

GOOD PRACTICE STANDARDS FOR EVALUATION OF MDB-SUPPORTED PUBLIC SECTOR OPERATIONS

I. INTRODUCTION

A. BACKGROUND AND CONTEXT

a. MDB Task Force Report

1. According to the March 1996 report of the Development Committee (DC) Task Force on Multilateral Development Banks (MDBs)

“currently, it is not possible to compare their operational results, or even to describe them in a common language. Many 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 and kept up to date over time, with best practices in evaluation techniques being identified and disseminated. 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.

The heads of the five MDB evaluation units have agreed recently to begin meeting on a regular basis. We suggest that they be charged with elaborating common evaluation standards, including performance indicators; exchange experience with evaluation techniques, share results; and become the repository of best evaluation practices. The immediate task would be to develop, within a specified period, methodology and criteria for assessing and rating the MDB's operational performance and development effectiveness.” (Underlining added for relevance to this paper).

b. MDBs response

2. In an immediate response to the Development Committee Task Force recommendationsⁱⁱ, the Evaluation Cooperation Group (ECG) was formed. It consisted initially of the heads of the evaluation units of the five MDBs referred to in the Task Force Report: African Development Bank (AfDB), Asian Development Bank (AsDB), European Bank for Reconstruction and Development Bank (EBRD), Inter-American Development Bank (IADB) and the World Bank (WB). The European Investment Bank (EIB) joined in 1998 (re. ECG governance, paragraph 3.). In March 1998 the original five MDBs reported to the DC:

“The [Evaluation Cooperation] Group will continue its efforts to make evaluation results comparable and to have their findings properly translated into operational standards. Meeting in Hong Kong in October 1997, the MDB presidents strongly endorsed further intensification of collaboration among MDB evaluation units in harmonizing evaluation standards and activities, defining more

effective linkages between independent and self-evaluation The harmonization dialogue will be extended to country evaluations, non-lending services, and evaluation of private sector operations.ⁱⁱⁱ

3. ECG governance provides that “the founding members [and their affiliates] are committed to the original harmonization mandate set by the MDB Heads in 1996. Other core voting members are not bound by this obligation and therefore do not vote with respect to harmonization matters.”^{iv}

B. PURPOSE, GUIDING FRAMEWORK, PRINCIPLES APPLIED, SCOPE AND DEFINITIONS

a. Purpose of this paper

4. In following up on the foregoing commitment by the MDBs, and consistent with widely accepted evaluation principles^v, this paper is to help define and establish good practice standards* for MDBs’ evaluation of public sector operations supported by a MDB loan with sovereign guarantee (for definitions see paragraph 10). The standards aim at the ECG objectives of harmonization of evaluation criteria and processes. They are also expected to help in the MDBs’ linkage between evaluation and operational standards in pursuit of corporate missions and objectives, and in the learning from experience between MDBs for improved results.

b. Guiding framework

5. The Members of the ECG’s Working Group on Evaluation Criteria and Ratings for Public Sector Evaluation (WGEC) - the Members - have developed this paper within the general framework of the OECD-DAC evaluation principles, the ECG principles and objectives, and the coherence principles of results-based management. The principal documents underlying this paper are listed in Attachment 1. The guiding (“core”) standards listed in this paper are currently in practice to some extent in all Members. They recognize that differences among them may affect the pace at which harmonization can be achieved. Further, harmonization is interpreted as comparable results to the extent possible and not standardization. It is seen as a means to an end. Namely, the reduction of transactions costs for borrowers/clients and the facilitation of joint evaluation among development assistance partners.

* Two sets of standards, “minimum” or “core” standards and “good-practice” standards, are being used in this paper. “Core standards” are those that have been adopted by the MDBs bound by the harmonization mandate (see paragraph 3); “good-practice standards” are seen by MDBs as desirable improvements over present practice and conforming with the ECG harmonization objectives. The ECG used this distinction of two types of standards in 1996¹ in the context of the “Comparative Analysis of MDB Completion Reporting and Performance Review”. True, the Development Committee Task Force Report on the MDBs referred to “best practices” (see paragraph 1). However, in the light of the practice used so far within the WGEC, it is suggested that the latter term be reserved for a future stage when further progress toward harmonization may adopt some higher than good practice standards. Thus, all standards spelled out below are “core standards” except where specifically referred to as “good-practice standards”.

c. Application of evaluation principles

6. The practice standards presented in this paper
- (i) derive from the DAC principles of impartiality and independence, credibility, transparency, participation, inter-agency cooperation, feedback and follow-up;
 - (ii) reflect good-practice design features in accordance with widely accepted professional standards, procedures and independent verification; are tailored to be responsive to the special characteristics and institutional learning and accountability needs of the MDBs regarding their public sector operations keeping in mind that they have to report their evaluation findings to their Board and to the public;
 - (iii) are designed to be consistent with the MDB's operational policies and standards that were in place at Board approval (for accountability) and as evolved up to the time of evaluation (for learning);
 - (iv) are informed by the MDB's performance measurement methods and experience to date in applying them in the evaluation of public sector operations; and
 - (v) recognize that while the individual MDBs' missions vary, the Working Group's harmonization agenda implies that evaluation practice standards must be sufficiently consistent in their design and application to achieve the objectives referred to which the MDBs have committed themselves in the light of the MDB Task Force Report.

d. Scope and limitations of paper

7. This paper covers the ECG's agenda for evaluation harmonization at the level of *individual public sector operations* and corporate level synthesis reporting in relation to them. At the same time, recognizing that harmonization involves a progression from first principles to details, these standards reflect a prioritization within that process and thus involve limitations.

8. *Scope:* To meet part of the ECG's primary agenda for harmonization for institutional accountability, the scope of this paper extends to good practice standards in the three principal areas that underlie an effective evaluation system:

- (i) governance;
- (ii) evaluation criteria; and
- (iii) processes.

A concluding section outlines briefly proposed good-practice implementation and monitoring arrangements.

9. *Limitations:* They emerge from the following:

- (i) these operation- and related synthesis-level standards are only part of the MDBs' public sector operations evaluation activities (viz. country, sector and other thematic specific evaluations) that are needed in support of broader accountability judgments of institutional effectiveness, or as may be judged efficacious by individual MDBs for meeting their mission-specific institutional accountability and learning needs^{vi};

- (ii) even for individual operations, relative to informing strategy, the Members recognize that the portfolio coverage may not necessarily provide a sufficient basis for learning needs down to all sector/sub-sector levels^{vii};
- (iii) most importantly to this paper's primary purpose, these standards do not provide a complete basis for harmonization and inter-MDB results transparency until the Members have agreed on more detailed sub-criteria and indicators (including on rating guidelines and on benchmarks).

II. GOOD PRACTICE STANDARDS

A. GOVERNANCE OF MDB EVALUATION SYSTEM

a. Principle of independent evaluation at MDBs

10. The Development Committee Task Force on Multilateral Development Banks recommended in 1996 that:

“the heads of the five MDB evaluation units should be allowed to issue final evaluation reports to the MDB President and Executive Directors without prior clearance by anyone outside the unit”^{viii}.

b. Structure and role of independent evaluation

11. The CED operates according to a Board-approved mandate (policy statement) that specifies its mission, scope of responsibilities, reporting structure and key operating principles. The Board oversees the CED's work through an evaluation oversight committee. The governance arrangements are designed to ensure CED's independence, its relevance to the MDB mission and the delivery of its corporate accountability and learning value-added. Specifically, CED's reporting line, staff and functions are organizationally independent from the MDB's operational, policy and strategy departments and related decision making^{ix}. The CED operates with full autonomy but in close consultation with the MDB's other departments to ensure, as far as possible (subject to the primacy of sound evaluative principles and practices), coherence of corporate standards (as among operations, portfolio and strategy analysis and evaluation), and good prospects for corporate ownership of CED's findings and recommendations for improvement.

12. The CED's role is to ensure the relevance, quality and impartiality of the products of the MDB's evaluation system. To meet this responsibility, under its mandate CED has a scope of responsibility that extends to all the determinants of the MDB's operational results. To help ensure that the independent evaluation work responds to the MDB's needs for information to guide policy and operational decisions, CED's annual work program, the principal determinant of the CED budget, is widely discussed during preparation with the Board, managers and Bank staff.

13. CED has unrestricted access to (a) the MDB's staff and records considered essential by CED in the context of an actual evaluation, and (b), to co-financiers and recipients of the MDB's loans and grants. CED's director holds grade-rank at least equal to the head of any major operational unit. CED staff, most of whom have had extensive development experience, may work in other departments before or after working in the CED; but they do not participate in evaluations or reviews of operations for which they were responsible before joining CED. CED transmits its final products to the President and Board without any clearance from management.

14. In relation to the evaluation of individual public sector operations, CED executes its mandate by:

- (i) defining, writing and refining the MDB's evaluation standards, instruments and related guidelines, in consultation with operations management, relevant corporate functional departments, and the Board's oversight committee, informed by internal experience with their application (e.g., on the relevance of sound monitoring and evaluation systems), feedback from stakeholders, and evolving internal and external good practice;
- (ii) coordinating with those MDB units responsible for strategy formulation and corporate scorecard development and related reporting (as relevant) to ensure that, consistent with the coherence principle of good-practice results-based management (1) the evaluation measures, standards and benchmarks will serve to reinforce the corporate mission, objectives and policies, inform corporate learning needs and align as far as possible with the general reporting framework, and (2) the annual review's synthesis ratings are included in the integrated corporate performance reporting;
- (iii) developing and carrying out a work program for evaluating completed individual public sector operations;
- (iv) maintaining an evaluation database including all relevant characteristics of evaluated operations to support independent evaluation studies and annual review analysis;
- (v) synthesizing the CED-verified CR findings, supplemented by its performance evaluation reports (and other evaluation studies) in annual reviews to the MDB's management and Board;
- (vi) disseminating the findings of all desk-reviews of CRs, of performance evaluation reports and annual reviews and studies through instruments that allow easy retrieval, absorption and application of lessons by policy and strategy departments, and by operational staff, and, as defined under the MDB's disclosure policies, to persons and institutions outside the agency;
- (vii) monitoring and reporting annually to management and the Board on the quality and efficacy of the MDB's evaluation system, including application of lessons in new operations.

c. Role of self-evaluation

15. In the interest of efficiency, ownership of CRs and annual review findings and rating judgments, learning from experience and providing accountability for results, the departments responsible for operations and portfolio management, supported by relevant functional specialist units:

- (i) contribute to the development of evaluation guidelines and standards, including providing process feedback on problems encountered in applying them;
- (ii) prepare a CR for every completed operation ;
- (iii) establish the CR delivery schedule with a view to spreading the review load for the agreed number of reports throughout the year;
- (iv) execute the CRs in accordance with evaluation guidelines;
- (v) ensure CR quality and delivery within agreed CR schedule; and
- (vi) identify and bring to bear relevant lessons and findings from the evaluation system in identifying and appraisal of new public sector operations and supervision/monitoring of portfolio.

B. EVALUATION CRITERIA

a. Scope of using evaluation criteria

16. CEDs assess the development effectiveness of operations having completed their implementation phase by systematically applying a set of evaluation criteria. In addition, CEDs evaluate the performance of the MDB and of the borrower. CEDs also evaluate the quality of CRs reviewed. These assessments typically result in ratings according to pre-defined scales. These evaluations, and any independent evaluation carried out by CEDs during implementation, are grounded in the same policies and criteria that are used in appraisal of new operations and in the self-evaluation of CRs.*

17. The scope of both the CR and the full performance evaluation reports is comprehensive in encompassing all performance attributes and dimensions that bear on the operation's success as considered against the MDB's mission, the operation's relevance, efficacy, efficiency , sustainability and institutional development impact. The operation's performance under each of these dimensions is analyzed according to sets of indicators, and the operation's performance in each is rated according to specific guidelines.

* This section on Evaluation Criteria is based on core standards evidenced by CED guidelines for their full performance evaluation reports, which are not necessarily fully consistent with CR guidelines when it comes to evaluation criteria. There is likely to be a need for internal harmonization to varying degrees between guidelines for the full performance evaluation reports and for CRs, to lead to a situation where the criteria addressed are covered in the guidelines for both stages/products. There remains the broader issue of harmonizing the criteria fully with upstream supervision reports and appraisal documents where the language may not be consistent with the language for later stage evaluation.

b. Performance dimensions evaluated

18. In their evaluation of public sector operations, MDBs use a set of evaluation criteria which the ECG approved in 1999. WGEC has enhanced the definitions for these evaluation criteria recognizing that, within the MDBs, evaluation criteria have evolved further since 1999. Most important, it aligned them with recent work done by the DAC* which should facilitate harmonization with bilateral agencies. In addition, WGEC was concerned that the definitions, though designed for application at the project level, should be broad enough for their eventual use in evaluations that go beyond individual operations with MDB financial assistance. Four of the core evaluation criteria form the basic evaluation building blocks; one is an aggregation of these four (see box 1). This group of core criteria, culminating in the aggregate indicator, represent the core standard towards achieving the most comparable operational results across MDBs. The remaining criteria are complementary to the core group and their formulation and application may be expected to differ across MDBs given varying institutional priorities, including two relating to the performance of the individual MDB and the Borrower rather than that of the operation itself. Together, *these evaluation criteria provide a systematic framework for assessing MDB supported public sector operations.*

19. In applying these criteria MDBs are cognizant of the need for giving systematic attention to the relative importance of each of the criteria and the interrelations among them. As to the latter, WGEC notes that among the four core criteria, the forward looking part of assessments under criteria 2. and 3. captures the essence of project sustainability, though the latter is seen as important enough in its own right to deserve specific consideration. Thoroughness in the analysis and transparency in addressing each criterion are seen as critical factors for ensuring minimum quality standards in the evaluations.

* Development Assistance Committee (DAC). Working Party on Aid Evaluation. Glossary of Evaluation and Results Based Management Terms. June 2001. All definitions listed in Box 1 correspond to the respective DAC definitions unless otherwise noted (though no DAC definitions are available for 6., 7., and 8.). However, throughout Box 1 the term “project” is used to reflect MDB financed operations rather than the much broader term “development intervention” found in the DAC Glossary.

BOX 1 Evaluation Criteria: WGEC Definitions

Core Criteria

1. RELEVANCE –The extent to which the objectives of a project are consistent with beneficiaries' requirements, country needs, global priorities and partners' and donors policies. *Note: Retrospectively, the question of relevance often becomes a question as to whether the objectives of a project or its design are still appropriate given changed circumstances.*

2. ACHIEVEMENT OF OBJECTIVES (“EFFICACY”) –The extent to which the project's objectives were achieved, or are expected to be achieved, taking into account their relative importance, while recognizing any change introduced in the project since Board approval.

3. EFFICIENCY – Extent to which project benefits/output are commensurate with resources/inputs (funds, expertise, time etc.).

4. SUSTAINABILITY –The probability of continued long-term benefits, and the resilience to risk of the net benefit flows over the intended useful project life.

5. AGGREGATE PROJECT PERFORMANCE INDICATOR – A single measure of overall project performance taking into account the evaluation findings under the criteria listed under 1. through 4. above.

Complementary Criteria

1. INSTITUTIONAL DEVELOPMENT – The extent to which a project improves or weakens the ability of a country or region to make more efficient, equitable, and sustainable use of its human, financial and natural resources, for example through: (a) better definition, stability, transparency, enforceability and predictability of institutional arrangements and/or (b) better alignment of the mission and capacity of an organization with its mandate, which derives from these institutional arrangements. Such impacts can include intended and unintended effects of a project.

2. OTHER IMPACTS – This criterion can include other areas of special focus proposed by WGEC Members, including:

- (i) **POVERTY REDUCTION** - extent to which project achieved planned poverty reduction impact; unintended impact should also be considered.
- (iii) **TRANSITION IMPACT**
- (iv) **ENVIRONMENTAL IMPACT**
- (v) **ADDITIONALITY**
- (vi) **OTHER IMPACTS**

3. BORROWER PERFORMANCE – Adequacy of Borrower's assumption of ownership and responsibilities during all phases. Main focus on effective measures taken by Borrower to establish basis for project sustainability, especially – and right from the identification stage – through fostering participation by the project's stakeholders, in addition to its own support.

4. MDB PERFORMANCE – Quality of services provided by MDB during all project phases. Main focus on MDBs' role in ensuring project quality at entry, and that effective arrangements were made for satisfactory implementation and future operation of the project.

20. WGEC considers the following applications of the above evaluation criteria as *good practice standards*:

- *Relevance*: evaluation focuses on consistency of project with (i) the needs of (a) the project's beneficiaries, (b) a country's development strategy and (c) the MDB's assistance strategy for that country, as reflected in a Country Assistance Strategy or equivalent document, and (ii) MDB's statutory requirements, comparative advantage

and policy priorities. "Relevance" refers to current circumstances, i.e., it is based on (i) and (ii) as they stand at the time of evaluation. If there have been significant changes under (i) or (ii) since Board approval, the evaluation will reflect these. Since evaluations cover both accountability and lesson learning, restructuring of project objectives in response to the operating environment is itself a feature that deserves careful evaluation.

- *Achievement of Objectives ("Efficacy")*: Evaluation against objectives enhances accountability. Different categories of objectives are normally specified, such as physical, social, economic, financial, institutional, and environmental, as well as policy changes. Often there are multiple objectives: the evaluator will have to make judgments about the weight to be placed on the separate objectives in determining a sound evaluation under this criterion.
- *Efficiency*: Economic and financial rates of return should be used where feasible. Special care will be necessary when treating costs and benefits for projects restructured during implementation. Transparency in evaluations is essential so as to avoid the introduction of doubtful sunk cost assumptions. Where calculation of economic and financial rates of return is not feasible, the evaluation should address explicitly cost-effectiveness (considering the cost of alternative ways to achieve project objectives; or unit costs for comparable activities), and timing (were objectives achieved on time; what were the benefits of early completion, or costs of late completion). Such evaluation may not be feasible in the case that project design or monitoring did not provide for cost indicators.
- *Sustainability*: for a systematic application of this criterion, the following factors should be considered (their priorities/weights will vary according to the nature of the project): (i) technical soundness, (ii) government commitment, including supportive legal/regulatory framework, (iii) socio-political/stakeholder support, (iv) economic viability, (v) financial viability, (vi) institutional, organizational and management effectiveness, (vii) environmental impact, and (viii) resilience to exogenous factors. MDBs will give special attention to the continuation of project benefits over time and after external financing ceases. Sustainability, together with institutional development impact discussed next, may already be included under the preceding three criteria, especially for projects that are subjected to an economic and financial rate of return assessment. Nevertheless given the need to take account of risks in a generally volatile operating environment, and the emphasis given by MDBs on environmental impact and other factors underlying project sustainability, the *explicit* assessment of these factors is essential for a high quality evaluation – though, as already noted, attention will have to be given to prevent double counting..
- *Aggregate Project Performance Indicator*: This is not an independent evaluation criterion. The goal sought by this criterion is to aggregate, through a single indicator, overall project performance. The challenge for the evaluator is to (i) ensure completeness, i.e., to account for all the projects' effects that can be measured under

any of the core evaluation criteria (listed above), and (ii) weigh the relative importance of all effects in a fully transparent fashion.

- *Institutional Development:* evaluation encompasses, as applicable, the project's effects on (i) the broad scope of institutional development: the formal laws, regulations and procedures, and informal norms and practices that govern social and economic interactions and exchanges between people, and on (ii) the organization that operates within these broader institutional arrangements.
- *Other Impacts:* learning from the experience with past operations, MDBs have introduced special areas of focus to improve their development effectiveness. At this stage, several areas have been identified and recognized on a sufficiently broad basis so as to require a good practice standard, e.g., *poverty reduction*. This will require further deliberations among its members before a general good practice standard can be developed.
- *Borrower Performance:* conceptually, this criterion (and the following one on MDB performance) is treated separately from the foregoing criteria that address project performance. It focuses on processes that underlie the Borrower's effectiveness in discharging its responsibilities as the owner of a project.. This includes, importantly, the Borrower's efforts and success in establishing a lasting support for a project by its beneficiaries, thereby laying the basis for project sustainability.
- *MDB Performance:* evaluation encompasses (separately) the quality, benchmarked against corporate good practice, of the MDB's at entry-screening, appraisal and supervision work, its role and contribution (the need for the MDB's participation relative to other available financing, and the quality of the MDB's delivered additionality over the operation's life from inception to evaluation). It considers compliance with basic operating principles, the operation's client capacity building objectives (as relevant), consistency with furtherance of the MDB's corporate, country and sector strategies, and its client service satisfaction.

c. Ratings

21. For each rated criterion, MDBs use an even number (mostly four, exceptionally six for greater differentiation) of rating scale points. For the sake of validity, credibility, transparency and comparability, they apply a clearly defined rating for each scale point that reflects a pre-defined set of ranked value terms. Scales are symmetrical in nature (with due regard to the need for eliminating non-evaluable and not applicable cases). Evaluators provide a justification for each rating, where necessary or useful.

22. A distinction is made between *indicators* and *ratings*. Ratings reflect the informed judgment by the evaluator, based on qualitatively or quantitatively captured indicator(s). Ratings for non-quantifiable indicators require that qualitative judgments be made. The criteria for these judgments should be clearly spelled out in the evaluation

guidelines, and reflect the extent to which performance has been consistent with MDB policies and good-practice operational standards.

23. When MDBs use an “aggregate project performance” rating they derive this rating through weighting the ratings attached to the core evaluation criteria that form the basic building blocks. Although an aggregate project performance indicator may be useful for corporate reporting and comparisons among projects, regions, sectors and MDBs, there is a risk that excessive focus on this indicator may diminish lesson learning by Operations.

d. Learning Lessons

24. Evaluation should draw concise, prescriptive lessons placed in context. They should relate the experience of evaluated operations to the pattern of past lessons in the country or sector, and focus on what the MDB might have done to obtain better results.

25. Lessons may be presented in a four question format such as the following:

- (i) what was expected at approval;
- (ii) what actually happened (providing information about the context);
- (iii) what went particularly well or wrong; and
- (iv) what is to be learnt.

C. PROCESSES AND DOCUMENTATION

a. Evaluation timing

26. For a reliable assessment of an operation’s development effectiveness, evaluative judgments must be based on full completion of the operation’s implementation phase and an initial knowledge about the prospects for the operational phase. For this reason, a CED evaluation is carried out in all cases after the operation has been fully implemented. Since core standards provide for the responsible operational units to prepare a CR after completion of the implementation phase of the operation which will be used by the CED as a source for its evaluation, MDBs’ performance evaluation reports will show a significant delay from the date of project completion (at present some three years or even more, except for EBRD with a much shorter interval^x).

27. Evaluation timing may involve a tradeoff between the desire of a CED for timely learning feedback (while benefiting from the fresh memory of MDB staff involved in project supervision) and the availability of sufficient operational experience, in particular in order to assess the operation’s sustainability prospects^{xi} [see also reference to impact evaluation in paragraph 38].

b. Coverage and selection criteria

28. MDB operational units prepare CRs for all completed operations. CEDs in turn carry out an independent review of completion reporting through a two stage

performance review process utilizing (1) desk reviews for a sample of CRs and (2) in-depth full performance evaluation report for selected projects. Having established with the operational units an annual program of CRs scheduled for preparation, the CED defines the total annual program for desk reviews and for full performance evaluation reports. In case of a CR backlog from previous year(s), allowance is made specifically to reduce it substantially or eliminate it altogether.

29. The sample size needed depends on the degree of the desired disaggregation of the overall results to be drawn from the total evaluation findings in the CED annual reporting. In order for the CED to establish a sufficient basis for such reporting, it determines a combination of (1) desk reviews and (2) full performance evaluation reports for a representative sample of sufficient size and coverage of the total number of completed and reviewed operations. The desk review sample is set in the CED's annual work program agreed with the Board or management.

30. The ratio of full performance evaluation reports to completed projects, as reflected by CRs, is set by the MDB Board or management as part of CED's annual work program. Ratios and balanced qualitative selection criteria for full performance evaluation reports should be clear and transparent; the same applies to the statistical robustness of the samples chosen, when random sampling is used (spelling out the confidence level and sampling error). Examples of selection criteria are: potential to yield significant lessons, potential for ongoing or planned country/sector/thematic or corporate evaluation, CRs of outlying quality, areas of special interest expressed by Board. Selection of individual full performance evaluation reports will be verified during the desk reviews of CRs. To increase the attention paid by the Board and senior management to CRs, CEDs regularly assess compliance with CR guidelines in terms of submission rates and quality.

c. Documentation and reviews

31. (1) Completion Reports (CR): The CED either drafts or reviews CR guidelines in order to ensure consistency among the evaluation criteria and ratings within the MDB's overall evaluation system. Close cooperation between CED and operational staff is needed to ensure sound application of these guidelines. For accountability and learning purposes, and to promote a results-orientation among staff, operational staff prepare the CR. For improved application or utilization, CRs are submitted for management review in order to encourage timely feedback into operations. Borrower participation in the completion reporting process for public sector operations, including a borrower's report and borrower's comment on draft report, is encouraged and monitored as a means of improving quality as well as borrower evaluation capacity and ownership.

32. CED assesses compliance with CR guidelines in terms of submission rates and quality. The CED may provide assistance through CR draft review that improves CR quality, especially for operational staff not yet very familiar with the task. Two areas where CRs merit special attention, are: (i) their effort at re-estimating economic performance, i.e., re-estimation of ERRs and FRRs or cost-effectiveness analysis for

public sector operations that are not subject to a cost-benefit analysis at appraisal; and (ii) their focus on how well future operation needs are addressed in order to increase the likelihood for sustainability.

33. *Good-practice:*

1. *Guidelines for preparing CRs for adjustment or policy-based lending operations* are elaborated specifically to meet the special evaluation requirements for such operations, if necessary;
2. *for co-financed operations, the joint preparation by MDBs of CRs enhances quality and partnership, and reduces costs to MDBs' borrowers/clients.* Stakeholders, particularly the ultimate beneficiaries, participate in the CR preparation;
3. elements of *plans for future operations* include a system for monitoring and evaluation, performance indicators, and proposed MDB follow-up actions.

34. (2) Full Performance Evaluation Reports : CED staff carries out full performance evaluation reviews leading up to a Performance Evaluation Report in accordance with CED guidelines which were prepared in consultation with the MDB's operational and functional departments, and the agreed criteria for selecting full performance evaluation reports. CED uses internal peer reviews to improve the full performance evaluation report's quality, and seek operations management comments. Borrowers participate in the preparation of the full performance evaluation reports, including through the provision of comments on draft reports. For improved application or utilization, full performance evaluation reports are submitted for management review to encourage feedback to operations.

35. *Good-practice:*

1. *Stakeholders, particularly the ultimate beneficiaries, participate in the preparation of full performance evaluation reports.*
2. *for co-financed operations, the joint preparation of full performance evaluation reports enhances quality and partnership, and reduces cost to MDBs' borrowers/clients.*

36. *Good-practice:* Primarily for lesson learning, CED carries out, on a selective basis, *impact evaluations*^{xii} after a project has been in operation for an extended period (normally not going beyond ten years after completion date because of the increasing problem of attribution of costs/benefits over time). This instrument helps CED in identifying issues, both problems and new opportunities, that are not yet evident at the time of a CR or PR, and thereby providing the MDB with feedback for future operations and policy development that might otherwise have been missed.

d. Annual reporting

37. CED prepares an annual review addressed to the MDB's management, staff and Board whose scope, *inter alia*, includes a synthesis of the CED's validated findings from CRs and its own full performance evaluation reports generated and reviewed during the

period covered. The ratings criteria used in these evaluations should be clearly spelled out. All ratings reported are those from the CED and where there are differences from those given in the CRs these should be disclosed.

38. From an analysis of (a) the ratings and lessons learned patterns, (b) other relevant CED evaluation studies, and (c) management's implementation progress on previous CED review and special study recommendations, the annual review identifies recommendations to management and for the Board for improving the MDB's development effectiveness. CED maintains a tracking system for recording follow-up steps taken on each endorsed recommendation, ensuring that follow-up is reported in subsequent annual report(s). Good Practice: the tracking system includes a limited number of actionable recommendations.

e. Dissemination and lesson application

39. The CED makes available to all MDB staff, a range of user-friendly, high quality dissemination products covering the CR desk review/full performance evaluation report findings, the annual review and CED special studies. Various approaches may be used to enhance dissemination and application:

- (i) printed evaluation reports with executive summaries
- (ii) electronic databases of evaluations and lessons learned
- (iii) syntheses of lessons learned
- (iv) presentations incorporating visual aids for decision-makers
- (v) workshops and meetings with operations units
- (vi) giving the Board and management a role in evaluation topic and issue selection
- (vii) review project designs for lessons learned
- (viii) open evaluation recommendations periodically reviewed with Board and management.

Among these approaches, CED places primary reliance on its intranet website for document posting, widely notifies staff of new items through the corporate website and maintains and enhances learning lessons system. The goal is to:

- (i) draw attention to the existence of relevant evaluation findings;
- (ii) make the material user-accessible and easily digestible; and
- (iii) have measures for monitoring of accessing lessons^{xiii}.

40. Members recognize that achieving awareness and internalization of lessons and their systematic application in new operations is the weakest link in the evaluation learning cycle^{xiv}:

- (i) operational managers are responsible for ensuring that past lessons have been systematically searched, identified and applied in new operations;
- (ii) standard processing documentation for new operations includes a prompt, in the early stage documents, for relevant past lessons, complemented by a prompt, in final decision-stage documentation, for how the past lessons have been addressed in the structuring of the new operation and in its appraisal ;

- (iii) in its annual evaluation reporting, CED reviews and reports to management and the Board on the quality of responsiveness to these prompts and other evidence of lesson application.

41. *Good-practice:* For ensuring lessons applications, MDB good-practice is that internal corporate reporting (up to the Board) is centered around an integrated performance measurement system that is (1) broadly aligned with the evaluative framework (viz. performance indicators for development impact and MDB effectiveness/efficiency); (2) applies coherent and consistent success standards (both as to scope and benchmarks) across the project cycle (identification, appraisal, supervision and evaluation), across operations, and up and down the reporting hierarchy (project-, department- and corporate-level); and (3) integrates evaluation findings (success rates) into the scorecards^{xv}.

f. Disclosure

42. Evaluation reports should normally be disclosed. The MDB's disclosure policy for evaluation products should be explicit and consistent with the MDB's general disclosure policy.

III. THE FUTURE: IMPLEMENTATION AND PROGRESS MONITORING

a. Harmonization of evaluation methods

43. Members intend to review periodically their evaluation methods, including rating systems, and guidelines with the aim of assessing and reporting the development effectiveness of their public sector operations according to consistent standards and advancing the ECG harmonization agenda. To this end, they share all evaluation documentation including completed evaluation results, self-evaluation instructions, rating guidelines, best-practice reports and annual reviews, taking into account the disclosure policies. As Members reach agreements on application of these standards to different instruments (e.g., technical assistance or policy-based lending) they may document them in subsequent refinements of this paper. Members also consider that harmonization facilitates joint evaluations, and vice versa. The harmonization process can be enhanced by taking into account the GPS evaluation criteria in the CED evaluation guidelines and in the design of the evaluations.

b. Progress monitoring

44. Each Member will report on implementation issues faced with the core standards and on its progress toward good-practice standards in successive WGEC meetings for the ECG's periodic reports to the MDB Presidents. They will also reference this paper and report on their harmonization progress in their annual evaluation reviews reported to their Boards. In addition to this self-reporting, the Members agree that they will arrange for periodic independent cross-cutting assessments of the extent to which these (and

subsequent) good-practice standards are being applied in each MDB's evaluations and annual reporting, and report the findings to the MDB Presidents and Boards.

ENDNOTES

ⁱ Development Committee, Task Force on Multilateral Development Banks, "*Serving a Changing World – Report of the Task Force on Multilateral Development Banks*", March 15, 1996, Page 18, paragraphs 50 and 51..

ⁱⁱ IBRD. Committee on Development Effectiveness. *Comparative Analysis of MDB Completion Reporting and Performance Review*. Memorandum from the Director-General Operations Evaluation, paragraph 2. CODE96-77. November 12, 1996.

ⁱⁱⁱ Development Committee, *Report from the Multilateral Development Banks on Implementation of the Major Recommendations of the MDB Task Force Report*, March 7, 1998, page 4.

^{iv} ECG Mandate. Proposed Amendments to the ECG Founding Document and Mandate. For November 2001 ECG Meeting.

^v Chapter VIII (Evaluation Programming), Section 31 of the DAC Principles: "...agencies that have not already done so should elaborate guidelines and/or standards for the evaluation process. These should give guidance and define minimum requirements for the conduct of evaluation and for reporting".

^{vi} [From WGPSE 2001 paper:] The subject of standards and scope for integrated corporate performance reporting deserves separate pursuit by ECG members. The present paper addresses the quality and comparability of the basic building blocks for part of that comprehensive mosaic, covering the evaluation of the core operations portfolio. These principles are not intended to apply to synthesis reporting in special thematic studies of parts of the portfolio (e.g., sectors, countries), since the predominant purpose of such special studies is learning, rendering sampling representativeness for accountability less of an issue than in annual reporting.

^{vii} [From WGPSE 2001 paper:] This empirical deficiency does not necessarily present a harmonization issue for inter-MDB performance comparison at the corporate level, so long as the statistical significance limits of all portfolio/group performance data reported to the Board are made transparent. The significance limits, due to resource constraints (limited coverage standard) and/or to available small numbers in relevant sub-groups, however, comprise both a learning and accountability drawback relative to each individual MDB's results-based management needs.

^{viii} Development Committee, Task Force on Multilateral Development Banks, "*Serving a Changing World. Report of the Task Force on Multilateral Development Banks*", March 15, 1996, Page xi, paragraph 33.

^{ix} [From WGPSE 2001 paper:] Good-practice mandates include independence related provisions specifying who has decision authority for (1) hiring/terminating the CED head and staff; (2) CED head's appointment terms and reporting structure; (3) CED head's and staff's grading, performance reviews and pay increases; and (4) the CED's budget.

^x Eriksson, 2001, p. 20.

^{xi} [From WGPSE 2001 paper - abbreviated] There is a recognized tradeoff between delaying evaluation long enough to have a fully reliable basis for making updated projections for outcome and impact judgment (favors later evaluation), and studying results as early as possible for their lessons, relevance to evolving corporate priorities and accountability uses (favors earlier evaluation). Another consideration is that since projections tend to prove optimistic (albeit decreasingly so) over an operation's life, earlier evaluations will tend to reach more favorable impact judgments (weighted as they are by assumptions of better future compared to past results) than will fully ex post judgments of the same operation's results.

^{xii} The DCD/DAC/EV(2001)3 glossary provides the following definition of "impact": The totality of positive and negative, primary and secondary effects produced by a development intervention, directly or indirectly, intended or unintended. *Note:* Also interpreted as the longer-term or ultimate result attributable to a development intervention (RBM term); in this usage, it is contrasted with output and outcome which reflect more immediate results from the intervention.

^{xiii} For useful suggestions on ways to disseminate lessons learned, see the DAC.

^{xiv} [From WGPSE paper:] From paragraph 25 of *Review of the DAC Principles for Evaluation of Development Assistance – Final Report* dated February 1998: "On the matter of lessons learned from evaluations, it is clear that the users will rarely draw on such material unless required by agency leaders. The demand for the results of evaluations and the lessons they provide in an environment that promotes organizational learning is key. At the same time, *the supply of lessons and other knowledge that would benefit operations needs to be easily accessed in usable form.* The costs of searching out relevant material from evaluations, even when required, is a major disincentive to the lesson learning process. 'Just in time' practices in providing this material can facilitate use" (italics added). The availability of webpage-intranet dissemination and searchable databases makes accessing relevant lessons much easier than in the past. It does, however, require the development of an MDB-specific thesaurus of terms and coding of each lesson using it to enable thematic subject searches.

^{xv} [From WGPSE paper:] This integration of evaluation scope, measurement standards, findings and reporting with corporate- and unit-level portfolio reporting caters for the principle that "what gets measured, gets done", and that properly "what gets done, gets measured coherently and consistently". Without this integration, there will likely be a disconnect between predominant operational/career incentives, and application of evaluation lessons for getting better outcomes, since competing short-term results pressures (viz. annual new approval targets) will likely prevail over longer-term outcome concerns.

Guiding Framework Documents

Development Committee, Task Force on Multilateral Development Banks, *Serving a Changing World – Report of the Task Force on Multilateral Development Banks*, March 15, 1996

Development Committee. *Report from the Multilateral Development Banks on Implementation of the Major Recommendations of the MDB Task Force Report*, March 26, 1998

OECD. Development Assistance Committee. *Principles for Evaluation of Development Assistance*, Paris 1991

OECD. Development Assistance Committee. DAC Working Party on Aid Evaluation. *Review of the DAC Principles for Evaluation of Development Assistance*. Paris 1998

OECD. Development Assistance Committee. *DAC Criteria for Evaluating Development Assistance*, Update 21 November 2000

OECD. Development Assistance Committee. Working Party on Aid Evaluation. *Glossary of Evaluation and Results Based Management Terms*, April 24, 2001

Eriksson, John. *Comparative Analysis of MDB Completion Reporting and Performance Review*. Prepared for the MDB Evaluation Cooperation Group (ECG). Washington, DC, 1996

Eriksson, John. *Review of Good Practice and Processes for Evaluation of Public Sector Operations by MDBs*. Prepared for Working Group on Evaluation Criteria and Ratings for Public Evaluation (WGEC) of the MDB Evaluation Cooperation Group (ECG). Washington, DC, 2001

Forss, Kim. *The Theory and Practice of Analysing Institutional Development in Evaluation*. Prepared for the MDB Evaluation Coordination Group (ECG). Strangnas, Sweden, 2001

Wyss, Hans. *Harmonization of Evaluation Criteria*. Report on Five Workshops. Prepared for the MDB Evaluation Coordination Group (ECG). Washington, DC, 1999