the company's business success is measured by the performance of the sub-projects, a proxy for the contribution to the business success of the financial institution or fund financed. (See following para.)

ANNEX 3: CRITERIA USED FOR APPLYING GOOD PRACTICE HARMONIZATION STANDARDS AND SUGGESTIONS FOR POSSIBLE CHANGES IN THESE STANDARDS

Std #	*Harmon-ization or Other Standards*	*Standard*	*Criteria Used for Applying Standards & (in bold font) Suggestions for Possible Changes in Standards*
		financial intermediary or investment fund; • For other projects: by the project's profit contribution and the achievements of the company's at-approval business objectives.	**The standard for financial markets projects, as written, is unrealistic. Financial institutions in developing or transition economies are unlikely to be able to provide information on the contribution to their own profitability of the projects financed by an MDB loan. Moreover, the standard would not apply to financial markets projects that do not finance a specific set of projects. The standard should be revised.** **Standard should specify that FRR should be in real terms.**
32	H	The project's development or transition outcome is based partly on the *project's contribution to the country's private sector development and/or its development of efficient capital markets and/or its transition to a market economy.*	
33	H	The project's development or transition outcome is based partly on its *economic viability.* For non-financial markets operations: Based mainly on the project's net quantifiable economic benefits and costs, as measured by the project's real economic rate of return (ERR). For financial market operations: Rated on whether the sub-projects financed are economically viable (as reflected, e.g., in sub-project ERRs or the portfolio credit or equity IRR performance combined with the absence of portfolio concentrations in protected industries); whether the project has led to use of economic viability criteria in the intermediary's or investment company's investment decisions; and benefits to the economy. In most cases, quantitative information on the economic viability of sub-projects is not available to the MDB. The judgment, therefore, relies on assessing portfolio financial performance and an assessment of the extent to which the intermediary or investment company invests in protected industries.	**To conform to wording in GPS 31, "non-financial markets operations" should be replaced by "capital expenditure projects."**
34	H	The project's development or transition outcome is based partly on its *contribution to the country's living standards*. This rating reflects the project's economic benefits and costs to those who are neither its owners nor its financiers, i.e., customers, employees, government, suppliers, competitors, local residents, etc.	MC can be assigned if contribution to the country's living standards is covered by GPS 33.
35	H	The project's development or transition outcome is based partly on its *environmental*	MC requires that performance be benchmarked against

ANNEX 3: CRITERIA USED FOR APPLYING GOOD PRACTICE HARMONIZATION STANDARDS AND SUGGESTIONS FOR POSSIBLE CHANGES IN THESE STANDARDS

Std #	*Harmon-ization or Other Standards*	*Standard*	*Criteria Used for Applying Standards & (in bold font) Suggestions for Possible Changes in Standards*
		sustainability (benchmarked against compliance with the MDB's specified standards in effect (i) at investment approval and (ii) at the time of the evaluation). "Environment" includes the physical environment and, to the extent covered by the MDB's policies, also includes social, cultural, and health and safety impacts.	standards in effect at approval and standards in effect at evaluation. **It may be worth considering whether this standard should be separated into a good practice standard, calling for comparisons with the standards at investment approval, and a best practice standard, calling for comparison with the standards at approval and at the time of evaluation.**
		Indicators for MDB's investment profitability:	
36	H	*Good practice.* MDB investment's profitability is based upon the investment's *gross contribution* in relation to corresponding at-approval standards for minimally satisfactory expected performance.	**Gross contribution needs to be defined to make clear that it is net of financing costs attributable to loan operations and net of loss provisions.** **(Net profit contribution, cited in best-practice standard, needs to be defined. Should it refer to cost of particular transaction or average cost? Should it refer to direct cost or fully loaded cost?)** **Eliminate unnecessary differences in wording.**
		Indicators for MDB's work quality:	
37	H	The rating for the MDB's work quality is based partly on *at-entry screening, appraisal and structuring work*; i.e., how effectively the MDB carries out its work prior to approval of the investment.	
38	H	The rating for the MDB's work quality is based partly on its *monitoring and supervision quality*, i.e., how effectively the MDB carries out its work after approval of the investment.	
39	H	The rating for the MDB's *additionality and work quality* is based partly on its role, and contribution, i.e., the need for the MDB's participation relative to other available financing and the quality of the MDB's additionality from inception to evaluation. The rating judgment considers compliance with basic operating principles, the MDB's contribution to client capacity building objectives (as relevant), it's the operation's consistency with furtherance of the MDB's corporate, country and sector strategies, and its clients' satisfaction with the MDB's service quality.	**See comments on GPS 26.**

ANNEX 3: CRITERIA USED FOR APPLYING GOOD PRACTICE HARMONIZATION STANDARDS AND SUGGESTIONS FOR POSSIBLE CHANGES IN THESE STANDARDS

Std #	***Harmon-ization or Other Standards***	***Standard***	***Criteria Used for Applying Standards & (in bold font) Suggestions for Possible Changes in Standards***
40	H	Assessments of the MDB's work quality should be made independently of the ratings assigned for development or transition outcomes and MDB's investment profitability. These assessments, which are benchmarked against corporate good practice, reflect the quality of the MDB's contributions to good or bad outcomes, not the good or bad outcomes themselves.	
		Performance ratings—principles and benchmarks:	
41	H	Within the rating scales (e.g., ranging from *unsatisfactory* to *excellent*), there should be balance between positive and negative characterizations (i.e., if there are four ratings, two are less than good and two are good or better). The words used to describe these ratings should accurately reflect whether the judgments are less than good or else good or better.	
42	H	Each of the evaluated performance attributes is assigned a rating using a 4- to 6-point scale for each indicator ratings and a 4- to 6-point scale for each synthesis rating.	
43	H	The ratings benchmarks for the project's financial rate of return after-tax (FRR) can be determined by *either* of the following methods: ***Good practice (transitional).*** Using an arbitrary scale, e.g., 20% or more after taxes in real terms for excellent, 10-19.9% for satisfactory, 5-9.9% for partly unsatisfactory, and <5% for unsatisfactory. *Best practice*. Comparing the re-estimated after-tax FRR in real terms with each project company's weighted average cost of capital at the time the project financing is committed. Although this approach is also subject to problems, it is clearly better than the alternatives. See Attachment 2 of GPS.	MC rating assigned for good practice standard if member used *either* of the two methods specified. **The standard should be revised to make clear that an MC rating is appropriate if the member used either of the two methods specified.**
44	H	The benchmarks for the project's real economic rate of return (ERR) are set in relation to the MDB's ERR benchmarks used in approving or rejecting projects. The ERR benchmarks may be universal or may vary by country or business sector. They are not, however, the same as the ERRs projected at appraisal for specific projects.	
45	H	*Good practice.* Loan performance benchmarks are set in relation to the MDB's expectations at approval.	**Should clarify that this standard applies also to guarantees.**

ANNEX 3: CRITERIA USED FOR APPLYING GOOD PRACTICE HARMONIZATION STANDARDS AND SUGGESTIONS FOR POSSIBLE CHANGES IN THESE STANDARDS

Std #	*Harmon-ization or Other Standards*	*Standard*	*Criteria Used for Applying Standards & (in bold font) Suggestions for Possible Changes in Standards*
46	H	*Good practice.* Equity investment rates of return are benchmarked (or discounted if an NPV measure is used) against standards for minimally satisfactory expected performance at approval.	Absent formal standards for minimally satisfactory expected performance at appraisal, a statement by MDB Management on *de facto* standards or a review of a random sample of new investment operations can provide the basis for determining minimally satisfactory expected performance at approval and, hence, benchmarks to be used for evaluation. Since few members are likely to adopt the best practice standard, a good practice standard for combining the loan and equity returns may be useful.
47	H	Ratings of non-quantitative indicators require that relative qualitative judgments be made. The criteria should reflect the extent to which performance has been consistent with the MDB's policies, prescribed standards for corporate sustainability and recognized good-practice standards. The criteria for the judgments should be clearly specified in the instructions for the preparation of XASRs and in the CED's annual review. .	
48	H	The synthesis ratings for the three dimension**s** (development or transition outcomes, profitability to the MDB, and the MDB's work quality) reflect summary qualitative performance judgments based on the underlying indicator ratings. They are not simple averages of the indicator ratings.	**"Dimensions" should be plural.**
		Standard XASR attachments: These attachments provide the basis for review and independent verification of the XASR's judgments and conclusions. They include:	
49	H	Details of the financial and economic rate of return derivations (with transparent assumptions and cash flow statements).	
50	H	For each safeguard dimension addressed in the MDB's environmental and social guidelines, a comprehensive summary of environmental, worker health and safety, and social outcome compliance information with sufficient evidence from a field visit and/or client reporting to support the assigned outcome and related MDB work quality ratings.	

ANNEX 3: CRITERIA USED FOR APPLYING GOOD PRACTICE HARMONIZATION STANDARDS AND SUGGESTIONS FOR POSSIBLE CHANGES IN THESE STANDARDS

Std #	*Harmon-ization or Other Standards*	*Standard*	*Criteria Used for Applying Standards & (in bold font) Suggestions for Possible Changes in Standards*
		Annual Reporting and Process Transparency	
		Annual synthesis reporting: Annual Review	
51	H	CED prepares an annual review addressed to the MDB's management, staff and Board of Directors. The scope of the annual review includes, *inter alia*, a synthesis of the CED's validated findings from all XASRs and PERs generated and reviewed during the period covered.	
52	H	The annual review should provide sufficient information to make the reader aware of possible biases in the sample of projects covered by the annual review. Consequently, the annual review: • Describes how the population was identified, how the sample was selected and, if stratification was applied or part of the sampling was non-random, states the rationale. • Reports on the number of XASRs and PERs for the year. • Includes an annex profiling the important characteristics of the evaluated sample against the population (cf. GPS 15). • Reports on the mean number of months between the dates projects reached early operating maturity, as defined in GPS 11 and 12) and the dates the corresponding XASRs or PERs were issued. • If less than 100% of the population has been covered, provides information on statistical confidence levels and states explicitly whether reported success rates can be attributed to population.	MC requires that practices be consistent with all bullet paras. **The WGPSE may wish to reconsider inclusion of the requirement that the MDBs report on the mean number of months between the dates projects reached early operating maturity, as defined in GPS 11 and 12) and the dates the corresponding XASRs or PERs were issued. The focus, rather, should be on confirming that all projects reaching early operating maturity are included in the population for the appropriate year.**
53	H	The annual review either (i) describes the ratings criteria and benchmarks in an annex or else (ii) refers to a website providing this information.	
54	H	The ratings reported should be those of CED.	
55	H	*Good practice.* CEDs should disclose the differences between CED and operating staff ratings and the materiality of the differences. Where CED ratings are reflected partly in XASR-As and partly in PERs, the CED should disclose the differences between CED and operating staff ratings separately for the XASRs and the PERs. The disclosure is made in global terms, not on a project-by-project basis and is limited to differences in binary outcome and work quality success ratings.	**EBRD considers that this standard should not be treated as a harmonization standard.**

ANNEX 3: CRITERIA USED FOR APPLYING GOOD PRACTICE HARMONIZATION STANDARDS AND SUGGESTIONS FOR POSSIBLE CHANGES IN THESE STANDARDS

Std #	***Harmon-ization or Other Standards***	***Standard***	***Criteria Used for Applying Standards & (in bold font) Suggestions for Possible Changes in Standards***
56	H	*Good practice:* For each rating dimension and indicator, the annual review shows the *proportion of the evaluated sample* in each performance-rating category.	For institutions with a small number of projects, I assigned MC rating even if raw numbers, rather than proportions, given.

		Process transparency: Annual Report. (The annual report can be included in the annual review if an MDB wishes to do so.)	
62	H	CED reports annually to the MDB's management and Board on the quality and efficacy of the MDB's evaluation system, including the self-evaluation system, any gaps in coverage of the MDB's operations, the work of CED, the generation and application of lessons learned in new operations, and any differences between the MDB's practices and the GPS.	MC requires that practices be consistent with all elements of standard. **EBRD suggests splitting this standard to separate out reporting on differences between MDB practices and the GPS.** **The word "harmonization" should be added before "GPS." CEDs need to provide information to readers to allow them to recognize the limits to the comparability of evaluation results. The standards relating to comparability are all harmonization standards.**