ECG
Evaluation Cooperation Group

Big Book on Evaluation Good Practice Standards

November 2012

Contents

Acronyms and Abbreviations

3ie	International Initiative for Impact Evaluation
AfDB	African Development Bank
APPI	aggregate project performance indicator
AsDB	Asian Development Bank
BSTDB	Black Sea Trade and Development Bank
CEB	Council of Europe Development Bank
CED	Central Evaluation Department
C-GPS	core-good practice standard
CR	completion report
CSP	country strategy and program
CSPE	country strategy and program evaluation
EBRD	European Bank for Reconstruction and Development
ECG	Evaluation Cooperation Group
EIB	European Investment Bank
EP	evaluation principle
ERR	economic rate of return
ESMS	environmental and social management system
FRR	financial rate of return
GEF	Global Environment Facility
GPS	good practice standard
IADB	Inter-American Development Bank
IFAD	International Fund for Agricultural Development
IFI	international financial institution
IsDB	Islamic Development Bank
MDB	multilateral development bank
MDG	Millennium Development Goal
MIGA	Multilateral Investment Guarantee Agency
NONIE	The Network of Networks on Impact Evaluation
OECD–DAC	Organisation for Economic Co-operation and Development–Development Assistance Committee
O-GPS	optional-good practice standard
OP	operational practice
PBL	policy-based lending or policy-based loan
PER	performance evaluation report

ROE	return on equity
TA	technical assistance
UN	United Nations
WACC	weighted average cost of capital
WBG	World Bank Group
WGPUB	Public Sector Working Group of the Evaluation Cooperation Group
XASR	expanded annual supervision report
XASR-A	XASR-Assessment

Terms and Definitions

Term	Definition
Aggregate Project Performance Indicator (APPI)	A single measure of overall project performance constructed from ratings on the core evaluation criteria.
Benchmark	A standard that serves as a point of reference by which performance is measured.
Benchmarking	The process by which an IFI's evaluation framework, methodology, policies and practices are judged compliant with the ECG Good Practice Standards.
Borrower performance	The adequacy of the Borrower's assumption of ownership and responsibility during all project phases, including government, implementing agency, and project company performance in ensuring quality preparation and implementation, compliance with covenants and agreements, establishing the basis for sustainability, and fostering participation by the project's stakeholders.
Broad economic and social goals	Sector-wide and/or economy-wide goals that are not included in the project's statement of objectives but nevertheless are of interest in the evaluation.
Cancelled Guarantee	A cancelled guarantee is one that has been issued, and been active, and then is cancelled prior to its expiry date. In respect of Political Risk Insurance guarantees, the project company is not party to the cancellation, which is at the unilateral initiative of the guarantee-holder.
Cancelled Investment	An undisbursed, committed balance of an equity investment or loan, cancelled by mutual consent of the IFI and the project company .
Central Evaluation Department (CED)	The corporate unit charged with supporting the self-evaluation system for investment operations and reviewing its main products (e.g., completion reports (CRs), expanded annual supervision reports (XASRs), or country strategy completion reports), in addition to producing performance reports (PERs), Annual Reviews of the IFI's evaluation results and other independent evaluation studies, and performing related dissemination responsibilities.
Closed Investment	A disbursed investment that has been fully repaid, sold, or written off. Guarantees are considered closed when they have expired or been cancelled.
Company	Generally, the legal entity owning and implementing the project; in most cases the IFI's investment counterparty. For financial markets operations, the company is: (a) the financial intermediary in the case of credit lines, bank equity investments, leasing companies, etc.; or (b) the fund management company (as distinct from the normally separately owned investment fund itself) in the case of funds. In the case of political risk insurance (PRIs), MIGA's only counterparty is the financier or investor and not the project company.

Term	Definition
Completion Report (CR)	A record of a public sector operation's performance at the end of its implementation phase, undertaken as a self-evaluation by an IFI operations unit.
Completion Report validation	A review of Completion Report findings by the Central Evaluation Department, normally as a desk study.
Core Evaluation Criteria	The principal criteria that form the basis for evaluating project performance. For evaluations of investment/TA loans, the core criteria are Relevance, Effectiveness, Efficiency, and Sustainability. For evaluations of policy-based lending/loans (PBLs), the core criteria are Relevance, Effectiveness, and Sustainability.
Corporate goals	Areas of special focus of the IFI, such as poverty reduction, rural poverty reduction, transition to open market economies, implement EU policy objectives, European social cohesion, etc.
Cost-benefit analysis (CBA)	A quantitative analysis performed to establish whether the present value of benefits of a given project exceeds the present value of costs.
Cost-effectiveness analysis (CEA)	A quantitative analysis that compares the relative costs and outcomes of two or more courses of action. Cost-effectiveness analysis can be used to show whether the outcomes were delivered at least cost compared to alternative ways of achieving the same outcomes.
Direct Evaluation	Evaluations undertaken directly by the CED (as opposed to indirectly by IFIs), such as Performance Evaluation Reports (PERs). (Otherwise referred to as independent evaluation, this term is used in GPS on Private Sector Operations to acknowledge the different terminologies used in the different IFIs.)
Disclosure	The systematic distribution of evaluation findings through various media (including the CED's website) to the public at large, normally subject to certain restrictions specified in a Board-approved disclosure policy.
Dissemination	The systematic distribution of evaluation findings through various media within the IFI, generally without restriction as to contents, with the aim of promoting awareness and reinforcement of corporate objectives, success standards, accountability, and use of lessons for improved results.
Dropped Investment	A proposed investment in the private sector approved by the IFI's Board of Directors that has failed to become a signed agreement.
Early Operating Maturity	Referring to the point in time at which an investment operation in the private sector is ready for evaluation.
Economic Rate of Return (ERR) or Economic Internal Rate of Return (EIRR)	The internal rate of return of the time series of the project's economic costs and benefits. The ERR is an absolute measure of project benefits in relation to costs, not a measure of efficiency *per se*.

Term	Definition
Economic Return on Invested Capital (EROIC)	The internal rate of return on the economic costs and benefits on a before-after, rather than a with-without, basis but taking into consideration also other material, documented costs and benefits to customers, employees, government, suppliers, competitors, local residents, etc.
Effectiveness	The extent to which the project achieved (or is expected to achieve) its stated objectives, taking into account their relative importance.
Efficiency	The extent to which the project has converted its resources economically into results.
Evaluability	The extent to which the value generated or the expected results of a project are verifiable in a reliable and credible fashion. A measure of how well a proposed strategy or program sets out criteria and metrics to be used in its subsequent evaluation.
Evaluation Principle (EP)	A key unit of the Good Practice Standards, which together form the framework that IFIs should follow if they are to be deemed to have a satisfactory evaluation system. Each Evaluation Principle is defined by a set of elements.
Expanded Annual Supervision Report - Assessment (XASR-A)	The CED's instrument for conveying the findings of its desk review of each XASR. Its scope includes a judgment of the XASR's quality (responsiveness to scope guidelines, research depth, application of guideline-prescribed standards, and objectivity), appropriateness of assigned performance ratings, appropriateness and completeness of identified lessons, and issues for discussion in a Management-led review meeting (if the CED recommends the XASR for such a review).
Expanded Annual Supervision Report (XASR)	Otherwise referred to as an Indirect Evaluation. A standard, one-time supervision report undertaken once the project reaches early operating maturity with an attached evaluative addendum (expanded refers to the evaluative addendum), prepared on investments selected for evaluation by the CED. The addendum is a concise document, executed in a standard template according to a set of instructions prepared by the CED. It features analysis of specified performance dimensions with rated indicators and lessons learned. CED-verified XASR findings and performance ratings form the core of the CED's annual synthesis report (the Annual Review).
Financial Rate of Return (FRR or Financial Internal Rate of Return (FIRR))	The internal rate of return of a time series of cash flows describing the project's or company's financial investments and returns.
Financial Rate of Return (FRR)	The internal rate of return of a series of cashflows describing the project's financial investments and returns over time.
Fund Weighted Average Cost of Capital (FWACC)	The cost of capital for a private equity or listed fund, estimated by calculating the average cost of debt based on the country composition of the fund, and then levying a premium for the combined equity instrument and project risk.
Gross Profit Contribution	The gross revenues generated for an IFI by an investment after deducting financing costs and loss provisions but before deducting administrative costs.

Term	Definition
IFI Performance	The quality of services provided by the IFI during all project phases, including the IFI's performance in ensuring project quality at entry, satisfactory implementation, and future operation.
Impact	Higher level of outcomes of projects or strategy. [GPS on Evaluation of Public Sector Operations avoids the use of the term because of the multiple and conflicting meanings of the term, instead the GPS calls for project objectives to focus on outcomes for which the project can reasonably be held accountable, avoiding objectives beyond the purview of the project. To the extent that higher-level social and economic objectives and corporate goals are included, they should be targeted at segments of the population that can reasonably be expected to be affected by the project, directly or indirectly. – GPS Annex on Impact and Impact.
Impact evaluation	An impact evaluation quantifies the net change in outcomes that can be attributed to a specific project or program, usually by the construction of a plausible counterfactual. (See the annexed note on Impact.)
Independent Evaluation	Otherwise referred to as Direct Evaluation. Evaluations undertaken by the IFI's CED, including Performance Evaluation Reports (PERs), XASR Assessments (XASR-As – assessments of expanded annual supervision reports), special studies and Annual Reviews, the latter based largely or in part upon the findings of CED-verified XASRs, PERs and relevant portfolio performance data.
Indirect Evaluation	Evaluations undertaken by staff of the IFI (as opposed to directly by the CED), such as XASRs. Indirect Evaluations are accompanied by independent verification of findings by the CED~~,~~ such as in XASR-As. (Otherwise referred to as Self-Evaluation; term is used in GPS Private Sector to acknowledge the different terminology in the different IFIs.).
International Financial Institutions (IFIs)	Collectively refers to the Bretton Woods institutions, regional and bilateral development banks and financial institutions that are members of the ECG which provides financing and advisory services for projects and programs in member countries. It does not include internationally-operating commercial banks.
Investment	The IFI's financial instrument specific to the operation being evaluated. Investments mainly consist of loans, loan guarantees, quasi-equity and equity investments. In the case of MIGA, the investment refers to MIGA's PRI instrument (see below).
MIGA Political Risk Insurance (PRI)	MIGA's PRI guarantees typically involve a bilateral contractual relationship between the insurer (MIGA) and the guarantee holder and do not involve the project company. There is therefore no project agreement, and MIGA does not have a relationship with, or recourse to, the project company. Other forms of guarantee (e.g., financial or partial risk guarantees), by contrast, involve a three-way contractual relationship between the guarantor, the project company and the guarantee holder. There is both a project agreement and a guarantee agreement.
Net present value	The sum of the present values of the time series of project costs and benefits.

Term	Definition
Net Profit Contribution	The net profit earned by an IFI on an investment in the private sector after deducting financing costs, loss provisions and administrative costs.
Operation	The IFI's objectives, activities and results in making and administering its investment as part of the overall financing / support of the borrower's project.
Operational Practice (OP)	Operational Practices describe the policies and procedures that the CED / IFI would typically need to adopt in order to be deemed compliant with the respective Evaluation Principle and its elements.
Outcome	Refers to levels in the results chain beyond "outputs".
Output	The tangible goods and services that the project activities produce, generally under the direct control of the implementing agency.
Performance Evaluation Report (PER)	A CED report prepared from an independent/direct evaluation of an individual investment operation and normally includes field work
Project (Public Sector)	A public sector investment, technical assistance activities, or program that is supported by an IFI loan. Note that under this definition, a PBL-supported program is called a "project".
Project (Private Sector)	Generally, the company's capital project or program and related business activity that have been partially financed or otherwise supported by the IFI's investment selected for evaluation. In financial markets operations, the project generally refers to the financial intermediary's lending or investment program that is partially financed or otherwise supported by the IFI.
Relevance	Consistency of the development interventions objectives with beneficiary needs, the country's development or policy priorities and strategy, and the IFI's assistance strategy and corporate goals; and the adequacy and coherence of the project's components (design) to achieve those objectives.
Results chain	A model that sets out the sequence of inputs, activities, and outputs that are expected to lead to the project's intended outcomes. Describes the causal relationships, indicators, and the assumptions or risks that may influence project success and failure. Alternatively called a "results framework", "causal chain", or "logical framework (log frame)".
Return on Invested Capital (ROIC)	The internal rate of return on invested capital in real terms i.e., the financial rate of return (FRR) on the costs and benefits to the company as a whole on a before-after, rather than a with-without, basis.
Self-Evaluation	Evaluation undertaken by IFI operations departments. [Otherwise referred to as indirect evaluation in the private sector.]
Sub-Project	Refers to the project(s) undertaken by sub-borrowers under an IFI credit line (or guarantee of such) to a Financial Intermediary, or by investee companies within a Fund subscribed to (or guaranteed) by the IFI.

Term	Definition
Sustainability	The likelihood of continued long-term benefits, and the resilience to risk of net benefit flows over time.
Theory-based evaluation	An analysis that establishes a plausible association between the various links in the project's results chain, using quantitative and qualitative evidence as well as evidence from other evaluations and academic literature.
Weighted Average Cost of Capital (WACC)	The weighted average after-tax cost to the company of the yields it must provide on its borrowings and the equity investors' minimally acceptable returns, all adjusted for inflation.
Counterfactual	An attempt to gauge what would have occurred in the absence of the intervention with what has occurred with the intervention implemented
Ex-post Evaluation	Evaluation of a development intervention after it has been completed. (Note: It may be undertaken directly after or long after completion. The intention is to identify the factors of success or failure, to assess the sustainability of results and impacts, and to draw conclusions that may inform other interventions.)
Finding	Refers to evidences collected in an evaluation.

Source: DAC Glossary of Terms, and ECG-approved GPS on (i) Independence of IFIs' CED, June 2010, (ii) Evaluation of Public Sector, February 2012, (iii) Evaluation of Private Sector Operations, November 2011, and (iv) Country Strategy and Program Evaluation, 2008.

I. Background and Context

Contents

Multilateral Development Banks' Evaluation Harmonization

1. **The Harmonization Challenge.** In March 1996, the Development Committee Task Force on Multilateral Development Banks (MDBs)[1] issued a report entitled *Serving a Changing World,* which called for harmonization of evaluation methodologies, performance indicators, and criteria by 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…MDB evaluation units…[should] be charged with elaborating common evaluation standards, including performance indicators; exchange experience with evaluation techniques [and] share results; and become the repository of best evaluation practices. The immediate task would be to develop, within a specified time period, methodology and criteria for assessing and rating the MDB's operational performance and development effectiveness."[2]

2. As a matter of priority, the Development Committee invited the five major MDBs to act on the Report's relevant recommendations to strengthen their policies and practices and to advise the Committee in about two years, on the progress achieved in implementing the recommendations.[3]

3. **MDBs' Response.** In response to the task force's recommendations, the Evaluation Cooperation Group (ECG) was formed in 1996.[4] The ECG mandate embodies and endorses MDB evaluation harmonization and provides that:

> "The ECG (i) works to strengthen cooperation among evaluators and (ii) seeks to harmonize evaluation methodology in its member institutions, so as to enable improved comparability of evaluation results while taking into account the differing circumstances of each institution. Harmonization in the ECG includes increased information sharing and improved understanding of commonalities and differences in evaluation policies, procedures, methods and practices and is not interpreted by members as 'standardization of evaluation policies and

---

[1] The Development Committee Task Force on MDBs was established in 1994 to undertake, for the first time, an assessment of the five major MDBs - the African Development Bank (AfDB), the Asian Development Bank (AsDB), the European Bank for Reconstruction and Development (EBRD), the Inter-American Development Bank (IADB), and the World Bank Group (WBG). (Source: Development Committee Task Force on MDBs. 1996. *Serving a Changing World*. Washington, DC, Foreword.)

[2] Development Committee Task Force on MDBs. 1996. *Serving a Changing World*. Washington, DC, p. 18.

[3] Development Committee Task Force on MDBs. 1996. *Serving a Changing World*. Washington, DC, Foreword.

[4] ECG was created by the 5 MDBs in October 1995 and was officially convened in February 1996. (Source: ECG Mandate in ECG website.)

practices'."[5]

4. The ECG consisted initially of the heads of the evaluation units of the five MDBs referred to in the task force's report: the African Development Bank (AfDB), Asian Development Bank (AsDB), European Bank for Reconstruction and Development (EBRD), Inter-American Development Bank (IADB), and World Bank Group (WBG). A number of International Financial Institutions (IFIs) have joined the ECG since then and by 2011, total membership has reached nine (9) comprising of the heads of evaluation units of the five founding MDBs, the European Investment Bank (EIB), the International Monetary Fund (IMF), Islamic Development Bank (IsDB), and the International Fund for Agricultural Development (IFAD).[6] In 2012, Black Sea Trade and Development Bank (BSTDB) and the Council of Europe Development Bank (CEB) have been accepted as full members and membership processing is underway.[7] Meanwhile, ECG has three observers who are the heads of evaluation units of the United Nations (UN), Development Assistance Committee of the Organisation for Economic Co-operation and Development (OECD-DAC), and the Global Environment Facility (GEF).[8]

5. The ECG Chairperson is selected annually on a rotational basis from among the members, and a co-chairperson is likewise selected, if deemed desirable. ECG meetings are held twice a year where the ECG Chairperson proposes agendas that have benefited from consultation with ECG members. In executing its activities, the ECG may also convene working groups and technical groups, as needed, with specific terms of reference. It can also organize workshops on specific themes where participants from outside ECG members can be invited.

6. **Evaluation Harmonization.** In March 1998, the original five ECG members reported to the Development Committee on the implementation of the recommendations of the Task Force:

> "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, nonlending services, and evaluation of private sector operations."[9]

7. In almost one-and-half decades after its first Report to the Development Committee, the ECG has remained focused on its commitments and continues to pursue harmonization and improvements in evaluation process. ECG milestone accomplishments include the support to evaluation capacity development, conduct of joint evaluations and joint thematic workshops on major and timely topics, assessment of evaluability of operations of selected members, and strengthening its oversight function over member MDBs. The ECG has also established collaboration mechanisms with other evaluation networks by sitting as an observer in: (i) the OECD-DAC Network on Development Evaluation for the purpose of improving the effectiveness

---

5 ECG. 2003. *Amended ECG Mandate*.
6 EIB joined ECG in 1998, IMF in 2001, and IsDB and IFAD in 2010.
7 Minutes of ECG Plenary Meeting in March 2012.
8 The UN is represented by the Director of UNDP's Evaluation Unit and/or the Chair of the United Nations Evaluation Group, and OECD-DAC is represented by the Head of the Secretariat and /or the chair of the Network on Development Evaluation. (Source: ECG membership in ECG website.)
9 Development Committee Task Force on MDBs. 1998. *Implementation of the Major Recommendations of the MDB Task Force Report*. p. 4.

of international development assistance, and (ii) meetings of the United Nations Evaluation Group (UNEG).[10] But more importantly, ECG has developed and implemented the Good Practice Standards (GPS) for four categories of MDB evaluations covering governance and independence of evaluation function, public and private sector operations, and country strategy and program (CSP).[11] The goal of documenting these standards is to harmonize evaluation practice among ECG members, not to evaluate functions.[12] The GPS are therefore consistent with ECG's mandate on harmonization and on improving understanding of evaluation practices. Derived from the evaluation principles of the OECD–DAC, these GPS were built on good evaluation practices, and were designed to be consistent with the MDBs' operational policies.

8. **Origin of Big Book**. In view of the increased interaction between the various working groups on the GPSs as well as to achieve greater "cross fertilization", the ECG members suggested in 2011to combine the four GPS into a single document called the GPS "Big Book". The "Big Book" contains the latest version of the approved GPS (Please see footnote#11) and will be presented for review in ECG's Fall Meeting in 2012. The "Big Book" shall be considered a living document that will be revised from time to time.

9. The Chapters of the "Big Book" attempt to organize the evaluation principles by type, i.e., general and specific, as well as to address overlaps noted in the GPSs and to resolve differences in terminologies, if any, without losing the original intent of the four GPS reports taken individually. Chapter II deals with the general evaluation principles and lays down the GPS on independence of MDBs' central evaluation department (CED) essential to ensuring MDB accountability and feedback mechanism from performance evaluation. Chapters III and VI present the specific evaluation principles. Chapters III to V cover the independent evaluation principles and standards by evaluation category, namely public and private sector operations, and CSPs, the results from which should be of value to further increase the effectiveness of MDB development interventions. Chapter VI provides key standards for MDB self evaluation aimed to help improve the evaluability of projects and investments and quality of completion reports (CRs), and to establish effective linkages with independent evaluation. The respective evaluation principles are grouped by standards, each of which is composed of a set of elements. The evaluation principles are also supported by the corresponding set of operational principles. In the case of GPS on evaluation at the country level, the operational principles are expressed in terms of Core-GPS (C-GPS) and Optional-GPS (O-GPS).[13]

---

[10] Based on traditional links to UNEG, some ECG members/observers participate in UNEG meetings.

[11] The approved GPS are on (i) Independence of IFIs' CED, June 2010, (ii) Evaluation of Public Sector, February 2012, (iii) Evaluation of Private Sector Operations, November 2011, and (iv) Country Strategy and Program Evaluation, 2008.

[12] ECG has developed a separate set of standards for the assessment of the evaluation functions of IFIs in ECG's *Review of Framework for the Evaluation Function in Multilateral Development Banks*, 2009.

[13] A core GPS is defined as one that establishes the key principles for CSPEs and is necessary to permit comparability of evaluation results, to the extent possible, among MDBs. While the core GPSs listed in this paper are currently in practice to some extent in all members, institutional differences may affect the pace at which harmonization can be achieved. An optional GPS is defined as one that is not strictly needed for comparability but is nonetheless designed to help improve accountability and learning within each institution.

II. Independence of International Financial Institutions' Central Evaluation Departments

Contents

Background

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."[14]

Implicit in this statement is the notion that the Central Evaluation Department (CED) of an International Financial Institution (IFI) should exercise independence in conducting and reporting on its evaluation work.

11. **Rationale for CED Independence.** The key function of the CED is to provide independent evaluation of the projects, programs, policies, and activities of the parent IFI. While independence is essential for the IFI to maximize the benefits from its evaluation system, the *raison d'être* of independence is not for its own sake but to provide for impartial, credible evaluation as a means to help improve the performance of an organization. Four principles should be borne in mind when considering independence:[15]

(i) The rationale for independence in its various dimensions is to provide for, and to protect, the impartiality of evaluations and to ensure that the ability of the evaluators to provide credible reports and advice is not compromised.

(ii) Independence does not mean isolation, as both operations and evaluation activities are enriched through cross-fertilization of knowledge and experience, and evaluators can help to introduce good practice and innovations by being aware of relevant developments outside the IFI. This has implications for evaluation work processes and issues such as the rotation of CED staff to and from other parts of the IFI and the mix of CED staff with experience inside and outside the IFI.

(iii) Independence does not imply any particular approach to evaluation. In particular, independence does not mean that evaluators should focus more on accountability than on learning.

(iv) Independence does not mean lack of accountability and responsibility or that CED is exempt from the same degree of transparency as any other part of the IFI. The mechanisms used to ensure adequate levels of accountability for the evaluators may be somewhat different from, and independent of, the mechanisms for the parts of the organization reporting to management.

12. **Establishing Independent Evaluation Offices.** Since 1996, the concept of evaluation independence has been articulated further. MDBs have formed their respective independent evaluation offices and the subsequent removal of oversight function of MDB management over evaluation offices. In 2003 the Evaluation Cooperation Group (ECG) adopted a framework for assessing evaluation independence along four dimensions: organizational, behavioral,

---

[14] Development Committee, Task Force on Multilateral Development Banks, *Serving a Changing World: Report from the Task Force on Multilateral Development Banks*, March 15, 1996, page xi, paragraph 33.

[15] Adapted from Evaluation Cooperation Group, *Peer Review of IFAD's Office of Evaluation and Evaluation Function: Final Report*, February 15, 2010, p. 5.

avoidance of conflict of interest, and protection from outside interference (Annex 1). ECG continues to advocate the importance of independent evaluation offices being accountable for quality of evaluation functions. The process of peer review of MDBs' evaluation systems has also been established to serve the purpose of providing judgment on the quality of evaluation functions.[16]

13. Notwithstanding these progress, in 2007, a finding from the study of the Task Force on GPS on CSPE emerged "that country evaluators are not completely independent of the country program as a total non-involvement would mean inability to contribute meaningfully to the evaluation."[17]

14. **Purpose of Good Practice Standards** In anticipation of any issue that may arise from the above finding, members agreed that ECG can play a role as a body of observers in strengthening independence of evaluation offices. Members also recommended strengthening the perception of independence of evaluation offices among outsiders. Further to these, members proposed the development of good practice standards (GPS) on governance and independence of evaluation that should be common to all evaluation activities. These GPS are intended to guide IFI policies and procedures for ensuring the independence of the CED, and to be used in self-assessments, peer reviews, and governing Board reviews of the CED.

15. **Formulation Process**. In 2008, the Working Group on GPS on Public Sector was tasked to develop the GPS on independence of evaluation function. Following the presentation of the draft GPS in 2009, the WB Group (WBG) was designated to take the lead in revising the GPS to reflect the "(i) importance of independence in evaluation within the architecture of the functioning of the MDBs, (ii) broader concept of independence from management, differing operations based on differing governance rules, and varying conception of independence; (iii) current thinking on issues in independence; and (iv) greater clarity on key constituents of independence and trade-offs between independence, relevance, and use of evaluation findings."[18] The GPS on Independence of IFIs' CED was approved in 2010.

16. **Organization.** The GPS on CED Independence are presented in terms of evaluation principles (EPs) which consist of 5 standards and 26 elements. A summary is presented below.

---

[16] Peer Review can be described as the systematic examination and assessment of the performance of one institution by another institution that has, in principle, a similar mandate and set of basic principles, policies and way of working. The ultimate goal is to help the reviewed institution improve its policy making, adopt best practices, and comply with established standards and principles. The examination is conducted on a non-adversarial basis, and it relies heavily on mutual trust among the institutions involved in the review, as well as their shared confidence in the process.
Since all MDBs have independent evaluation offices, a Peer Review of the evaluation function in a given institution could theoretically be carried out by heads of evaluation in the other institutions. The credibility of such a review would be substantially enhanced, however, by including in the Peer Review process external peers from the development evaluation community. Therefore, Peer Reviews must be conducted by a Peer Review Panel, composed of heads of evaluation units from the participating MDBs and of respected independent evaluators in the international evaluation community/industry. To secure its independence, the majority of the Peer Review Panel should be external, independent evaluators. It is important that the Peer Review Panel can benefit from the experience of the Evaluation Heads of the MDBs. However, the independence and credibility of the system is best served with a minority of MDB peers in the Panel. With these elements in place, Peer Review is designed to create, through this reciprocal evaluation process, a system of mutual accountability and an independent external review mechanism that serves Boards of Directors and Shareholders of the MDBs.
(*Source: A Proposal for an Evaluation Peer Review System for Multilateral Development Banks in Documents Accompanying ECG Minutes of 2007 Meeting in ECG Website.*)

[17] Task Force on GPS on CSPE presentation to ECG in 2007 Fall Meeting.

[18] Minutes of ECG 2009 Spring Meeting.

The EPs, on the other hand, are supported by a total of 18 operational practices (OPs) which describe the policies and procedures that should be adopted to be compliant with the EPs. The detailed EPs and OPs are presented in the next section.

Summary of Standards and Elements of EPs on CED Independence

Standards	Elements	No. of OPs	Page
1. Governance and Independence of the CED	A. CED Mandate B. Mandate Coverage C. Structural Independence D. Oversight E. Consultative Framework F. Scope of Responsibility G. Rights of Access	4	20
2. Independent Leadership of CED	A. Appointment B. Contract Renewal C. Termination D. Authority & Remuneration E. Performance Assessment	4	21
3. Independence of CED Staff	A. Selection B. Skills C. Opportunities D. Conflict of Interest	3	22
4. The CED Work Program and Budget	A. Work Program B. Determination of Budget C. Adequacy of Budget D. Accountability and Transparency	2	22
5. Independent Reporting and Disclosure by CED	A. Reporting Line B. Primary Stakeholder C. Other Stakeholders D. Recommendations of CED E. Disclosure Policy F. Dissemination of Products	5	23
Total no. of standards: **5**	**Total no. of elements**: **26**	**Total no. of OPs: 18**	

Source: GPS on Evaluation of Private Sector Operations.

GPS on CED Independence

Evaluation Principle (Standards and Elements)	Standard Operational Practices	Element Link	Notes
1. Governance and Independence of the CED: A. *CED Mandate:* The CED's mandate is specifically approved through a Board resolution. B. *Mandate Coverage:* The CED mandate establishes its mission, scope of responsibilities and independence. C. *Structural Independence:* The CED's governance, organization and resources make it independent from the IFI's Management. D. *Oversight:* The CED mandate establishes that the Board oversees the CED's work. E. *Consultative Framework:* The CED has full autonomy, but works in consultation with the IFI's operational departments. F. *Scope of Responsibility:* The CED reports on all determinants of the IFI's operational results. G. *Rights of Access:* The CED has unrestricted access to the IFI's records, staff and counterparties.	1.1 The CED operates according to a Board-approved mandate that specifies its mission, scope of responsibilities, reporting structure and key operating principles. The governance arrangements are designed to ensure the CED's independence, its relevance to the IFI's mission, and the delivery of its corporate accountability and learning value-added. To ensure organizational independence, the CED does not report to IFI Management, is located organizationally outside the line and staff management function, and is independent of the IFI's operational, policy, and strategy departments and related decision-making. Where the IFI has a monitoring and evaluation policy, it should make specific provision for the organizational and behavioural independence of the CED and its protection from interference by Management. The policy should reflect the Board-approved mandate of the CED.	*CED Mandate* *Mandate Coverage* *Structural Independence*	Note: In respect of IDB's Office of Evaluation & Oversight, reference to the Board is to the IDB Board. While MIF and IIC contract for the services of the Office of Evaluation & Oversight, their Boards have no jurisdiction over its mandate or operations.
	1.2 The CED's work is overseen by the Governing Board, a designated committee of the Board, or an independent governing body; for purposes of these GPS such bodies are referred to as the governing Board.	*Oversight*	
	1.3 The CED operates with full autonomy but in close consultation with the IFI's other departments to ensure, as far as possible (subject to the primacy of sound evaluative principles and practices), both: (a) coherence of corporate standards among operations, portfolio and strategy analysis, and evaluation; and (b) good prospects for corporate ownership of the CED's findings and recommendations for improvement. To help ensure that the independent evaluation work responds to the IFI's needs for information to guide policy and operational decisions, the CED's annual work program – the principal determinant of the CED's budget – is widely discussed during preparation with the Board, managers and IFI staff.	*Consultative Framework*	
	1.4 The CED's role is to ensure the relevance, quality and impartiality of the products of the IFI's evaluation system. Under its mandate the CED scope of responsibility extends, without restriction, to all the determinants of the IFI's operational results.	*Scope of Responsibility*	In some private-sector activities, the mandate may allow for restrictions on access to clients and projects where an evaluator's direct contact could prejudice the IFI's financial

Evaluation Principle (Standards and Elements)	Standard Operational Practices	Element Link	Notes
	The CED has unrestricted access to: (a) the IFI's staff and records in the context of an actual evaluation; and (b) co-financiers and recipients of the IFI's loans, grants, and equity investments. The CED also has access to project, program, and activity sites, as well as other stakeholders.	*Rights of Access*	interests or materially increase the risk of litigation. Should client access be restricted in such jeopardy cases, the number of such cases should be reported in the CED's annual report and/or annual evaluation review.
2. Independent leadership of the CED: A. *Appointment:* The CED's head is selected and appointed by the Board or representative thereof. B. *Contract Renewal:* Renewal of the CED head's contract can only be authorised by the Board. C. *Termination:* Only the Board is able to terminate the contract of the CED's head on the basis of predefined policy. D. *Authority & Remuneration:* The CED's head holds grade-rank and remuneration comparable to the level immediately below Vice-President or equivalent. E. *Performance Assessment:* The performance of the CED's head is assessed by the Board.	2.1 The head of the CED is appointed by the governing Board or the Board Committee that oversees the evaluation function, through procedures approved by that body. These procedures may include a search committee on which IFI Management is represented, as well as the use of outside search firms or consultants, provided that the governing Board retains final decision-making authority. When the IFI does not have a resident Board, the minimum requirement regarding the appointment of the head of the CED is the presence of at least one Board representative in the selection process. The CED head's appointment normally is for a fixed term, but may include an option for renewal at the end of that term. If renewal of the CED's appointment is allowed, the governing Board has the authority to approve such a renewal.	*Appointment* *Contract Renewal*	
	2.2 Only the Board may terminate the head of the CED; any such termination should be for cause, based on performance or conduct grounds. A policy on termination should be in place. To preserve independence, upon termination of service as the CED head, the individual is not eligible for staff positions within the IFI.	*Termination*	
	2.3 The head of the CED holds a grade-rank equal at minimum to the level immediately below Vice-President or equivalent, with commensurate compensation.	*Authority & Remuneration*	
	2.4 The CED head's performance is assessed by the governing Board or an individual or body designated by it for this purpose. To preserve independence, IFI Management, including the President, may provide inputs into this process by way of feedback, but is not the assessor.	*Performance Assessment*	

Evaluation Principle (Standards and Elements)	Standard Operational Practices	Element Link	Notes
3. Independence of the CED's staff: A. *Selection:* The CED's staff are appointed by the CED's head or designee. B. *Skills:* The CED's staff should have adequate skills to conduct evaluations. C. *Opportunities:* Staff should not be career disadvantaged by having worked in the CED. D. *D. Conflict of Interest:* The CED ensures that its staff have no conflict of interest in their evaluation work.	3.1 The staff of the CED are selected by the CED's head or his/her designee, in accordance with overall personnel policies of the IFI. Such staff should have or be required to acquire specific evaluation skills; the CED should provide training needed to meet these requirements.	*Selection* *Skills*	For example, the skills required by evaluation staff can be defined using a competency framework, which can also serve as a guide for career progression within the CED.
	3.2 The CED's staff should not be disadvantaged because of the judgments and findings they report, and policies should be in place to ensure against such disadvantage. These should include policies that permit (but not necessarily require) the use of separate processes for assessing the CED's staff for changes in compensation, promotions, and job tenure, and for handling human resource issues. Such processes may be parallel to those for other staff of the IFI, but should protect the CED's staff from potential career limitations for findings and recommendations in their evaluations. Unlike the CED's head, CED staff may be permitted to rotate out of evaluation into other IFI units, subject to the conflict of interest limitations.	*Opportunities*	
	3.3 The CED has policies and procedures to ensure against conflicts of interest involving CED staff. Staff are prohibited from evaluating projects, programs, or other activities for which they previously held responsibility. The CED also has a policy regarding movement of evaluation staff into other IFI units to ensure that they are not subject to conflicts of interest while seeking or being sought for such positions.	*Conflict of Interest*	
4. The CED's work program and budget: A. *Work Program:* The CED consults on its work priorities, but determines its work program independently of Management. B. *Determination of Budget:* The CED's budget is approved by the Board. C. *Adequacy of Budget:* The CED's budget is commensurate with its work program.	4.1 The CED develops its own work program, which may be endorsed by the governing Board. The CED may consult with IFI staff and Management, as well as the Board and outside organizations or experts, in constructing its work program, but Management does not exercise direct control over the work program.	*Work Program*	
	4.2 The CED's budget is approved by the governing Board, commensurate with the work program. The CED may be required to follow IFI processes of general applicability in presenting its budget and in accounting for the use of budget resources. However, Management does not have approval authority over the CED's budget. The CED is subject to the institutional auditing requirements of the IFI. However, audits must be conducted by an auditor independent of Management, and approved by the relevant governing body or bodies.	*Determination of Budget / Adequacy of Budget* *Accountability & Transparency*	

Evaluation Principle (Standards and Elements)	Standard Operational Practices	Element Link	Notes
D. *Accountability and Transparency:* The CED is accountable for its application of financial resources.			
5. Independent reporting and disclosure by the CED: A. *Reporting Line:* The CED transmits its products to the Board, without Management clearance or Management-imposed restrictions on content. B. *Primary Stakeholder:* The CED's primary stakeholder is the Board.[19] C. *Other Stakeholders:* The CED is also guided by the interests of other relevant internal and external stakeholders.[7] D. *Recommendations:* The CED monitors and reports on the implementation of CED recommendations by Management. E. *Disclosure:* The CED's disclosure policy is explicit, and consistent with the IFI's general disclosure policy. F. *Dissemination:* The CED employs an appropriate range of dissemination activities for its disclosed products.	5.1 The CED transmits evaluation products to the governing Board, normally after review and comment by Management, but without any Management clearance or Management-imposed restrictions on the scope and content of the products.	*Reporting Line*	
	5.2 The CED's major stakeholder is the governing Board to which it reports. The Board is responsible for ensuring the efficient use of resources and achieving results on the ground with sustainable development impact.	*Primary Stakeholder*	
	5.3 The CED also serves a wide range of internal and external stakeholders. Major internal stakeholders may include, but are not limited to: - IFI Management, which is responsible for acting on and following up evaluations, and for how evaluation findings might influence the IFI's future directions; - operations staff concerned with the feedback of evaluation lessons and findings, and how those might affect future operations; and - other IFI staff concerned with knowledge management, dissemination of evaluation findings, lessons and recommendations, and evaluation capacity development. Major external stakeholders may include, but are not limited to: - governments, executing agencies, and institutions responsible for implementing IFI-supported projects in borrowing countries; - beneficiaries and targeted populations directly affected by IFI support; - co-financiers and other partner institutions, including NGOs, civil society organizations, development research centres, and evaluation networks that are engaged in CED-financed operations; and - multilateral and bilateral institutions concerned with harmonizing evaluation methods and practices, and other development partners with whom the CED may undertake joint evaluations of programs, projects, policies and strategies, disseminate best practices, and organize evaluation seminars and workshops.	*Other Stakeholders*	

[19] Adapted from Asian Development Bank. 2008. Review of the Independence and Effectiveness of the Operations Evaluation Department. Manila. pp. 12-13.

Evaluation Principle (Standards and Elements)	Standard Operational Practices	Element Link	Notes
	5.4 Management has responsibility for implementing CED recommendations. However, the CED is responsible for a system to monitor and report to the governing Board Management's record of adoption of and response to recommendations, including its success in remedying any problems found in evaluations.	*Recommendations*	
	5.5 Disclosure of evaluation findings is an important component of IFI accountability to stakeholders, and of behavioural independence on the part of the CED. Therefore, written reports and other evaluation products are disclosed in accordance with the CED's Board-approved disclosure policy. Such a policy should be explicit, consistent with the IFI's general disclosure policy, and cover all evaluation products. The CED head may determine the appropriate types and level of external activities to promote the dissemination of disclosed evaluation reports and other evaluation products, within the limitations of the applicable disclosure policy and without Management interference.	*Disclosure* *Dissemination*	To protect client company confidentiality, promote the candour needed for effective corporate learning, and reduce risk to the IFI's credit rating that partial release of investment portfolio data (and related standards and benchmarks) might entail, the IFI may decide not to disclose individual evaluation reports or the full text of the CED's annual review for private sector operations.

Source: GPS on Evaluation of Private Sector Operations, Fourth Edition, November 2011.

Annex II.1: Template for Assessing the Independence of Evaluation Organizations

Criterion	Aspects	Indicators
I. Organizational independence	The structure and role of evaluation unit	Whether the evaluation unit has a mandate statement that makes clear its scope of responsibility extends to all operations of the organization, and that its reporting line, staff, budget and functions are organizationally independent from the organization's operational, policy, and strategy departments and related decision-making
	The unit is accountable to, and reports evaluation results to, the head or deputy head of the organization or its governing Board	Whether there is a direct reporting relationship between the unit, and a) the Management, and/or b) Board or c) relevant Board Committee, of the institution
	The unit is located organizationally outside the staff or line management function of the program, activity or entity being evaluated	The unit's position in the organization relative to the program, activity or entity being evaluated
	The unit reports regularly to the larger organization's audit committee or other oversight body	Reporting relationship and frequency of reporting to the oversight body
	The unit is sufficiently removed from political pressures to be able to report findings without fear of repercussions	Extent to which the evaluation unit and its staff are not accountable to political authorities, and are insulated from participation in political activities
	Unit staffers are protected by a personnel system in which compensation, training, tenure and advancement are based on merit	Extent to which a merit system covering compensation, training, tenure and advancement is in place and enforced
	Unit has access to all needed information and information sources	Extent to which the evaluation unit has access to the organization's a) staff, records, and project sites; b) co-financiers and other partners, clients; and c) programs, activities, or entities it funds or sponsors
II. Behavioral Independence	Ability and willingness to issue strong, high quality, and uncompromising reports	Extent to which the evaluation unit: a) has issued high quality reports that invite public scrutiny (within appropriate safeguards to protect confidential or proprietary information and to mitigate institutional risk) of the lessons from the organization's programs and activities; b) proposes standards for performance that are in advance of those in current use by the organization; and c) critiques the outcomes of the organization's programs, activities and entities
	Ability to report candidly	Extent to which the organization's mandate provides that the evaluation unit transmits its reports to the Management/Board after review and comment by relevant corporate units but without management-imposed restrictions on their scope and comments
	Transparency in the reporting of evaluation findings	Extent to which the organization's disclosure rules permit the evaluation unit to report significant findings to concerned stakeholders, both internal and external (within appropriate safeguards to protect confidential or proprietary information and to mitigate institutional risk). Who determines evaluation unit's disclosure policy and procedures: Board, relevant committee, or management.

Criterion	Aspects	Indicators
	Self-selection of items for work program	Procedures for selection of work program items are chosen, through systematic or purposive means, by the evaluation organization; consultation on work program with Management and Board
	Protection of administrative budget, and other budget sources, for evaluation function	Line item of administrative budget for evaluation determined in accordance with a clear policy parameter, and preserved at an indicated level or proportion; access to additional sources of funding with only formal review of content of submissions
III. Protection from outside interference	Proper design and execution of an evaluation	Extent to which the evaluation unit is able to determine the design, scope, timing and conduct of evaluations without Management interference
	Evaluation study funding	Extent to which the evaluation unit is unimpeded by restrictions on funds or other resources that would adversely affect its ability to carry out its responsibilities
	Judgments made by the evaluators	Extent to which the evaluator's judgment as to the appropriate content of a report is not subject to overruling or influence by an external authority
	Evaluation unit head hiring/firing, term of office, performance review and compensation	Mandate or equivalent document specifies procedures for the a) hiring, firing, b) term of office, c) performance review, and d). compensation of the evaluation unit head that ensure independence from operational management
	Staff hiring, promotion or firing	Extent to which the evaluation unit has control over: a) staff hiring, b) promotion, pay increases, and c) firing, within a merit system
	Continued staff employment	Extent to which the evaluator's continued employment is based only on reasons related to job performance, competency or the need for evaluator services
IV. Avoidance of conflicts of interest	Official, professional, personal or financial relationships that might cause an evaluator to limit the extent of an inquiry, limit disclosure, or weaken or slant findings	Extent to which there are policies and procedures in place to identify evaluator relationships that might interfere with the independence of the evaluation; these policies and procedures are communicated to staff through training and other means; and they are enforced
	Preconceived ideas, prejudices or social/political biases that could affect evaluation findings	Extent to which policies and procedures are in place and enforced that require evaluators: a) to assess and report personal prejudices or biases that could imperil their ability to bring objectivity to the evaluation; b) and to which stakeholders are consulted as part of the evaluation process to ensure against evaluator bias
	Current or previous involvement with a program, activity or entity being evaluated at a decision-making level, or in a financial management or accounting role; or seeking employment with such a program, activity or entity while conducting the evaluation	Extent to which rules or staffing procedures that prevent staff from evaluating programs, activities or entities for which they have or had decision-making or financial management roles, or with which they are seeking employment, are present and enforced
	Financial interest in the program, activity or entity being evaluated	Extent to which rules or staffing procedures are in place and enforced to prevent staff from evaluating programs, activities or entities in which they have a financial interest
	Immediate or close family member is involved in or is in a position to exert direct and significant influence over the program, activity or entity being evaluated	Extent to which rules or staffing procedures are in place and enforced to prevent staff from evaluating programs, activities or entities in which family members have influence

Source: U.S. General Accounting Office, *Government Auditing Standards*, Amendment 3 (2002); OECD/DAC Working Party on Aid Evaluation, *Glossary of Key Terms in Evaluation and Results Based Management* (2002); OECD/DAC, *Principles for Evaluation of Development Assistance* (1991); INTOSAI, *Code of Ethics and Auditing Standards* (2001); Institute of Internal Auditors, *Professional Practices Framework* (2000); European Federation of Accountants, *The Conceptual Approach to Protecting Auditor Independence* (2001); Danish Ministry of Foreign Affairs, *Evaluation Guidelines* (1999); Canadian International Development Agency, *CIDA Evaluation Guide* (2000).

wb160946
O:\KM\Lydia\ECG\ECG Independence Criteria Template.rev - March 2004.doc
March 17, 2004 1:30 PM

A. Independent Evaluation Principles

III. GPS on Evaluation of Public Sector Operations

Contents

Background

17. **Formulation Process**. The GPS on evaluation of public sector operations, prepared by the ECG Public Sector Evaluation Working Group (WGPUB), were first adopted in 2002.[20] This followed a series of workshops and analyses designed to take stock of current practices among the IFIs and identify good practices that should be promoted.[21] In 2007 an effort was made to benchmark ECG members against these standards, but the consultant undertaking that work ultimately recommended that the GPS themselves needed to be updated before a meaningful benchmarking could be completed.[22]

18. The process of revising the 2002 GPS began with a stocktaking of current IFI practices in evaluation of public sector operations, conducted in 2010.[23] At its November 2010 meeting in London, the ECG WGPUB reviewed and discussed the findings of the stocktaking paper to: (a) assess current practices of ECG members in relation to the current public sector evaluation GPS; and (b) identify key issues for consideration in the revision of the GPS, based on the comparison of current practices of the members and interviews with ECG member evaluation staff.

19. The consensus of the meeting was that the public sector evaluation GPS should be revised, building on the stocktaking paper and the model of the ECG Private Sector Evaluation Working Group (WGPSE) GPS in identifying core principles that are the basis for harmonization for ECG members, while also spelling out options in terms of operating procedures that would allow some flexibility in implementation. The work was guided by an advisory group with a representative of each member of the ECG WGPUB.

20. A draft revision of the ECG public sector evaluation GPS was circulated and subsequently discussed by the WGPUB at the March 2011 meeting of the ECG in Manila. The meeting generated additional suggestions for improvements and the public sector evaluation GPS was updated in its third draft, dated May 2011.[24] The question remained, however, as to whether the draft GPS as designed and articulated would be workable for benchmarking members' practices. To this end, the Working Group commissioned a pilot benchmarking exercise that was undertaken at the World Bank and African Development Bank, and reported on at the November 2011 ECG meeting.[25] Based on that exercise, members agreed on final revisions to these GPS.

21. **Objectives and Organization.** ECG's Good Practice Standards for the Evaluation of Public Sector Operations aim mainly to: (i) establish standards for the evaluation of IFI interventions that meet good evaluation practices generally accepted in the evaluation literature and backed by the experience of ECG members; and (ii) facilitate the comparison of evaluation results across ECG members, including the presentation of results in a common language. The

---

[20] ECG, *Good Practice Standards for Evaluation of MDB Supported Public Sector Operations*, 2002. Standards covering policy-based lending later were added as an annex.

[21] Hans Wyss, *Harmonization of Evaluation Criteria*: *Report on Five Workshop*, prepared for the Evaluation Cooperation Group, Washington, DC, 1999; John Eriksson, *Review of Good Practice and Processes for Evaluation of Public Sector Operations by MDBs*, prepared for the Working Group on Evaluation Criteria and Ratings for Public Evaluation of the Evaluation Cooperation Group (ECG), Washington, DC, 2001.

[22] V. V. Desai, "Benchmarking of MDB Evaluation Systems Against the GPS for Public Sector Operations," 2007.

[23] Kris Hallberg, *Multilateral Development Bank Practices in Public Sector Evaluation*. Final Report, March 3, 2011.

[24] Kris Hallberg, "Good Practice Standards for the Evaluation of Public Sector Operations: 2011 Revised Edition." Third Draft, May 22, 2011.

[25] Patrick G. Grasso, "Benchmarking Pilot for the Draft Public Sector GPS," 2011. The pilot was based on a review of evaluation guidelines of the two IFIs, as opposed to reviewing performance.

GPS also attempt to improve the identification and dissemination of best practices in evaluation; and improve the sharing of lessons from evaluation. The standards are applicable to projects supported by IFI investment loans, technical assistance loans, and policy-based lending. The GPS that define more effective linkages between independent and self-evaluation are presented in Chapter on Self-Evaluation.

22. The goal of documenting these standards is to harmonize evaluation practice among the ECG members, not to evaluate their evaluation functions. ECG has developed a separate set of standards for the assessment of the evaluation functions of international financial institutions.[26]

23. These GPS are organized into three sections, dealing with, report preparation and processes, evaluation approach and methodology, and dissemination and utilization. The Preparation and Processes section contains standards related to the planning, timing, coverage, selection, consultation, and review of evaluation reports. The Evaluation Approach and Methodology section contains standards relating to the objectives that form the basis of evaluations, as well as evaluation criteria and ratings. The Dissemination and Utilization section includes CED reporting and disclosure standards.

24. Within each topic area, the GPS groups the standards under a number of Evaluation Principles (EPs) which articulate the concept or purpose underlying the standards (the “what”). The EPs on public sector evaluation are composed of 7 standards and 27 elements. Each EP is supported by one or more “Operational Practices” (OPs) that describe the policies and procedures that would normally need to be adopted in order to be deemed consistent with the respective EP (the “how”). Unless otherwise noted, EPs and OPs apply to investment loans, technical assistance loans, and policy-based lending (PBL). The summary of EPs and OPs are presented below:

Summary of Standards and Elements on EPs and Number of OPs on Evaluation of Public Sector Operations

Evaluation Principles		**Number of OPs**	**Page**
Standards	**Elements**		
A. Report Preparation and Processes			
1. Timing	A. Performance Evaluation Reports (PERs)	2	33
2. Coverage and Selection	A. Accountability and Learning	1	33
	B. Sample Size	2	33
	C. Additional Sample Size	1	33
	D. Sampling Methodology	3	34
3. Consultation and Review	A. Stakeholders’ Consultation	3	34
	B. Review	3	34
B. Evaluation Approach and Methodology			
4. Basis of the Evaluation	A. Objective-based.	8	35

[26] ECG, *Review Framework for the Evaluation Function in Multilateral Development Banks*, 2009.

Evaluation Principles		Number of OPs	Page
Standards	**Elements**		
	B. Project Objectives Used in Assessments	1	36
	C. Unanticipated Outcomes	3	36
	D. Evaluations of PBLs	2	37
5. Criteria	A. Scope of Evaluation	2	37
	B. Relevance	7	38
	C. Effectiveness	3	39
	D. Intended Outcomes	4	39
	E. Efficiency	6	40
	F. Sustainability.	4	41
	G. IFI Performance	2	42
	H. Borrower Performance	2	42
6. Ratings	A. Criteria Rating	2	43
	B. Rules	2	43
	C. Aggregate Project Performance Indicator (APPI)	6	44
C. Dissemination and Utilization			
7. Dissemination and Utilization	A. Synthesis Report	5	45
	B. Accessibility of Evaluation Results	3	45
	C. Disclosure	2	45
	D. Dissemination	1	46
	E. Utilization of Recommendations	3	46
Total no. of standards: 7	**Total no. of elements: 27**	**Total no. of OPs: 83**	

Source: GPS on Evaluation of Public Sector Operations, Revised Edition, 2012.

25. In addition to the EPs and OPs, these GPS on evaluation of public sector operations also provide the following documents in the last section of this Chapter: (i) a guide on benchmarking against the GPS (Annex A.1); (ii) a note on impact and impact evaluation in the GPS(Annex A.2); and (iii) Guidance Notes that provide detailed options in three areas: attributing outcomes to the project, analyzing project efficiency, and special considerations for evaluating Policy-Based Lending (PBL)(Annexes A.3-A.5).

GPS on Evaluation of Public Sector Operations

Report Preparation and Processes

Evaluation Principle (Standards and Elements)	Operational Practices	Notes
1. Timing *A. Performance Evaluation Reports (PERs)*: Subject to the constraints and specific needs of the CED, PERs are scheduled to ensure that sufficient time has elapsed for outcomes to be realized and for the sustainability of the operation to be apparent.	1..1 PERs are scheduled to ensure that sufficient time has elapsed for outcomes to be realized, recognizing that outcomes higher in the results chain may take more time to materialize. PERs may be conducted before project closing if needed to inform the design of subsequent operations or to provide case studies for higher-level evaluations – but if this is done, the project is not rated.	
	1.2 PBLs in a series are evaluated at the end of the series.	Relevant for IFIs that provide PBLs.
2. Coverage and Selection *A. Accountability and Learning:* The CED has a strategy for its mix of evaluation products that balances the two evaluation functions of accountability and learning.	2.1 The mix of CR validations and PERs reflects the need for both accountability and learning, taking into account the quality of the IFI's CRs, the CED's budget, and the size of the population of projects ready for evaluation.	CEDs may differ in the relative emphasis they place on the two functions (accountability and learning).
B. Sample Size of Projects: For purposes of corporate reporting (accountability), the CED chooses a sample of projects for a combination of CR validations and PERs such that the sample is representative of the population of projects ready for evaluation.	3.1 The sample size for a combination of CR validations and PERs is sufficiently large to ensure that sampling errors in reported success rates (Effectiveness ratings or APPI ratings) at the institutional level are within commonly accepted statistical ranges, taking into account the size of the population of operations ready for evaluation.	
	3.2 If the sample for CR validations and PERs is less than 100% of the population of CRs and projects ready for evaluation, a statistically representative sample is selected. If the annual sample has too large a sampling error or the population is too small to yield reasonable estimates, the results from multiple years can be combined to improve the precision of the results.	A stratified random sample may be chosen. Examples of strata are regions, sectors, and types of operations.
C. Additional Sample Projects: If an additional purposive sample of projects is selected for learning purposes, it is not used by itself for corporate reporting.	4.1 In cases where an additional purposive sample of projects is selected for PERs independent from a statistically representative sample used for corporate reporting, the PER ratings are not included in aggregate indicators of corporate performance.	Relevant for IFIs that choose an additional purposive sample of projects for evaluation. Examples of selection criteria are: potential to yield important lessons; potential for planned or ongoing country, sector, thematic, or corporate evaluations; to verify CR validation ratings; and areas of special interest to the Board.
D. Sampling Methodology: The sampling methodology and significance of trends are	5.1 The CR validation sample and the PER sample are set in the CED's annual work program. Ratios and selection criteria are clearly stated.	

Evaluation Principle (Standards and Elements)	Operational Practices	Notes
reported.	5.2 In corporate reporting the confidence intervals and sampling errors are reported.	
	5.3 The significance of changes in aggregate project performance and how to interpret trends are reported.	
3. Consultation and Review *A. Stakeholders' Consultation:* Stakeholders are consulted in the preparation of evaluations.	6.1 PERs are prepared in consultation with the IFI's operational and functional departments. The criteria for selecting projects for PERs are made transparent to the stakeholders.	
	6.2 As part of the field work for PERs, the CED consults a variety of stakeholders. These may include borrowers, executing agencies, beneficiaries, NGOs, other donors, and (if applicable) co-financiers.	
	6.3 The CED invites comments from the Borrower on draft PERs. Their comments are taken into account when finalizing the report.	
B. Review: Draft evaluations are reviewed to ensure quality and usefulness.	7.1 To improve the quality of PERs, draft PERs are peer reviewed using reviewers inside and/or outside the CED.	
	7.2 To ensure factual accuracy and the application of lessons learned, draft PERs are submitted for IFI Management comments.	
	7.3 To ensure factual accuracy and the application of lessons learned, draft CR validations are submitted for IFI Management comments.	

Evaluation Approach and Methodology

Evaluation Principle (Standards and Elements)	Operational Practices	Notes
4. Basis of Evaluation *A. Objective-based:* Evaluations are primarily objectives-based.	8.1 Projects are evaluated against the outcomes that the project intended to achieve, as contained in the project's statement of objectives.	IFIs may choose to add an assessment of the achievement of broad economic and social goals (called "impacts" by some IFIs) that are not part of the project's statement of objectives. If such a criterion is assessed, it is not included in the calculation of the APPI (i.e., it falls "below the line). See also EP #3C and OP # 22.1 and # 22.2.
	8.2 Broader economic and social goals that are not included in the project's statement of objectives are not considered in the assessment of Effectiveness, Efficiency, and Sustainability. However, the relevance of project objectives to these broader goals is included as part of the Relevance assessment.	
	8.3 The project's statement of objectives provides the intended outcomes that are the focus of the evaluation. The statement of objectives is taken from the project document approved by the Board (the appraisal document or the legal document).	
	8.4 If the objectives statement is unclear about the intended outcomes, the evaluator retrospectively constructs a statement of outcome-oriented objectives using the project's results chain, performance indicators and targets, and other information including country strategies and interviews with government officials and IFI staff	
	8.5 The focus of the evaluation is on the achievement of intended outcomes rather than outputs. If the objectives statement is expressed solely in terms of outputs, the evaluator retrospectively constructs an outcome-oriented statement of objectives based on the anticipated benefits and beneficiaries of the project, project components, key performance indicators, and/or other elements of project design.	Intended outcomes are called "impacts" by some IFIs. Evaluations of countercyclical operations also focus on the achievement of outcomes. The intended outcomes may need to be constructed from sources of information other than the project documents, including interview evidence from government officials and IFI staff.
	8.6 If the evaluator reconstructs the statement of outcome-oriented objectives, before proceeding with the evaluation the evaluator consults with Operations on the statement of objectives that will serve as the basis for the evaluation.	

Evaluation Principle (Standards and Elements)	Operational Practices	Notes
	8.7 The anticipated links between the project's activities, outputs, and intended outcomes are summarized in the project's results chain. The results chain is taken from the project design documents. If the results chain is absent or poorly defined, the evaluator constructs a retrospective results chain from the project's objectives, components, and key performance indicators.	Intended outcomes are called "impacts" by some IFIs.
	8.8 PBL evaluations focus on the program of policy and institutional actions supported by the PBL, and the resulting changes in macroeconomic, social, environmental, and human development outcomes. The PBL's intended outcomes are taken from the program's statement of objectives and results chain.	Relevant for IFIs that provide PBLs.
B. Project Objectives used in Assessments: If project objectives were revised during implementation, the project is assessed against both the original and the revised objectives.	9.1 If project objectives and/or outcome targets were changed during implementation and the changes were approved by the Board, these changes are taken into account in the assessment of the core criteria. The CED defines a method for weighting the achievement of the original and revised objectives in order to determine the assessment of the core criteria.	The CED may apply the same method to projects with changes in objectives and/or outcome targets that were not approved by the Board. The evaluator may need to judge whether such changes were valid. Options for weighting include (i) using the original and revised objectives by the share of disbursements before and after the restructuring; (ii) weighting by the share of implementation time under each set of objectives; and (iii) weighting by the undisbursed balances on the loan before and after restructuring.
C. Unanticipated outcomes: The evaluation includes consideration of unanticipated outcomes.	10.1 Unanticipated outcomes are taken into account only if they are properly documented, are of significant magnitude to be consequential, and can be plausibly attributed to the project.	Unanticipated outcomes are called "unanticipated impacts" by some IFIs. Unanticipated (or unintended) outcomes are defined as positive and/or negative effects of the project that are not mentioned in the project's statement of objectives or in project design documents. Excluding consideration of unanticipated outcomes in the Effectiveness and Sustainability assessments ensures the accountability of the project for effective and sustainable achievement of its relevant

Evaluation Principle (Standards and Elements)	Operational Practices	Notes
		objectives.
	10.2 Unanticipated outcomes are taken into account in the assessment of Efficiency. The calculation of the project's *ex post* ERR includes unanticipated positive outcomes (by raising benefits) and unanticipated negative outcomes (by raising costs). The assessment of the project's cost-effectiveness includes unanticipated negative outcomes (by raising the costs of achieving the project's objectives).	
	10.3 Unanticipated outcomes, both positive and negative, are discussed and documented in a separate section of the evaluation.	
D. Evaluation of PBLs: Evaluations of PBLs assess the performance of the reform program as a whole.	11.1 Evaluations of a programmatic series of PBLs assess the performance of the entire program (the series) in addition to assessing and rating the individual operations in the series.	Relevant for IFIs that provide PBLs.
	11.2 PBL evaluations assess the results of the overall program, regardless of the sources of financing.	Relevant for IFIs that provide PBLs.
5. Criteria *A. Scope of Evaluation:* Evaluations encompass all performance attributes and dimensions that bear on the operation's success.	12.1 Investment and technical assistance operations are assessed according to a minimum of six criteria: four core criteria related to project performance (Relevance, Effectiveness, Efficiency, and Sustainability) along with IFI Performance and Borrower Performance. This applies to both CR validations and PERs.	Definitions of the criteria are given in the Glossary of Terms. IFIs may choose to assess additional criteria such as the quality of the CR, the quality of the project's monitoring and evaluation framework, social impacts, environmental impacts, institutional development impact, etc.
	12.2 PBLs are assessed according to a minimum of five criteria: three core criteria related to project performance (Relevance, Effectiveness, and Sustainability) along with IFI Performance and Borrower Performance. This applies to both CR validations and PERs.	Relevant for IFIs that provide PBLs. IFIs may choose to assess additional criteria such as the quality of the CR, the quality of the project's monitoring and evaluation framework, social impacts, environmental impacts, institutional development impact, etc.
B. Relevance; The assessment of Relevance covers both the relevance of objectives and the relevance of project design to achieve those objectives.	13.1 The relevance of objectives is assessed against beneficiary needs, the country's development or policy priorities and strategy, and the IFI's assistance strategy and corporate goals. Projects dealing with global public goods also assess relevance against global priorities.	For further guidance on assessing Relevance for PBLs, see Guidance Note 3 (Annex A.5).
	13.2 The assessment also considers the extent to which the project's	

Evaluation Principle (Standards and Elements)	Operational Practices	Notes
	objectives are clearly stated and focused on outcomes as opposed to outputs.	
	13.3 The realism of intended outcomes in the country's current circumstances also is assessed.	
	13.4 The relevance of design assesses the extent to which project design adopted the appropriate solutions to the identified problems. It is an assessment of the internal logic of the operation (the results chain) and the validity of underlying assumptions.	
	13.5 The assessment also considers the relevance of modifications to project design.	
	13.6 Whether the project's financial instrument was appropriate to meet project objectives and country needs also is assessed.	
	13.7 The relevance of objectives and design is assessed against circumstances prevailing at the time of the evaluation.	
C. Effectiveness: The assessment of Effectiveness evaluates the extent to which the project achieved (or is expected to achieve) its stated objectives, taking into account their relative importance.	14.1 The assessment of Effectiveness tests the validity of the anticipated links between the project's activities, outputs, and intended outcomes (the results chain).	Intended outcomes are called "impacts" by some IFIs.
	14.2 Both the actual and the expected results of an operation are included in the assessment of Effectiveness.	
	14.3 In evaluations of PBLs, achievement of outcomes is measured against development objectives; prior actions taken and triggers met do not by themselves provide sufficient evidence of the achievement of outcomes.	Relevant for IFIs that provide PBLs.
D. Intended Outcomes: Subject to information and CED resource constraints, the assessment of Effectiveness uses appropriate methods to determine the contribution of the project to intended outcomes in a causal manner.	15.1 Outcomes are evaluated against a counterfactual. When feasible and practical, evaluations use a combination of theory-based evaluation and impact evaluation. If an impact evaluation is not feasible or practical, evaluators at a minimum use a theory-based approach, and discuss factors other than the project that plausibly could have affected outcomes.	Intended outcomes are called "impacts" by some IFIs. Other IFIs include causality in the definition of "impact". See Guidance Note 1 (Annex A.3) for a menu of quantitative and qualitative approaches to attributing outcomes to the project.
	15.2 In rare cases where there are no other plausible explanations of the change in an outcome indicator other than the project, a "before-and-after" evaluation method is sufficient. In these cases, the evaluator presents the arguments why outside factors were unlikely to have	

Evaluation Principle (Standards and Elements)	Operational Practices	Notes
	affected outcomes.	
	15.3 In CR validations, the method used to construct a counterfactual depends on the quality of evidence in the CR. At a minimum, the evaluator uses a theory-based approach to validate the CR's conclusions regarding the links between project activities, outputs, and outcomes. Other non-project factors that plausibly could have contributed to observed outcomes are discussed.	
	15.4 PBL evaluations attempt to separate the effects of the program supported by the PBL from the effects of other factors.	Relevant for IFIs that provide PBLs. See also Guidance Note 3 (Annex A.5).
E. Efficiency: The Efficiency assessment attempts to answer two questions: (i) Did the benefits of the project (achieved or expected to be achieved) exceed project costs; and (ii) Were the benefits of the project achieved at least cost?	16.1 To address the first question (Did the benefits of the project, achieved or expected to be achieved, exceed project costs?), cost-benefit analysis is carried out to the extent that data is available and it is reasonable to place a monetary value on benefits. An economic rate of return (ERR) higher than the opportunity cost of capital indicates that project was a worthwhile use of public resources. Therefore, when an ERR is calculated, it would normally need to be greater than the opportunity cost of capital for a fully satisfactory assessment of Efficiency. Other thresholds may be used -- varying for example by sector -- but if so are explicitly defined by the CED.	See Guidance Note 2 (Annex A.4) for further detail on options for assessing Efficiency. Note that Efficiency is assessed for investment and TA loans but not for PBLs (see OPs # 12.1 and # 12.2).
	16.2 The methodology and assumptions underlying the calculation of an economic rate of return or net present value are clearly explained and transparent. *Ex post* estimates are compared with the *ex ante* estimates in the project documents.	Relevant when ERRs/NPVs are estimated.
	16.3 Sensitivity tests on ERRs based on possible changes in key assumptions are carried out. These assumptions reflect any concerns in the assessment of Sustainability.	Relevant when ERRs/NPVs are estimated.
	16.4 To address the second question (Were the benefits of the project achieved at least cost?), cost-effectiveness analysis is carried out. The analysis considers the cost of alternative ways to achieve project objectives, unit costs for comparable activities, sector or industry standards, and/or other available evidence of the efficient use of project resources.	
	16.5 In addition to the traditional measures of efficiency (cost-benefit analysis and cost-effectiveness analysis), the Efficiency assessment considers aspects of project design and implementation that either contributed to or reduced efficiency. For example, implementation	

Evaluation Principle (Standards and Elements)	Operational Practices	Notes
	delays -- to the extent they are not already captured in the evaluation's cost-benefit or cost-effectiveness analysis -- would have an additional negative impact on Efficiency.	
	16.6 For evaluations of TA operations, if project design includes a pricing policy or pricing guidelines for TA, the Efficiency assessment considers the degree to which these policies were implemented.	Relevant for IFIs that provide lending for TA.
F. Sustainability: The assessment of Sustainability is based on the risk that changes may occur that are detrimental to the continued benefits associated with the achievement or expected achievement of the project's objectives, and the impact on that stream of benefits if some or all of these changes were to materialize.	17.1 The Sustainability assessment considers several aspects of sustainability, as applicable: technical, financial, economic, social, political, and environmental. It also considers the degree of government ownership of and commitment to the project's objectives; the ownership of other stakeholders (e.g., the private sector and civil society); and the degree of institutional support and the quality of governance. The risk and potential impact of natural resource and other disasters is also considered.	
	17.2 Sustainability is determined by an assessment of both the probability and likely impact of various threats to outcomes, taking into account how these have been mitigated in the project's design or by actions taken during implementation. The evaluator takes into account the operational, sector, and country context in projecting how risks may affect outcomes.	
	17.3 The Sustainability assessment refers to the sustainability of intended outcomes that were achieved or partially achieved up to the time of the evaluation, as well as intended outcomes that were not achieved by the time of the evaluation but that might be achieved in the future. To avoid overlap with Effectiveness, Sustainability is not downgraded based on incomplete achievement of objectives *per se*.	Intended outcomes are called "impacts" by some IFIs.
	17.4 The time frame for the sustainability assessment depends on the type of project being evaluated, but is clearly stated in the evaluation. For investment operations, the time frame for the Sustainability assessment is the anticipated economic life of the project. For PBLs, the time frame may need to be longer to encompass the persistence of results from policy and institutional actions adopted under the operation. For some types of investment projects, the starting point of the sustainability analysis may not be the time of the evaluation, but rather the start of operation of the project.	For PBLs, see also Guidance Note 3 (Annex A.5).
G. IFI Performance: The assessment of IFI Performance covers the quality of services	18.1 The assessment of IFI Performance at project entry covers the IFI's role in ensuring project quality and in ensuring that effective	

Evaluation Principle (Standards and Elements)	Operational Practices	Notes
provided by the IFI during all project phases.	arrangements were made for satisfactory implementation and future operation of the project. This includes: – the quality of the analysis conducted to identify problems and possible solutions; – the consideration of alternative responses to identified problems; – the degree of participation of key stakeholders; the use of lessons learned from previous operations; – the quality of risk analysis and the adequacy of proposed risk mitigation measures; – the adequacy of institutional arrangements for project implementation; – the identification of safeguards relevant to the project; and – the IFI's efforts to ensure the quality of the monitoring and evaluation framework.	
	18.2 The assessment of IFI performance during project supervision is based on the extent to which the IFI proactively identified and resolved problems at different stages of the project cycle, including: – modifying project objectives and/or design as necessary to respond to changing circumstances; – enforcing safeguard and fiduciary requirements; and – ensuring that the monitoring and evaluation system was implemented.	
H. Borrower Performance: Borrower Performance assesses the adequacy of the Borrower's assumption of ownership and responsibility during all project phases.	19.1 The assessment of Borrower Performance focuses on processes that underlie the Borrower's effectiveness in discharging its responsibilities as the owner of a project, including: – government and implementing agency performance in ensuring quality preparation and implementation; – compliance with covenants, agreements, and safeguards; – provision of timely counterpart funding; – provision of timely counterpart funding; and – measures taken by the Borrower to establish the basis for project sustainability, particularly by fostering participation by the project's stakeholders.	
	19.2 The assessment covers the performance of the government as well as the performance of implementing agencies.	

Evaluation Principle (Standards and Elements)	Operational Practices	Notes
6. Ratings *A. Criteria Rating:* Each of the six criteria (five for PBLs) is assigned a rating.	20.1 For Relevance, Effectiveness, Efficiency, and Sustainability, the criterion is rated on the degree of achievement, for example from “negligible” to “high”. Normally a four-point rating scale is used. Ratings may be either categories or numbers.	Additionally, ratings of “non-evaluable” and “not applicable” may be used.
	20.2 For IFI Performance and Borrower Performance, the number of rating scale points is either four or six. The ratings measure degrees of satisfactory or unsatisfactory performance, for example ranging from “Highly Successful” to “Highly Unsuccessful”. The rating scale is symmetric. Ratings may be either categories or numbers.	Additionally, ratings of “non-evaluable” and “not applicable” may be used.
B. Rules: Rules for assigning criteria ratings are clearly spelled out.	21.1 If the rating for a given criterion is constructed from ratings on sub-criteria or from ratings on different elements of the criterion, the rules for the aggregation are clearly spelled out in evaluation guidelines.	For example: (i) the Relevance rating may be based on separate ratings for the relevance of objectives, design quality and preparation, institutional arrangements, and the relevance of modifications; (ii) the Effectiveness rating may be based on separate ratings for the achievement of each of the project objectives; (iii) the Efficiency rating may be based on separate ratings for overall economic and financial performance, cost-effectiveness, and timeliness of outputs and outcomes.
	21.2 In evaluation reports, evaluators provide a justification for each rating.	
C. APPI: An Aggregate Project Performance Indicator (APPI) is constructed from the core criteria.	22.1 For investment and TA loans, the APPI is constructed from the four core criteria: Relevance, Effectiveness, Efficiency, and Sustainability.	If additional (non-core) criteria are included in the evaluation (see Note to OP # 12.1 above), their ratings are not used in the calculation of the APPI (i.e., they are “below the line”). A second aggregate indicator, including these additional criteria, may be constructed.
	22.2 For PBLs, the APPI is constructed from the three core criteria: Relevance, Effectiveness, and Sustainability.	Relevant for IFIs that provide PBLs.
	22.3 In constructing the APPI, the component criteria are normally given equal weights. The relative ratings of the core criteria are reviewed for logical consistency. If there are inconsistencies, the evaluator may choose to assign unequal weights to the component criteria, explaining the reasons behind them.	For example, for an ineffective project to have a high rating on Sustainability would be unusual. Similarly, for a project to be given a highly successful rating if its sustainability was in doubt or if its relevance was poor at

Evaluation Principle (Standards and Elements)	Operational Practices	Notes
		project completion would be unusual.
	22.4 If criteria ratings are given numerical values, the rules for constructing the APPI rating category (e.g., by rounding or by using threshold values) are clearly spelled out in evaluation guidelines.	
	22.5 For the APPI, the number of rating scale points is either four or six. The rating scale is symmetric. Ratings may be either categories or numbers.	
	22.6 If, in addition to the APPI, a second aggregate indicator is calculated, the component criteria and rules for constructing the second indicator are clearly spelled out in evaluation guidelines. Both the APPI and the second aggregate indicator are presented in corporate reports.	Relevant for IFIs that construct a second aggregate indicator.

Dissemination and Utilization

Evaluation Principle (Standards and Elements)	Operational Practices	Notes
7. Dissemination and Utilization *A. Synthesis Report:* The CED prepares a periodic synthesis report.	23.1 At least every three years, the CED prepares a periodic synthesis report addressed to the IFI's Management, staff, and Board. The frequency of reporting depends on the significance of changes in aggregate ratings and recommendations year-to-year.	
	23.2 The review includes a synthesis of CR validations and PERs produced during the period covered. The criteria and rating systems used in the evaluations are clearly spelled out. All ratings reported are those from the CED; differences in aggregate ratings between CR validations/PERs and the CRs are disclosed.	
	23.3 The CED reports periodically (at least every three years) to the IFI's Board of Directors and Management on the quality of the IFI's self-evaluation system, including the application of lessons in new operations.	
	23.4 The CED's synthesis ratings are included in integrated corporate performance reporting.	
	23.5 Since the APPIs for investment/TA loans and PBLs are based on different criteria and thus are not strictly comparable, they are reported separately in corporate performance reporting.	

Evaluation Principle (Standards and Elements)	Operational Practices	Notes
B. Accessibility of Evaluation Findings: The CED makes evaluation findings and lessons easily available to IFI staff.	24.1 The CED makes available to all IFI staff a range of user-friendly dissemination products covering all of its evaluation products along with the periodic synthesis report.	
	24.2 The CED relies primarily on its intranet website for document posting and notifies staff of new items through the corporate website.	
	24.3 The CED maintains a searchable lessons-learned system to assist Operations staff to find lessons applicable to new projects. The entries include specific lessons along with contextual material to allow the lessons to be readily applied.	
C. Disclosure: Within the guidelines of the IFI's overall disclosure policy, the CED discloses all evaluation products.	25.1 The CED's disclosure policy for evaluation products is explicit and consistent with the IFI's general disclosure policy.	
	25.2 The CED discloses the full report of all of its evaluation products. Only in exceptional cases is some measure of confidentiality warranted. In these cases, if possible, evaluation reports are redacted and then disclosed.	Examples of exceptional cases would be (i) an evaluation of an operation with a semi-public/semi-private entity, for which the relevant disclosure standard may be that of the Private Sector GPS; and (ii) an evaluation of a PBL for which the disclosure of evaluation results would be likely to seriously compromise the process of policy change.
D. Dissemination: The CED pro-actively reaches the public with its evaluation results.	26.1 The CED has a strategy for disseminating its evaluation products according to the types of products it produces and the audiences it intends to reach: IFI staff, member governments, other client stakeholders, civil society organizations, academia, and others.	Options include evaluation summaries, inclusion of evaluation findings in IFI annual reports; hosting conferences, training sessions, and public consultations on evaluation methods and findings and methodologies; and the use of websites, public media, and social media.
E. Utilization of Evaluation Recommendations: The CED follows up on IFI Management's implementation of recommendations made by the CED.	27.1 Based on its PERs and higher-level evaluations, the CED makes recommendations to IFI Management and the Board to improve the IFI's effectiveness. These include a specific, time-bound set of actions to be taken by IFI Management which can reasonably be taken in the short term and can be monitored by IFI Management and the CED.	
	27.2 The CED maintains a tracking system for recording and following up on steps taken to respond to each recommendation that was endorsed by IFI Management.	

Evaluation Principle (Standards and Elements)	Operational Practices	Notes
	27.3 The CED reports to the Board on IFI Management follow-up to its recommendations.	

Annex III.1: Benchmarking Against the GPS

1. In benchmarking IFI adherence to these GPS, the following will be taken into account:

- Consistency with the GPS is measured on the basis of the IFI's consistency with the EPs.
- An IFI can achieve consistency with an EP by implementing all of the associated OPs. Alternatively, an IFI may also achieve consistency with an EP by adopting other practices that it believes (and the benchmarking consultant agrees) meet the intent of the EP.
- Unless stated otherwise in the wording of an OP, the judgment that an OP is being implemented should be based on evidence of actual implementation of the OP; it is not sufficient for the IFI to document procedures and policies if these are not carried out in practice.
- The benchmarking scheme allows for "partial" scores. In other words, OPs may be found to be "fully implemented", "partially implemented", or "not implemented". Consistency with the corresponding EP may be found to be "fully consistent", "partially consistent", or "not consistent". The determination of the EP rating is left to the judgment of the benchmarking consultant, but it would be unusual for an EP to be rated "fully consistent" if one or more of the associated OPs were found to be "partially implemented".
- EPs and OPs that are not relevant for an IFI are not included in that IFI's benchmarking assessment. For example, the GPS contains one EP (#4D) that is relevant only for IFIs that provide policy-based lending. This EP would be ignored for IFIs that do not provide PBLs. In addition, within a particular EP, some OPs may be relevant only for IFIs that provide PBLs. For IFIs that do not provide PBLs, consistency with the EP is based on consistency with the remaining OPs. For example, if an EP has four OPs and one OP is relevant only for IFIs that provide PBLs, the IFIs that do not provide PBLs would be assessed against only the remaining three OPs to determine their consistency with the EP.

2. The following schematic illustrates the benchmarking process. Benchmarking will be used to characterize the extent of harmonization across institutions at the level of each EP. However, the ECG may decide at a future date to apply a scoring mechanism for comparative purposes.

Examine the IFI's Operational Practices under each EP	→	Determine Consistency with the EP
OPs listed under each EP:		Score each EP:
Fully Implemented Partially Implemented Not Implemented		Fully Consistent Partially Consistent Not Consistent

Annex III.2: A Note on "Impact" and "Impact Evaluation" in the GPS

1. The term "impact" has multiple meanings. It is understood to mean different things by different IFIs, and even within a given definition, there are sometimes multiple meanings. According to the OECD-DAC glossary of evaluation terms, "impacts" are defined as:

> *"positive and negative, primary and secondary long-term effects produced by a development intervention, directly or indirectly, intended or unintended."*[27]

2. The OECD-DAC definition encompasses several concepts:

- a time dimension ("long-term effects"). The long-term nature of "impacts" is sometimes contrasted with the short- and medium-term nature of "outcomes;"
- two levels of results ("primary" and "secondary");
- the pathways of transmission of results ("directly or indirectly"); and
- results captured in the project's results chain ("intended effects") as well as others outside of the results chain ("unintended effects").

3. Not surprisingly, and partly as a result of the multiple meanings in the OECD-DAC definition, various IFIs understand "impact" to mean different things.

- Sometimes it is understood to include both causal linkages (results attributable to the intervention) and long-term effects.
- Some IFIs consider "impact" to mean "the final level in the results chain" or "the highest level of objectives that an intervention has identified "(i.e., the ultimate intended results of the intervention).
- Some IFIs understand "impacts" to mean overall development goals (e.g., Millennium Development Goals in various sectors), and use the term "outcomes" to mean "project purposes/objectives." This definition of the term "impact" is not the same as the "final level in the results chain."
- Some IFIs define "impact" to mean specific types of results that are of particular interest to the IFI – such as institutional development, environmental impact, and/or social impact – that may or may not appear in the project's results chain.
- "Impact" sometimes also has the meaning of the unit of analysis examined (that is, results measured at the sector, region, or country level versus those measured for project beneficiaries).

These meanings are at variance with each other, and a number with the OECD-DAC definition, which includes both intended and unintended results.

4. *Because of the multiple and sometimes conflicting meanings that are associated with the word "impact," the GPS avoids using the term* – either as a level in the results chain or as an evaluation criterion. Instead, the various meanings of the OECD-DAC and other definitions are stated directly – e.g., "long-term effects"'; "results attributable to the project", and "broad social and economic goals". The final level of the results chain – reflecting the ultimate intended results of the project – is generally called "intended

[27] Organisation for Economic Co-operation and Development (2002). *Glossary of Key Terms in Evaluation and Results Based Management.* Paris. The Network of Networks on Impact Evaluation (NONIE) uses the same definition of "impacts": see Leeuw, Frans and Jos Vaessen (2009), *Impact Evaluations and Development: NONIE Guidance on Impact Evaluation.*

outcome" in the GPS. Where relevant, the equivalent use of the term "impact" by some IFIs is shown in the Notes column.

5. Note that a results chain can have several levels of outcomes. In the evaluation literature, these are sometimes distinguished by using the terms "intermediate outcomes" and "final outcomes". However, to avoid introducing new terminology, the GPS simply uses the word "outcome" to refer to levels in the results chain beyond "outputs."

6. In avoiding the term "impact," the GPS does not intend to suggest that IFI projects should not aim at achieving broad economic and social goals. However, the GPS calls for project objectives to focus on outcomes for which the project can reasonably be held accountable, avoiding objectives beyond the purview of the project. To the extent that higher-level social and economic objectives and corporate goals are included, they should be targeted at segments of the population that can reasonably be expected to be affected by the project, directly or indirectly (See OP #1.1 in GPS on Self-Evaluation of Public Sector in Chapter VI).

7. As with the term "impact," there is confusion regarding the term "impact evaluation." For some, "impact evaluation" means "an evaluation of impact", however "impact" is defined. For example, some IFIs understand "impact evaluation" to mean an assessment of the achievement of the objectives reflected in the final level of the results chain; others interpret it to mean the project's effect on broad social and economic indicators that are not included in the results chain. For other IFIs, "impact evaluation" means an evaluation that establishes causality, i.e., attributes results to the project.

8. The OECD-DAC Glossary does not have a definition of "impact evaluation". The definition used by the International Initiative for Impact Evaluation (3ie) is based on attribution and methodology:

"Impact evaluations measure the net change in outcomes amongst a particular group, or groups, of people that can be attributed to a specific program using the best methodology available, feasible and appropriate to the evaluation question(s) being investigated and to the specific context."[28]

The 3ie definition is consistent with that of the Network of Networks on Impact Evaluation (NONIE), which focuses on two underlying premises for impact evaluation: (i) attribution: attributing impacts to interventions, rather than just assessing what happened; and (ii) counterfactual: an attempt to gauge *what would have occurred* in the absence of the intervention with *what has occurred* with the intervention implemented.[29]

9. Consistent with the 3ie and NONIE definitions, the GPS defines "impact evaluation" as:

An evaluation that quantifies the net change in outcomes that can be attributed to a specific project or program, usually by the construction of a plausible counterfactual.[30]

---

[28] International Initiative for Impact Evaluation. *3ie Principles for Impact Evaluation*. See www.3ieimpact.org.
[29] See Leeuw and Vaessen (*op. cit.*).
[30] See Guidance Note 1 on attributing outcomes to a project.

10. Under this definition, an impact evaluation usually employs quantitative methods to measure the net change in outcomes attributable to the project. However, qualitative methods can also be used to infer causality (see EP #5D and Annex A.3). Together, impact evaluation and qualitative methods to establish causality might be called "attribution analysis". However, the GPS avoids introducing this new term.

11. Note that, in principle, impact evaluation can be applied to any of the levels of outcome in the results chain ("intermediate outcomes" and "final outcomes"). For higher-level outcomes, determining the appropriate counterfactual becomes more difficult because of the greater influence of factors external to the project. For lower-level outcomes (closer to the level of outputs), the counterfactual approaches the "before project" situation. In those cases, the results of an impact evaluation would be the same as the results of a naïve (before-and-after) approach.

Annex III.3: Guidance Note 1: Attributing Outcomes to the Project

1. The definition of "impact evaluation" used in the 2012 GPS is "an evaluation that quantifies the net change in outcomes that can be attributed to a specific project or program, usually by the construction of a plausible counterfactual. "[31] Thus, impact evaluation focuses on quantifying the incremental contribution to results that is attributable to the intervention.

2. In theory, a comparison with a counterfactual can be done for any level in the project's results chain.[32] But the question of attribution becomes trivial at lower levels of the results chain: the counterfactual of "what would have happened in the absence of the project" becomes "before-project". A good example is the impact of a water supply project on the time household members spend collecting water.[33] The average water collection time falls after the project. The only plausible explanation is the improved proximity or predictability of water. In this case, the counterfactual (what would have been the time spent gathering water, without the project) is that the time would have remained the same as before the project. A before-and-after approach is sufficient to determine the change in outcome attributable to the project. At higher levels in the results chain, however, construction of a counterfactual is more difficult.

3. A variety of quantitative methods can be used in impact evaluations (see below). Even if it is not possible or desirable to conduct an impact evaluation, qualitative methods can be used to construct a plausible counterfactual and make an informed judgment (but not quantify) the additionality of the project to the intended outcomes.

4. Establishing a causal relationship between the project and its outcomes starts with the project's results chain that links project activities with intended outputs and outcomes. In other words, the starting point is to build up the program theory. This is sometimes referred to as a "theory-based" evaluation framework. This approach maps out the channels through which the activities, inputs, and outputs are expected to result in the expected outcomes. It is a systematic testing of all of the links (assumptions) in the results chain. It also allows for the identification of unintended effects.

5. The Network of Networks on Impact Evaluation (NONIE) provides the following overall guidance for impact evaluation:

   Carefully articulate the theories linking interventions to outcomes. Address the attribution problem. If possible, use quantitative approaches, embedding experimental and quasi-experimental designs in a theory-based evaluation framework. Qualitative techniques should be used to evaluate attribution issues for which quantification is not feasible or practical. Preference is to use mixed-methods designs. Use existing research relevant to the results of the intervention.[34]

---

[31] Note that this definition of impact evaluation is not the same as "an evaluation that focuses on the final level in the causal chain" (e.g., social and economic outcomes such as poverty reduction, which are sometimes called "impact").

[32] See White, Howard (2007). *Evaluating Aid Impact*. World Institute for Development Economics Research, Research Paper No. 2007/75, November.

[33] Example from White, Howard (2009). *Some Reflections on Current Debates in Impact Evaluation.* International Initiative for Impact Evaluation Working Paper No. 1, April.

[34] Leeuw, Frans and Jos Vaessen (2009). *Impact Evaluations and Development: NONIE Guidance on Impact Evaluation*. Network of Networks on Impact Evaluation.

6. The remainder of this Guidance Note discusses various quantitative methods that can be used to attribute results to project activities, what to do when quantitative techniques are not feasible, or practical, and how to construct a counterfactual for policy-based operations.

Quantitative Methods

7. The main designs for impact evaluation include the following:[35]

- Randomized assignment: Randomized assignment of treatment essentially uses a lottery to decide who among the equally eligible population receives the project treatment and who does not. Under specific conditions, randomized assignment produces a comparison group that is statistically equivalent to the treatment group.
- Difference-in-differences: Estimates the counterfactual for the change in outcome for the treatment group by calculating the change in outcome for the comparison group. This method takes into account any differences between the treatment and comparison groups that are constant over time.
- Matching Estimators: Uses statistical techniques to construct an artificial comparison group by identifying, for every possible observation under treatment, a non-treatment observation (or set of non-treatment observations) that has the most similar characteristics possible. These "matched" observations then become the comparison group that is used to estimate the counterfactual. A common way to match different units is to model how likely each unit is to be treated, based on observed variables, and then to match treated and untreated units based on this likelihood (or "propensity score").
- Regression approaches: An alternative to matching, usually done by observing both treated and control units and to "control" for as many pre-program covariates as possible. This method is similar to matching techniques in that it uses observed characteristics of treated and untreated units to try to make them "similar".
- Instrumental variables: A method used to control for selection bias due to unobservables. Certain variables are chosen that are believed to determine program participation but not outcomes. These instrumental variables are first used to predict program participation; then, the predicted values are used to see how outcome indicators vary with the predicted values.
- Regression Discontinuity: An impact evaluation method that can be used for programs that have a continuous eligibility index with a clearly defined cutoff score to determine who is eligible and who is not. The regression discontinuity measures the difference in post-intervention outcomes between the units just above and just below the eligibility cutoff.
- Modeling the theory: The determinants of outcomes are estimated using regression models. The determinants of these determinants are also modeled, working down the results chain until the link is made to project inputs.

8. Experimental and quasi-experimental methods should be used to construct a comparison group when they are appropriate, feasible, and practical. In many situations, however, they are not possible – for example, when the project is comprehensive in scope (such as economy-wide policy reforms) or works with a small number of entities

[35] See Paul J. Gertler and others. (2001), *Impact Evaluation in Practice*. World Bank; and Yuri Soares (2011), *Note on the Practice and Use of Impact Evaluation in Development: Reflections for the Evaluation Cooperation Group (ECG) Conference in Manila*; March.

(such as institutional reforms). Random assignment also may not be possible for political or ethical reasons.[36]

9. Often, baseline data are not available. Possible alternative designs include (i) single difference methods (after-project comparisons of participants and non-participants), if the groups are drawn from the same population and some means is found to address selection bias; and (ii) using another dataset to serve as a baseline.

Qualitative Methods

10. If quantitative methods are not feasible, or practical, the evaluation should employ “causal contribution analysis” by building a strong descriptive analysis of the causal chain. The evaluator attempts to provide evidence that assumed links in the chain in fact occurred, or identify breaks in the chain so as to argue that expected results did not occur. Arguments can be strengthened by triangulation, i.e., drawing on a variety of data sources and approaches to confirm that a similar result obtained from each.

11. To analyze the links in the causal chain, the evaluator:

- Assesses the causal chain in relation to the needs of the target population, collaborating with stakeholders and experts.
- Examines the critical assumptions and expectations inherent in the project’s design, reviewing the logic and plausibility of the results chain. Again, this is done in collaboration with stakeholders.
- Uses available research evidence and practical experience elsewhere, comparing the project with projects based on similar concepts.
- Observes the project in operation, focusing on interactions that were expected to produce the intended outcomes.[37]

12. Beneficiary surveys, focus groups, structured interviews, and other instruments are other techniques commonly used to provide qualitative evidence for causal contribution analysis.

13. Case studies are useful as a complementary method. They can describe what the implementation of the project looked like on the ground and why things happened the way they did. Not only are case studies more practical than large national studies, they also provide in-depth information that is often helpful to decision makers.

14. Most evaluation textbooks and guidelines advocate a mixed-method approach, combining quantitative and qualitative methods when possible. This is because some impact evaluation methods give results out of a “black box” – i.e., they can be used to quantify the results of a project but do not necessarily explain why the results occurred. For other reasons, it may be useful to compare the results of before-and-after comparisons with the results of using other methods to determine causality.

---

[36] See Soares (op. cit.) for the limitations of randomized control trials and of impact evaluation in general.
[37] Morra-Imas, Linda, and Ray C. Rist (2009). *The Road to Results: Designing and Conducting Effective Development Evaluations.* World Bank.

Policy-Based Lending[38]

15. It is more difficult to assess and attribute the results of PBL operations than investment loans. PBLs support a program of policy and institutional changes, and often operate at the economy-wide level. Assessing PBL outcomes is complicated by the interaction of IFI-supported reforms with contemporaneous changes in other public policies, shocks, cyclical factors, and changes in market conditions. Isolating and attributing change to any particular set of PBL-supported policy and institutional actions is information-intensive and analytically demanding.

16. Quantitative approaches are available to isolate PBL outcomes and to compare performance with a counterfactual scenario. Some PBL evaluations have employed simple growth decomposition methods to isolate the effects of policy change from major shocks and changes in the terms of trade. Others have used cross-country regression models to distinguish the effects of policy change from starting points and structural characteristics of borrowers.

17. A number of qualitative approaches also may be useful for separating the effects of the program supported by the PBL from other factors, and assess the influence of the PBL on program outcomes. These include:

- A review of performance indicators, activity surveys, and structured interviews with key stakeholders can be used to assess whether or not the implementation of PBL-supported measures actually gave rise to the outputs and outcomes expected of them.
- Beneficiary satisfaction surveys can be conducted, with the results of policy and institutional change being "scored" directly by stakeholders.
- In some cases, PBL appraisal reports contain a "without reform" scenario for certain key outcome variables.
- It may also be possible to use the performance of policy and impact variables in similar countries that did not undertake PBL-supported reforms as a baseline comparator.
- Observed outcomes can be benchmarked against regional or international standards of public policy and institutional performance to assess the significance of PBL-supported actions to transforming policy settings.
- Advantage should also be taken of previous evaluations and research, including comparative studies of experiences with structural adjustment. They can suggest factors that have been associated with successful adjustment.
- The insights obtained from other sources of information, including key informant and group interviews and mini-surveys, can shed further light on attribution issues. Individuals intimately involved in a reform process can often identify the counterfactual.

18. In some cases, a qualitative assessment of the linkages between the PBL and the desired outcomes is sufficient to identify what elements were missing, or could have been better designed. With adequate benchmarks and *ex post* performance information, simulations, cost-benefit, cost-effectiveness, and other quantitative techniques can be used to inform such judgments. Evaluations of PBL poverty outcomes draw on a variety

[38] Tabor, Steven R., and Stephen Curry (2005). *Good Practices for the Evaluation of Policy-Based Lending by Multilateral Development Banks.* Asian Development Bank report prepared for ECG, March.

of techniques and survey instruments to assess changes in living standards, livelihoods, and benefit incidence.

19. Precisely attributing the contribution of any single PBL is nearly impossible when many stakeholders have a hand in policy change, but evaluators can assess what additional value the PBL had to the policy change process, beyond the provision of financial support. For example, the additionality of a PBL can be evaluated in terms of whether or not it (i) accelerated (or delayed) reform, (ii) strengthened the hand and credibility of reformers, (iii) raised the perceived political returns to reform in terms of easing budget constraints and positive reputation effects, (iv) fostered policy learning, (v) built domestic capacity to design policy, and (vi) spurred debate and dialogue on new approaches to meeting development objectives.

Annex III.4: Guidance Note 2: Efficiency Analysis

1. Efficiency is a measure of how well the project used resources in achieving its outcomes. It is measured in economic terms because it examines whether the project was an efficient use of resources for the country.

   - Cost-benefit analysis is a way for the IFI to verify to its stakeholders that its operations are having a net positive effect on the standard of living in member countries. It is a quantitative analysis performed to establish whether the present value of benefits of a given project exceeds the present value of costs. Such analysis usually also produces both a net present value (NPV) calculation and an economic rate of return (ERR) calculation.[39] When compared with the opportunity cost of capital, the ERR can show whether or not the project was a worthwhile use of public resources. It can be used to compare the costs and benefits of projects in different sectors, using a common "cost of capital" benchmark.
   - Cost effectiveness analysis attempts to compare different alternatives for achieving the same result. It can be used to show whether the outcomes were delivered at least cost compared to alternative ways of achieving the same outcomes.

2. Consistent with the evaluation guidelines of several ECG members, the GPS adopts the standard that evaluations should conduct cost-benefit analysis if data are available and it is reasonable to place a monetary value on project benefits. When compared with the opportunity cost of capital, the ERR can show whether the project was a worthwhile use of public resources. Cost-effectiveness analysis should be conducted in all evaluations. Both types of analysis should present their assumptions and methodology in a transparent way. In addition to these traditional measures of efficiency, the evaluation should discuss aspects of project design and implementation that either contributed to or reduced efficiency, and factor these into the rating on Efficiency.

3. Some projects include the objective of increasing the efficiency of a sector – for example, increasing the efficiency of provision of health care services. The achievement of the sector objective is relevant to the assessment of the project's effectiveness, but is not in and of itself evidence of the project's efficiency.

Cost-Benefit Analysis

4. Applying generally accepted good practices for cost-benefit analysis of projects suggests that:

   - Benefits and costs should be measured against the situation without the project.
   - Macroeconomic, institutional, behavioral, and financial assumptions underlying the analysis should be clearly spelled out.
   - Benefit and cost estimates should not merely re-state the assumptions and values used in *ex ante* ERR estimates. Up to the year of the evaluation, actual costs and

[39] A project's economic rate of return (ERR) is the internal rate of return of a time series of the project's economic costs and benefits. The ERR is an absolute measure of project benefits in relation to costs. A project's net present value (NPV) is the sum of the present values of the time series of project costs and benefits. A project's financial rate of return (FRR) is the internal rate of return of a time series of cash flows describing the project's financial investments and returns. The ERR and NPV incorporate measures of externalities and public goods, whereas the FRR does not.

benefits should be used, while new projections should be made for the remaining useful life of the project. The assumptions underlying projected costs and benefits should be based on the lessons of experience with the country, the sector, and the Borrower.

- Sensitivity tests on ERRs based on possible changes in key assumptions should be carried out as part of the evaluation. These assumptions reflect any concerns raised in the assessment of Sustainability – for example, increases in key prices, operation and maintenance expenditures being less than assumed at appraisal, or changes in government policies.
- The analysis should take into account any domestic or cross-border externalities.

5. *Ex post* ERRs should be compared with *ex ante* appraisal estimates. The evaluation should indicate the components and the percentage of total project costs covered by the analysis, noting any differences from the analysis at appraisal.

Cost-Effectiveness Analysis

6. Cost-effectiveness analysis asks whether the project achieved its outcomes at least cost compared to alternative ways of achieving the same outcomes. The analysis can use either the cost per unit of input or cost per unit of output. Whether benefits are measured in monetary or other terms, flows should be adjusted to reflect real use of resources. When distortions are considerable, values should be adjusted to reflect social opportunity costs. Cost per beneficiary is often used in sectors such as education, health, and urban development.

7. Comparators for cost-effectiveness analysis may be drawn from similar projects in the same country implemented by the Government or other development partners; from similar projects in other countries; or from other local, national, or regional benchmarks.

Implementation Costs

8. The evaluation of Efficiency also should include aspects related to the project's design and implementation that either contributed to or reduced efficiency. Implementation delays in and of themselves may not necessarily reduce efficiency if they have an equal impact on discounted costs and benefits. However, additional administrative costs that arise from delays and extensions of closing dates would have a negative impact on efficiency.

Annex III.5: Guidance Note 3: Special Considerations for Policy-Based Lending

1. This guidance note suggests ways in which the EPs and OPs in this document may be applied to Policy-Based Lending. The guidance is taken from the PBL GPS,[40] the background paper for the PBL GPS prepared by the Asian Development Bank,[41] as well as current guidelines and practices of the four ECG members that provide policy-based lending (Asian Development Bank, African Development Bank, World Bank, and Interamerican Development Bank).[42]

2. Policy-based lending is a form of lending in support of a time-bound set of discrete changes in public policy. PBLs may be directed at the economy-wide or sectoral level to support global, regional, national, sector, or thematic objectives. PBLs may occur as individual operations or a series of operations (a programmatic series), and disbursements made in single or multiple tranches. The agreed-upon policy and institutional actions determine a set of conditions or prior actions, against which tranches are disbursed or subsequent operations in the series are approved. Functionally, two main categories of PBLs can be distinguished: (i) stabilization operations, focusing on macroeconomic measures aimed at restoring short-term balance of payments and fiscal equilibrium; and (ii) development policy operations, focusing on policy and institutional reforms aimed at improving the medium-term structural, sector, and sub-national enabling environment for growth, poverty reduction, and sustainable development.

3. PBL inputs are the agreed-upon policy conditions, finance, and technical assistance (TA); PBL outputs refer to the implementation of reforms and the disbursement and utilization of PBL finance. PBL intermediate outcomes are the changes in the policy or institutional environment that occur as a result of the implementation of reforms. PBL final outcomes are changes in economic, social, environmental, and human development performance attributable to PBL-supported policy and institutional reforms.[43]

Timing

4. The evaluation principle on the timing of ex-post evaluations is the same for PBLs as investment loans: PERs are timed to ensure that sufficient time has elapsed for outcomes to be realized and for the sustainability of the operation to be apparent. However, the outputs and outcomes generated by PBLs may become apparent only some time after a program has been completed. In addition, time is needed after an operation has been completed to assess whether or not the government has stayed the course and implemented agreed upon reforms. Thus, compared to investment loans, the appropriate time for post-evaluation of PBLs may be longer after project completion. Where there is a series of overlapping and related PBLs, the timing question is more

---

[40] ECG (2004), *Good Practices for the Evaluation of Policy-Based Lending.* Addendum to the 2002 GPS.

[41] Asian Development Bank (2005). *Good Practices for the Evaluation of Policy-Based Lending by Multilateral Development Banks*. Prepared for the Evaluation Cooperation Group of the Multilateral Development Banks, March.

[42] In particular, Asian Development Bank (2006), *Guidelines for Preparing Performance Evaluation Reports for Public Sector Operations*, Addendum 1 on “Evaluating Program Lending”; and African Development Bank (2001), *Revised Guidelines on Project Completion Report (PCR) Evaluation Note and Project Performance Evaluation Report (PPER)*, section on “Special Considerations in Evaluating Adjustment Lending Operations”. These references also contain guidance on sub-criteria under Relevance, Effectiveness, Efficiency, and Sustainability for PBLs.

[43] ECG 2004, p. 2.

complicated, since the policy reform effort will have been supported by several operations. In such instances, the appropriate timing of an evaluation depends on a judgment of when the outcomes of a sequence of related PBLs are likely to be realized.

Objectives and Results Chain

5. The statement of objectives in PBL appraisal or legal documents may be very general, and the evaluator may need to construct the operation's statement of objectives from various sources. The design logic for a PBL operation is often implicit in the policy matrix, the description of the reform measures, or the development policy letter. As with investment loans, the operation's key performance indicators may be used (with caution) to infer objectives or elements of the results chain.

Prior Actions

6. "Prior actions" or "triggers" in PBLs correspond to the inputs and outputs in the program's results chain. Applying the principle that evaluations focus on outcomes rather than inputs and outputs means that prior actions taken and triggers met do not by themselves provide sufficient evidence of achievement of program objectives. In assessing Effectiveness, PBL evaluations provide evidence on the achievement of intermediate and final intended outcomes, i.e., changes in the policy and institutional environment (rules of the game and incentives) and the resulting changes in the intended social, environmental, and human development outcomes contained in the program's statement of objectives.

Other Topics

7. PBL evaluations cover the following topics, either as part of the assessment of the six core criteria (Relevance, Effectiveness, Efficiency, Sustainability, IFI Performance, and Borrower Performance) or in other sections of the report:

- <u>Macroeconomic stability:</u> A PBL is evaluated with respect to its contribution to improved macroeconomic balances, whether or not this is stated in the operation's statement of objectives. This would include an assessment of the reasonableness of the macroeconomic assumptions and program and the performance of the authorities in correcting macroeconomic imbalances. It would also include any unintended outcomes.
- <u>Political economy:</u> PBL evaluations examine the degree to which the political economy of decision making was conducive to reform. Included in the assessment are the process of policy decision making, the role of reform champions, the likely winners and losers in the reform process, and the incentives facing those affected by reform. In addition, (i) PBL evaluations assess the Government's commitment to reform. A variety of methods can be used: leadership analysis, stakeholder analysis, institutional assessment, and reform readiness analysis; and (ii) PBL evaluations assess the adequacy of political support for reform, the degree to which reform objectives and likely effects were communicated to the public, and the extent to which PBL design reinforced national ownership.
- <u>Complementary reforms:</u> PBL evaluations include a thorough evaluation of the extent to which PBL outputs (the implementation of reforms) were achieved. Evaluators assess not only the extent to which inputs were delivered (i.e., agreed-upon reforms took place), but also the degree to which complementary measures necessary for

their implementation occurred. These may include changes in legislation, regulation, public awareness, and institutional arrangements.

- Adjustment costs: For PBLs that earmark resources to defray budget or adjustment costs, the PBL evaluation assesses the extent to which adjustment costs materialized, and the extent to which PBL-provided resources were sufficient to meet these obligations.
- Institutional development: The extent to which a PBL and its associated TA contributed to fostering institutional development is covered in PBL evaluations. This can be evaluated in terms of whether or not improved governance practices or improved skills, procedures, incentives, structures, or institutional mechanisms came into effect. Evaluating the contributions made by a PBL to building the capacity to lead and manage the policy reform process is also important.
- Impact on the poor and other specific groups: PBL evaluations need to assess whether or not a reform operation could have been more pro-poor in its design and implementation. Intended and unintended socioeconomic impacts on the poor and other specific and /or targeted beneficiary groups are assessed. The adequacy of measures planned at appraisal to protect the welfare of vulnerable groups during the adjustment process is analyzed.

IV. GPS on Evaluation of Private Sector Operations

Contents

Background

26. **Formulation Process.** The Evaluation Cooperation Group's Good Practice Standards for Evaluation of Private Sector Investment Operations (ECG-GPS) were originally formulated in response to the call for harmonization of evaluation methodologies by the Development Committee Task Force in 1996. In 2001, the ECG issued the first edition of the GPS, followed by second and third editions in 2003 and 2006 respectively. Each subsequent edition was informed by the findings and recommendations of a benchmarking exercise, which assessed members' practices against the GPS. Following the 2010 benchmarking exercise against the third edition of the GPS, the ECG has issued a fourth edition.

27. The fourth edition builds and improves upon the previous GPS to reflect the evolution in evaluation practices and in the scope of investment operations undertaken by International Financial Institutions (IFIs). It takes into account feedback from the benchmarking exercise, and the experiences of members in implementing the previous set of standards. In particular, the fourth edition addresses the following issues identified with the GPS Third Edition:

- some standards were too narrowly defined and did not recognize the variation in IFI mandates and operational procedures, particularly given the expanding membership of the ECG;
- the GPS did not differentiate between standards that could be implemented unilaterally by the evaluation departments and those that relied in part or wholly on management action or cooperation;
- some standards – the experimental standards in particular – were overly complicated both in design and in their implementation;
- there was unnecessary overlap between different standards, which created ambiguity and/or duplication;
- the GPS comprised a range of different types of standard (harmonization, other, good practice, best practice, and not universally applicable), which complicated both interpretation and periodic benchmarking;
- the definition of rating benchmarks was in certain cases too complex and went beyond the objective of harmonization;
- the GPS were narrowly focused on financial and non-financial sector investments (mainly project finance) and had limited relevance to the increased variety in investment operations (e.g., equity funds, working capital facilities, corporate finance, trade finance, political risk insurance etc.); and
- there was scope for the GPS to promote more innovation in evaluation and dissemination through, for example, web-based media.

28. **Objectives and Organization**. Although the premise of the Task Force's 1996 decree was to enable comparability of results, the guidance stressed other objectives including identifying and disseminating best practices in evaluation, sharing lessons from evaluations, and describing results in a common language. The decree also acknowledged that harmonization efforts should take into account the differing circumstances of each institution. The GPS Fourth Edition is responsive to these wider objectives.

29. These GPS are organized into generic standards on reporting and corporate learning, evaluation guidance and rating systems that apply across all IFIs, and standards specific to IFIs

supporting private sector investment.[44] The standards themselves are formulated as EPs. Each EP is defined in terms of its key components or “Elements”. To guide IFIs in their efforts towards meeting the Elements of the EPs, each EP is supported by a set of standard OPs. The OPs describe the policies and procedures that the IFI would typically need to adopt in order to be deemed compliant with the respective EP.

30. The EPs on evaluation of private sector operations comprise a total of 17 standards and 75 elements. The OPs include 63 policies and procedures.[45] The summary of the EPs, in terms of standards and elements, and OPs is presented below in a simplified and more logically consistent framework, as follows:

Summary of Standards and Elements on EPs and Number of OPs on Evaluation of Private Sector Operations

Evaluation Principles		No. of OPs	Page
Standards	**Elements**		
Generic Principles: Reporting and corporate learning Concern the scope and timing of periodic reporting, the disclosure of evaluation reports and products, and the capture, dissemination and application of lessons from the evaluation system.			
1. Annual reporting of corporate results	*A. Corporate Results* *B. Reporting Rating Results* *C. Analysis* *D. Recommendations* *E. Disclosure*	5	66
2. Periodic reporting on evaluation systems	*A. Periodic Reporting* *B. Quality Efficacy* *C. Alignment* *D. Evaluability* *E. Lessons Application*	5	66
3. Lessons and findings from evaluation	*A. Coverage* *B. Relevance* *C. Accessibility*	3	67
Generic Principles: Evaluation guidance and rating systems Concern the features of an evaluation rating system, and the preparation and dissemination of evaluation guidance.			
4. Guidance for project evaluation	*A. Preparation* *B. Content* *C. Dissemination*	2	68
5. Performance rating scales	*A. Range and Balance*	3	68

[44] The generic standards on independence of evaluation departments were excluded from the GPS on Evaluation of Private Sector Operations as these already constitute Chapter II of Big Book which contain GPS on Independence of IFIs' CED (June 2010). The same treatment was done on specific standards on indirect evaluation (otherwise referred to as self-evaluation) which are presented in Chapter VI on Self-Evaluation. Relatedly, GPS on Evaluation of Private Sector Operations use the terms “direct and indirect” evaluation to refer to self- and independent evaluation, respectively, to acknowledge the differences in terminology used in the different IFIs which reflect the nature of CED's interface with an actual project (based on clarification from consultant of ECG WGPSE.).

[45] GPS on Evaluation of Private Sector Operations, Fourth Edition is a refined and simplified version of GPS 3rd Edition where multi-faceted standards were disaggregated into their individual components resulting in a total of 22 EPs (or standards), each defined by between three and eight elements, making 101 Elements in total. The EPs and their Elements together cover much the same scope as GPS 3rd Edition.

Evaluation Principles		No. of OPs	Page
Standards	**Elements**		
	B. Descriptive *C. Binary Reporting*		
Private Sector Principles: **Planning and executing a project evaluation program** Concern the determination of when an operation is ready for evaluation, the selection of a sample from a defined population, and the process of direct evaluation and self-evaluation with independent verification.			
6. Defining the population of projects for evaluation	*A. Coherence and Objectiveness* *B. Qualifying Projects* *C. Screening* *D. Non-Qualifying Projects* *E. Exclusions* *F. Disclosure*	6	69
7. Selecting a sample of projects for evaluation	*A. Representative Random Sampling* *B. Sample Aggregation* *C. Disclosure* *D. Purposeful Sampling*	4	71
8. Process of direct evaluation by CED	*A. CED's Options* *B. Reporting* *C. Desk-Based* *D. In-Depth* *E. Transparency* *F. Review Process*	4	72
9. Scope of Independent verification by the CED	*A. Verification* *B. In-Depth Verification* *C. CED Reporting* *D. Review Process*	3	73
Private Sector Principles: Evaluation metrics and benchmarks Define the scope of measurement and benchmarks for rating each performance indicator within the evaluation framework.			
10. Rating project outcomes	*A. Synthesis rating* *B. Benchmark* *C. Financial Criteria* *D. Economic Criteria* *E. IFI Mandate Criteria* *F. E&S Criteria*	2	74
11. Outcome Indicator 1 - Financial performance and fulfilment of project business objectives	*A. Stakeholder Analysis* *B. Time Span* *C. Fulfilment of Project Business Objectives* *D. Methods* *E. Benchmarks*	5	76
12. Outcome Indicator 2 – Economic	*A. Stakeholder Analysis*	5	80

Evaluation Principles		No. of OPs	Page
Standards	**Elements**		
sustainability	*B. Time Span* *C. Net Benefits* *D. Methods* *E. Benchmarks*		
13. Outcome Indicator 3 – Contribution to IFI mandate objectives	*A. Method* *B. Balanced* *C. Benchmark*	2	83
14. Outcome Indicator 4 – Environmental and social performance	*A. E&S Performance* *B. E&S Capacity* *C. Sub-Project Performance* *D. Benchmark*	3	84
15. Rating the IFI's investment profitability	*A. Scope* *B. Net Method* *C. Gross Proxy Method* *D. Benchmark*	5	86
16. Rating IFI work quality / bank handling	*A. Scope* *B. Stand-alone* *C. Pre-Commitment* *D. Post-Commitment* *E. Benchmark*	4	87
17. Rating the IFI additionality	*A. Counterfactual* *B. Financial Additionality* *C. Non-Financial Additionality* *D. Benchmark*	2	90
Total No. of Standards: 17	**Total No. of Elements: 75**	**Total No. of OPs: 63**	

31. Where appropriate, these GPS make reference to a project typology (Annex B.2) so that OPs can be tailored to the wider range of projects and clients now supported by IFIs. The project typology is particularly important when defining early operating maturity, and rating the indicators of project business success, economic sustainability, and environmental and social performance. For projects exhibiting a mix of project types, it is recommended that IFIs use a range of metrics as appropriate.

GPS on Evaluation of Private Sector Operations[46]

Generic Principles: Reporting and corporate learning

Evaluation Principle (Standards and Elements)	Standard Operational Practices	Element Link	Notes
Annual reporting of corporate results: A. *Corporate Results:* The CED reports to the Board annually on the IFI's independently verified outcome results. B. *Reporting Rating Results:* The CED reports the IFI's results in all rating dimensions and indicators. C. *Analysis:* The CED analyses the results to discern performance drivers. D. *Recommendations:* The CED formulates recommendations based on the findings. E. *Disclosure:* The CED discloses its synthesis evaluation results externally.	On an annual basis, the CED reports to the governing Board on the IFI's corporate-wide performance, based on the findings from project-level evaluations and, if required, thematic evaluations.* The report can be stand-alone or incorporated in other reports to the IFI's Board. The ratings reported should be those independently verified or directly assigned by the CED.	*Corporate Results*	* Thematic evaluations could include country, sector or other studies of the aggregate results across a defined group of projects.
	For each rating dimension and indicator, the CED reports the number and proportion (by number of operations) of the evaluated cohort in each performance-rating category.	*Reporting Rating Results*	The CED may also choose to report the results weighted by project or investment size, to indicate the quantum of impact.
	The CED provides a synthesis description of the ratings patterns and their cross-cutting performance drivers under each indicator. It also provides the dimension and indicator ratings for the previous few years or cohorts thereof (where such data exists) to show how performance is evolving over time.	*Analysis*	
	Where feasible, the CED makes recommendations to the IFI's Management based on the evaluation findings.	*Recommendations*	The qualification here allows for an exception in the case of CEDs with insufficient evaluated projects to substantiate recommendations.
	The CED publishes its findings after appropriate redaction to protect commercial confidentiality, and posts on a webpage accessible via the IFI's external website the full text or an abstract of its report that accurately summarizes its essential findings.	*Disclosure*	The webpage can be on the CED's own site, provided that the CED's site can be accessed via a link on the IFI's main pages.
Periodic reporting on evaluation systems: A. *Periodic Reporting:* At least once every three years, the CED reports on aspects of the IFI's evaluation systems, including: B. *Quality & Efficacy:* The CED reports to the Board on the quality and efficacy of evaluation systems.	The CED reports to the Board at least once every three years on the functioning and effectiveness of the IFI's evaluation systems, as detailed below. The report can be stand-alone or incorporated in other reports to the Board.	*Periodic Reporting*	The review of evaluation systems in the IFI could be undertaken by the CED directly, or by an external independent body under commission from the CED.
	The CED reviews and reports on the quality and efficacy of the IFI's evaluation systems. As part of this reporting, the CED submits to the IFI's Management and Board the periodic benchmarking reviews of the consistency of the IFI's practices	*Quality & Efficacy*	

[46] See footnote #44.

Generic Principles: Reporting and corporate learning

Evaluation Principle (Standards and Elements)	Standard Operational Practices	Element Link	Notes
C. *Alignment:* The CED reviews and reports on the alignment of Management reporting systems with the evaluation framework. D. *Evaluability:* The CED reviews and reports on the evaluability of the IFI's operations. E. *Lessons Application:* The CED reviews and reports on the application of lessons learned from evaluation.	with the ECG Good Practice Standards (or provides a summary thereof).		
	The CED reviews and reports the extent to which internal Management and corporate reports (up to Board level) are broadly aligned with the evaluative framework. For example, the CED should review: (i) to what extent the IFI applies coherent and consistent benchmarks to gauge project performance at relevant stages of the project cycle; and (ii) whether Management's reporting of results includes project outcome and additionality ratings based on the ECG GPS.	*Alignment*	
	The CED assesses and reports on the evaluability of the IFI's operations i.e., the extent to which the value generated or the expected results of a project are verifiable in a reliable and credible fashion. In practical terms, the CED should assess whether the IFI had specified relevant indicators at approval and made sufficient provision to collect the data required for monitoring during project supervision. The CED need not report on every operation, or undertake such reviews at the time of project approval.	*Evaluability*	
	The CED assesses and reports evidence of the extent to which lessons of experience are being applied in new operations. It is not required that the CED report on every operation individually, or undertake such reviews at the time of project approval.	*Lessons Application*	Examples of methodologies for such an assessment include surveys or interviews of origination staff, or a CED review of appraisal documents.
Lessons and findings from evaluation: A. *Coverage:* Lessons of experience are identified for all project-level evaluations. B. *Relevance:* Lessons are relevant to new operations. C. *Accessibility:* Lessons and evaluation findings are made readily available to IFI staff.	All direct and indirect project-level evaluation reports should contain a prompt or template for the author(s) to identify and articulate one or more lessons from the operation.	*1.*	
	Lessons should be concise, prescriptive, and placed in the context of a material issue that was encountered in the evaluation so that its relevance to new operations can be determined easily, on a stand-alone basis. The point of view and selectivity should focus on what the IFI might have done to obtain better results from the operation.	*Relevance*	
	The CED maintains a database or library of operational lessons from project-level evaluation reports, which is freely accessible to IFI staff. Alternatively, the CED contributes lessons from project-level evaluations (or a summary thereof) to a	*Accessibility*	

Generic Principles: Reporting and corporate learning

Evaluation Principle (Standards and Elements)	Standard Operational Practices	Element Link	Notes
	database maintained by IFI Management. The CED makes available to IFI staff a range of easily accessible dissemination products covering evaluation findings from projects and/or synthesis CED reports. This could include, *inter alia*, access to the full reports, electronic notification of new items, and presentations of findings.		

Generic Principles: Evaluation guidance and rating systems

Evaluation Principle (Standards and Elements)	Standard Operational Practices	Linked to	Notes
Guidance for project evaluation: A. *Preparation:* The CED develops guidance for staff undertaking direct and indirect project evaluations. B. *Content:* Guidance is self-standing, current, and comprehensive in key aspects of the evaluation process. C. *Dissemination:* Guidance is easily accessible and supplemented by training and/or good practice examples.	The CED develops, in conjunction with Management as necessary, guidance for CED and operational staff undertaking direct and indirect project evaluations. The evaluation guidelines should be consistent with prevailing ECG Good Practice Standards and at a minimum include: (i) the key steps in the evaluation process, in the preparation and signing-off of reports, and in independent verification by the CED as necessary; (ii) the scope of measurement and the benchmarks for assigning ratings for each performance indicator and dimension; and (iii) standard reporting templates that include a performance ratings matrix.	*Preparation* *Content*	Where separate guidance is prepared for self-evaluations and independent direct evaluations and/or verifications, these should be completely coherent in terms of the prescribed metrics and benchmarks.
	The CED makes the evaluation guidelines and supporting information readily available on its website and/or the IFI's website in respect of guidance for self-evaluation. The CED undertakes dissemination activities to familiarize staff preparing project evaluations with the requirements and supporting documentation. This may include the showcasing of evaluation reports regarded as good-practice examples.	*Dissemination*	
Performance rating scales: A. *Range & Balance:* Each indicator is rated on a performance scale from most negative to most positive, with the scale balanced between positive and negative ratings. B. *Descriptive:* Each rating category accurately	The rating scale for each indicator should encompass performance ranging from the most negative to most positive. 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; or if there are six ratings, three are less than good and three are good or better).	*Range & Balance*	
	The words used to describe these ratings should accurately reflect whether the judgments are less than good or else good or better, and should clearly reflect the	*Descriptive*	

Generic Principles: Evaluation guidance and rating systems

Evaluation Principle (Standards and Elements)	Standard Operational Practices	Linked to	Notes
describes the extent of positive or negative performance. C. *Binary Reporting:* Binary ratings use the first positive rating within the performance scale as their benchmark.	graduation from worst to best. For example: Four-point scale: *unsatisfactory, partly (un)satisfactory, satisfactory, excellent*; or *unsuccessful; partly (un)successful; successful; highly successful*. Six-point scale: *highly unsuccessful, unsuccessful, mostly unsuccessful; mostly successful; successful; highly successful*; or *highly unsatisfactory; unsatisfactory; marginal; satisfactory; good; excellent*.		
	Where the CED reports success rates based on a binary simplification of the rating scale, the binary benchmark should be the first positive rating within the chosen scale i.e., a *satisfactory* or *successful* rating (in the case of the four-point scale cited above) or a *mostly successful* or *satisfactory* rating (in the case of the six-point scale cited above).	*Binary Reporting*	

Private Sector Principles: Planning and executing a project evaluation program

Evaluation Principle (Standards and Elements)	Standard Operational Practices	Linked to	Notes
Defining the population of projects for evaluation: A. *Coherence & Objectiveness:* All projects in the population share common characteristics based on a coherent set of criteria. B. *Qualifying Projects:* The population includes all projects that have reached early operating maturity (or are unlikely ever to do so), and all closed projects. C. *Screening:* The CED determines projects' early operating maturity according to GPS criteria. D. *Non-Qualifying Projects:* Projects that are not operationally mature are reconsidered in subsequent years. E. *Exclusions:* The population may exclude other classes of projects where the CED determines that	The CED defines the population of projects according to a coherent and objective set of criteria appropriate to the type of report. The full application of these criteria will determine whether or not an operation is to be included in the population. For example: Corporate reporting: the population should comprise projects with the same year (or defined range of years) of origin, based on the approval, commitment or disbursement date for the IFI's associated investment. Alternatively, the population should comprise projects that have reached early operating maturity within a defined timeframe.* Part-portfolio reporting: the population should comprise projects sharing a common time-basis (as above), and the same country, region, sector or other thematic characteristics as desired.	*Coherence & Objectiveness*	While this OP allows CEDs to define a sub-portfolio within the overall corporate portfolio, EP6 continues to call for annual reporting of corporate-level results. The onus therefore rests with the CED to determine how best to comply with EP6 should it opt for sub-portfolio sampling in respect of this OP. * This alternative approach is appropriate for CEDs that screen the entire portfolio every year to determine which projects have reached early operating maturity.
	Projects should be included in a designated population only once and only at such time as (but not necessarily as soon as) they have reached early operating	*Qualifying Projects*	* This does not prohibit projects from being included in different

Private Sector Principles: Planning and executing a project evaluation program

Evaluation Principle (Standards and Elements)	Standard Operational Practices	Linked to	Notes
individual evaluations have limited utility. F. *Disclosure:* The CED discloses its criteria for defining the population and any excluded class of project.	maturity.* The population (the boundaries of which are defined under OP11.1) also includes all closed projects (i.e., where the associated IFI investment has been repaid, sold or written off, or the guarantee has been cancelled) regardless of whether or not they had reached early operating maturity by the time of closing. The population should also include projects that are deemed unlikely ever to achieve early operating maturity.**		populations relating to different studies, for example if the CED was to undertake both a corporate and country-level evaluation. ** This could include, for example, a project that has failed or stagnated such that it is unlikely ever to establish a trading record, yet the IFI's investment has not been sold, cancelled or written off and so has not been officially closed. In such cases, there is little value in postponing evaluation, hence they should be included in the current population.
	The CED establishes which projects have reached early operating maturity, taking into consideration information on project status provided by operational departments and by applying the guidance in *Annex B.3 : Lookup Table for Determining Early Operating Maturity*.	*Screening*	
	Where the CED determines that projects have not yet reached early operating maturity (but are likely to do so in the future), they should be omitted from the current evaluation year's population. Instead, the CED should consider them for inclusion in the population in a future year when they will have reached early operating maturity. In cases where the IFI is involved in litigation, foreclosure or other legal process where evaluation could prejudice the IFI's legal position, the CED may choose to omit these projects from the current population and instead roll them forward for consideration in a future year.	*Non-Qualifying Projects*	Rolling projects forward for consideration in future years' populations may render them incongruous with the year(s) of origin of those later populations. Where the CED reports an aggregation of several years of consecutive evaluation findings, such projects might legitimately be included. Otherwise, the CED should disclose the incidence of older projects in the reported population, or report their results separately.
	The CED may choose to exclude altogether from the population the following classes of project: (i) those that did not proceed with IFI support and where the associated IFI investments were dropped or guarantees never signed, activated or utilized*; or (ii) those involving subscribed rights offerings or follow-up investments /	*Exclusions*	* CED may exclude projects where the IFI never incurred any exposure under the guarantee (e.g., because the beneficiary of the guarantee never made any advances to the end client).

Private Sector Principles: Planning and executing a project evaluation program

Evaluation Principle (Standards and Elements)	Standard Operational Practices	Linked to	Notes
	guarantees undertaken for substantially the same purpose as before (e.g., to help finance cost overruns or restructurings).**		** Repeat investments in an existing client company should normally be included in the population, unless they meet the criteria in (6).
	The CED discloses how it defined the population and its criteria for including or excluding projects in line with the operational practices above.	*Disclosure*	
Selecting a sample of projects for evaluation: A. *Representative Random Sampling:* Either all projects in the population are evaluated or the CED selects a random sample whose characteristics are representative of the population. B. *Sample Aggregation:* The CED reports the results of one or more years of evaluated random samples. C. *Disclosure:* The CED discloses its sampling methodology, how it defined the reported cohort, and sampling errors in reported results. D. *Purposeful Sampling:* The CED may self-select a purposeful sample to serve specific evaluative needs, but not for overall corporate reporting purposes.	If evaluation coverage is less than 100%, the CED should select a random sample of projects for evaluation from the established population. The sample should be as representative as practicable insofar as it reflects the distribution of important characteristics throughout the population as relevant to each institution.*	*Representative Random Sampling*	* Relevant characteristics for testing the representativeness of the sample could include: industry sector; country; region; project size; investment size; IFI instrument of support; incidence of loan impairment or equity write-down.
	For reporting purposes, the CED may report the results of a sample of projects evaluated in one year, or use a cohort comprising the evaluated samples from several consecutive years in order to increase the granularity of data and its statistical significance. In deciding how many years of data to combine, the CED should balance the desire to report on a meaningful number of observations against the currency of findings, particularly if using evaluation data more than three years old.	*Sample Aggregation*	
	Where sampling is used, the CED should report details of the sampling technique used and the extent to which the sample's characteristics reflect those of the population. When reporting the aggregate results of a cohort comprising samples from more than one evaluation year, the CED should disclose how the reported cohort is defined. The CED should calculate and disclose the sampling errors (at the 95% confidence interval) in the reported success rates for each of the evaluated indicators and outcome ratings.*	*Disclosure*	* Disclosure of sampling errors enable observers to judge the relevance, usefulness and comparability of success rates reported by different institutions.
	The CED may select a purposeful (self-selected) sample of projects to be evaluated.* The CED should not use the results of purposeful evaluations for	*Purposeful Sampling*	* Reasons for selecting a purposeful sample could include:

Private Sector Principles: Planning and executing a project evaluation program

Evaluation Principle (Standards and Elements)	Standard Operational Practices	Linked to	Notes
	overall corporate reporting purposes, unless projects in the purposeful sample are also selected as part of a random sample as determined under OP 7.1 .		the potential for learning; the high profile of an operation; credit and other risks; the sector is a new one for the IFI; the likelihood of replication; or the desirability of balanced country and sector coverage.
Process of direct evaluation by the CED: A. *CED's Options:* At its own discretion, the CED can select projects on which to conduct its own direct evaluations. B. *Reporting:* The CED conveys its findings in a Project Evaluation Report (PER). C. *Desk-Based:* As a minimum, the PER is based on internal IFI data, staff consultations and market research. D. *In-Depth:* For selected projects, the CED conducts on-the-ground research and stakeholder consultations. E. *Transparency:* The basis for the CED's findings are fully transparent in the PER, including financial / economic calculations and environmental and social effects. F. *Review Process:* Management and staff have the opportunity to comment on the draft PER, but the final assessment is determined solely by the CED.	The CED can undertake a direct evaluation of a project on its own volition, acting with consideration to on-going legal process in line with OP 6.4 . The scope of evaluation and indicator ratings should be consistent with the GPS. The CED reports its findings in a Project Evaluation Report (PER).	*CED's Options* *Reporting*	
	As a minimum, the research for PERs draws from a file review, discussions with available staff involved with the operation since its inception, and external market research. On a more rigorous basis, the CED may choose to conduct in-depth research (in the field as necessary) for the PER, based on consultations with stakeholders who are knowledgeable about the country, company and project.*	*Desk-Based* *In-Depth*	* Such stakeholders could include: IFI specialists, the company's management, employees, auditors, suppliers, customers, competitors, bankers, any relevant government officials, industry associations, community representatives and local NGOs.
	The basis for the CED's findings and ratings are made fully transparent in the PER. The PER should also cite which stakeholder groups were consulted as part of the process. Where ex-post financial and/or economic rates of return for the project are cited in the PER, the document includes an attachment providing details supporting these calculations such as the key assumptions and underlying financial / economic time-series data. The PER includes a summary of environmental, worker health and safety, and social performance information, for each of the IFI's environmental and social safeguards that apply to the project. Evidence from on-the-ground observations and/or client reporting should be sufficient to support the assigned outcome and IFI work quality ratings. The information can be incorporated as an attachment to the PER if preferred.	*Transparency*	
	The CED provides an opportunity to Management and operational staff to review and comment on the PER's draft findings, though the final content and ratings in	*Review Process*	

Private Sector Principles: Planning and executing a project evaluation program

Evaluation Principle (Standards and Elements)	Standard Operational Practices	Linked to	Notes
	the report remain the decision of the CED. Findings from the PER can be used in synthesis reporting without further verification.		
Scope of independent verification by the CED (see footnote #44): *A Verification:* The CED conducts an independent review of XASRs based on internal IFI data and independent research. *B. In-Depth Verification:* The CED conducts detailed verifications for selected projects. *C. CED Reporting:* The CED reports its independent findings in an XASR-Assessment (XASR-A), which records any rating differences to those in the XASR. *D. Review Process:* Management and staff have the opportunity to comment on the draft XASR-A, but the final content is determined by the CED.	The CED conducts an independent review (which may be desk-based) of the XASR to verify its scope, responsiveness, evident reliability of the analysis, impartiality and consistency in ratings judgments, and appropriateness and completeness of the identified lessons. As a minimum, the independent review draws from a file review, discussions with available staff involved with the operation since its inception, and external market research. Depending on the coverage of the population by XASRs, either: (a) If the IFI has prepared XASRs for a representative sample selected in accordance with EP 7 , then the CED should conduct independent reviews for all the XASRs in such sample; or (b) If the IFI has prepared XASRs for all projects in the population of operationally mature projects (as defined in accordance with EP 6), then the CED may choose to conduct independent reviews either for all XASRs or for a representative sample of XASRs selected in accordance with EP 7. If a sample is preferred, only the ratings from CED-verified XASRs are valid for corporate reporting purposes.	*Verification*	
	On a more rigorous basis, the CED conducts detailed reviews on selected XASRs to verify the self-evaluation findings. The CED should have a clear policy for selecting projects for in-depth verification and should implement the policy consistently. Selection criteria might include: poor quality / reliability of the XASR; apparent significant differences between self-evaluation ratings and CED ratings; projects exhibiting performance at the extremes; projects that will contribute to corporate learning; or projects of relevance to corporate strategy or development imperatives more widely. In-depth verifications have the same scope of research as in-depth PERs (per OP 8.2) and where deemed necessary by the CED are conducted through field-based research. The CED discloses its policy for selecting XASRs for in-depth verification along with the number and/or proportion of projects subjected to such a review.	*In-Depth Verification*	Note that this OP does not prescribe the number or proportion of XASRs that should be subjected to in-depth verification. However, in the interests of evaluative rigor, it is desirable that the CED performs some degree of in-depth verification (see Annex Note EP 8 / 9). Depending on the availability of resources, in-depth verification could be focused on a project's environmental and social effects, (which are most reliably determined through field-based research), rather than on the full range of project impacts.
	The CED prepares an XASR-Assessment (XASR-A) on the final-edition XASR that records the CED's findings from its verification and its independent judgments on	*CED Reporting*	

Private Sector Principles: Planning and executing a project evaluation program

Evaluation Principle (Standards and Elements)	Standard Operational Practices	Linked to	Notes
	the project's results and appropriate ratings in relation to GPS guidelines. The XASR-A is shared in draft form with the XASR team and their comments solicited and considered by the CED. For transparency, the final XASR-A should communicate the CED's final independent judgments highlighting any differences between its performance ratings and those of the XASR, and cite the comments received from the XASR team.	*Review Process*	

Private Sector Principles: Evaluation metrics and benchmarks

Evaluation Principle (Standards and Elements)	Standard Operational Practices	Linked to	Notes
Rating project outcome: A. *Synthesis Rating:* The project's outcome is based on a qualitative synthesis of underlying indicator ratings. B. *Benchmark:* The rating measures how well the project serves the IFI's institutional mandate. C. *Financial Criteria:* It reflects the project / company's financial performance and achievement of project business objectives. D. *Economic Criteria:* It reflects the project / company's contribution to economic growth*. E. *IFI Mandate Criteria:* It reflects the project / company's contribution to the IFI's mandate objectives. F. *E&S Criteria:* It reflects the project / company's environmental and social performance.	Scope of Measurement: The rating of project outcome reflects summary qualitative performance judgments based on a synthesis of all the following underlying indicator ratings, taking into consideration the sustainability of results: - the project / company's **financial performance** (i.e., the project's contribution to the company's financial results, or the company's financial results where the project is indistinguishable from the company). This also considers the extent of **fulfilment of project business objectives**; - the project / company's **economic sustainability** (i.e., the project and/or project company's contribution to growth in the economy)*; - the project / company's **contribution to the IFI's mandate objectives**, be they to stimulate development of the private sector, development of efficient financial / capital markets, or transition to a market economy; - the project / company's **environmental and social performance**.	*Synthesis Rating* *Financial Criteria* *Economic Criteria* *IFI Mandate Criteria* *E&S Criteria*	See Annex B.5: Guidance Paper: *Technical Note on IFC's Methodology for Assigning Development Outcome Ratings.* * Not applicable to EBRD.
	Binary Benchmark: As a minimum, for a positive project outcome rating, the project should have a clear preponderance of positive results (i.e., it may exhibit some minor shortcomings though these should be clearly outweighed by positive aspects). The guiding principle should be that if all the IFI's projects exhibited this level of performance, the IFI should be able to demonstrate the successful execution of its institutional mandate.	*Benchmark*	The Binary Benchmarks defined herein refer to the first positive rating within the chosen scale. In the example scales cited in OP 5.2, the binary benchmark therefore refers to that of a satisfactory / successful rating (in the case of a

Private Sector Principles: Evaluation metrics and benchmarks

Evaluation Principle (Standards and Elements)	Standard Operational Practices	Linked to	Notes
			four-point scale) or mostly successful/satisfactory rating (in the case of a six-point scale). See Annex B.5, note OP 10.2 for guidance on an extended rating scale.

Private Sector Principles: Evaluation metrics and benchmarks

Evaluation Principle (Standards and Elements)	Standard Operational Practices	Linked to	Notes
Outcome Indicator 1 – Financial performance and fulfilment of project business objectives: A. *Stakeholder Analysis:* The indicator measures the incremental effect of the project on all key financial stakeholders in the project and/or company. B. *Time Span:* The rating is based on historic and projected future financial performance. C. *Fulfilment of Project Business Objectives:* The rating considers the achievement of process and business objectives articulated at approval. D. *Methods:* The CED applies a range of evaluation methods appropriate to the project type, with an emphasis on quantitative metrics wherever possible. E. *Benchmarks:* The rating is based on benchmarks appropriate to the project type and evaluation methods applied.	In evaluating financial performance, the incremental effect of the project on the company is assessed on a with vs. without project basis, or a before vs. after project basis. The effect of the project on all financial stakeholders in the project and/or company should be considered.* Both historic and, where relevant, projected performance should be taken into consideration. The rating also considers fulfilment of project business objectives, that is the extent to which the project has delivered on the process and business objectives stated at approval.	*Stakeholder Analysis* *Time Span* *Fulfilment of Project Business Objectives*	* Such stakeholders should include as relevant: the owners (shareholders); senior lenders; junior lenders; and trade creditors.
	Scope of Measurement for projects of types A, B and C: The rating of financial performance and fulfilment of project business objectives is determined through the application of the methods set out below. The choice of method should be appropriate to the project type, and should use quantitative metrics wherever possible. At a minimum, methods 3, 4 and 5 should be used. <u>1. Quantitative Method:</u> The rating is based on the project's after-tax financial rate of return in real terms (FRR), or on the time-adjusted after-tax return on invested capital in real terms (ROIC i.e., the costs and benefits to the whole company on a before vs. after basis).* <u>2. Achievement of Appraisal Projections:</u> The evaluation compares actual performance with appraisal projections. This is only valid provided that the appraisal projections demonstrate sufficient profitability to: (i) service the project's debt obligations and meet creditor payments when due; and (ii) generate the minimally acceptable return to the project company's shareholders commensurate with the risk.** <u>3. Achievement of Project Business Objectives:</u> The assessment concerns the extent to which the project has, or is judged likely to, fulfil the process and business goals that were articulated at approval.*** <u>4. Analysis of Financial Statements:</u> An appropriate range of performance indicators in project financing are considered such as: sales, net profit, debt service coverage, and financial internal rate of return (FIRR). Suitable project return analysis should supplement balance sheet and income related indicators. <u>5. Business Prospects:</u> The project company's overall profitability, adaptability and prospects for sustainability and growth are considered, taking into account its performance relative to the market or sector peers.	*Methods*	For further guidance, see Annex B.5, note OP 11.2. * In general, an FRR should be calculated where the financial cashflows of the project can be separated from those of the company's other activities. A ROIC may be more appropriate in the case of corporate investments or expansion projects. ** The "minimally acceptable return" to shareholders can be derived from the IFI's own profit objectives if it is itself a shareholder. Alternatively, indicators could include the original financing plan or current expectations of investors in similar projects. *** For example: business objectives could be those related to carrying out an investment plan in respect of plant and equipment and the establishment of a strong management team; process objectives could be the introduction of an IAS accounting system or for a financial institution the improvement of credit manuals and the training of staff. For EBRD,

Private Sector Principles: Evaluation metrics and benchmarks

Evaluation Principle (Standards and Elements)	Standard Operational Practices	Linked to	Notes
			achievement of project objectives does not incorporate the transition impact objectives, which are captured separately.
	Binary Benchmark for projects of types A, B and C: As a minimum, for a positive rating of the project / company's financial performance and fulfilment of project business objectives, it should achieve the following benchmarks. Where more than one method is applied, each of the relevant benchmarks should be met: 1. Quantitative Method: The project's FRR or ROIC is equal to or greater than the project company's Weighted Average Cost of Capital (WACC). The WACC should be calculated using accepted principles and based on company- or sector-specific data.* The use of fixed IFI-wide assumptions or hurdle rates in place of the WACC is not good practice. 2. Achievement of Appraisal Projections: Actual performance meets or exceeds appraisal projections such that the project has demonstrably met its obligations to lenders and creditors, and has yielded the minimally acceptable return to its shareholders commensurate with the project risk.** 3. Achievement of Project Business Objectives: The project's process and business goals articulated at approval are broadly achieved or are deemed within reach albeit with some risk to their realisation. 4. Analysis of Financial Statements: Performance indicators are in line with appraisal estimates. 5. Business Prospects: The project company's overall profitability and prospects for sustainability and growth are sound, such that it is expected to remain competitive in relation to the market and its sector peers.	*Benchmarks*	See Annex B.5 , note OP 11.3 for guidance on an extended rating scale. * See guidance paper: *Using the FRR to Rate Project Business Success.* ** The "minimally acceptable return" to shareholders can be derived from the IFI's own investment outcome rating if it is itself a shareholder. Alternatively, indicators could include the original financing plan, current expectations of investors in similar projects, or evidence that shareholders are satisfied with their returns (e.g., the company has attracted additional investment and/or executed a successful rights issue to fund future growth).
	Scope of Measurement for intermediation projects of type D1 and D2: The rating of financial performance and fulfilment of project business objectives is determined through the application of the methods set out below. The choice of method	*Methods*	

Private Sector Principles: Evaluation metrics and benchmarks

Evaluation Principle (Standards and Elements)	Standard Operational Practices	Linked to	Notes
	should be appropriate to the project type, and should use quantitative metrics wherever possible. At a minimum, methods 3 and 4 should be used. 1. Performance of Sub-Portfolio: An assessment should be made of the financial impact of the sub-portfolio on the financial intermediary's viability. Where a calculation of the profit contribution of the sub-portfolio is not possible, proxies can be used, for example: sub-loan spreads (relative to the rest of the FI portfolio), FI sub-loan risk ratings, and/or incidence of arrears or write-offs among the sub-loans.* 2. Performance of Fund Portfolio: The rating is based on the project portfolio's projected or realized contribution to the fund's net return on equity (RoE) or net IRR to the investors (i.e., after management fees, carried interest and other administrative costs). 3. Achievement of Project Business Objectives: The assessment concerns the extent to which the project has, or is judged likely to, fulfil the process and business goals that were articulated at approval.** In particular, it should consider the project's success in reaching certain sub-borrower or investee groups if such groups were specified as targets at approval. 4. Performance of Intermediary: The financial intermediary / local fund management company's overall profitability, adaptability and prospects for sustainability and growth are considered, taking into account its performance relative to the market or sector peers.		* See guidance paper: *Additional Business Indicators for Financial Intermediaries.* ** For EBRD, achievement of project objectives does not incorporate the transition impact objectives, which are captured separately.
	Binary Benchmark for projects of types D1 and D2: As a minimum, for a positive rating of the project / company's financial performance and fulfilment of project business objectives, it should achieve the following benchmarks. Where more than one method is applied, each of the relevant benchmarks should be met: 1. Performance of Sub-Portfolio: There is adequate evidence (quantitative or qualitative) that the sub-portfolio has had a positive effect on the financial intermediary's profitability, and helped improve its viability. 2. Performance of Fund Portfolio: The projected or realized net return on equity (RoE) or net IRR to the fund's investors is equal to or greater than the fund's weighted average cost of capital (FWACC)*. 3. Achievement of Project Business Objectives: The project's process and business goals articulated at approval are broadly achieved or are deemed within reach with some risk to their realisation. The intermediary has succeeded in reaching sub-borrowers or investee groups that were specified as targets at approval.**	*Benchmarks*	See Annex B.5~~,~~ note OP 11.5 for guidance on an extended rating scale. * Annex B.5, note OP 11.5 demonstrates how the FWACC is estimated for multi-country funds, using a combined project and equity risk premium of 600bpts over the cost of debt for the fund. Alternatively, the CED can establish its own RoE benchmark, provided that it too reflects an appropriate equity risk premium over the cost of debt to the fund.

Private Sector Principles: Evaluation metrics and benchmarks

Evaluation Principle (Standards and Elements)	Standard Operational Practices	Linked to	Notes
	4. Performance of Intermediary: The intermediary's overall profitability, adaptability and prospects for sustainability and growth are sound, such that it is expected to remain competitive in relation to the market and its sector peers.		** Since project type D1 concerns credit lines designed to target specific groups of sub-borrower (rather than a more general corporate investment in a financial intermediary – project type B), the IFI should make adequate provision to track at a minimum the broad sector groups reached through the intermediation. In the absence of such information, the CED may choose to assign a rating of No Opinion Possible for the project's business success.

Private Sector Principles: Evaluation metrics and benchmarks

Evaluation Principle (Standards and Elements)	Standard Operational Practices	Linked to	Notes
Outcome Indicator 2 – Economic sustainability: A. *Stakeholder Analysis:* The indicator measures the incremental effect on all key economic stakeholders in the project. B. *Time Span:* The rating is based on historic and projected economic effects. C. *Net Benefits:* The rating considers both benefits <u>and</u> costs associated with the project, including economic distortions. D. *Methods:* The CED applies a range of evaluation methods appropriate to the project type, with an emphasis on quantitative metrics wherever possible. E. *Benchmarks:* The rating is based on benchmarks appropriate to the project type and evaluation methods applied.	In evaluating the project's economic sustainability (i.e., the project and/or project company's contribution to growth in the economy), the incremental effect of the project on stakeholders is assessed on a with vs. without-project basis, or before vs. after-project basis. Both historic and, where relevant, projected economic effects should be taken into consideration. The effect of the project on all key economic stakeholders (including and beyond the project company's owners and financiers) should be considered.* Economic distortions conveying trade protectionism should also be considered, for example: quotas; administrative barriers; import / export restrictions, tariffs or subsidies; anti-dumping legislation; exchange-rate manipulation; or protectionist use of patent systems.	*Time Span* *Stakeholder Analysis* *Net Benefits*	This EP is not relevant for EBRD. * Such stakeholder should include, as relevant: customers; suppliers; producers of complementary goods; competitors; new market entrants; employees; tax-payers (government); and neighbours. This EP does not prescribe the methodology by which the CED should measure economic impacts on different stakeholders or how to verify the attribution of economic effects to the project. In practice, CEDs may wish to examine this in the form of a synthesis study or by assessing the aggregate effects of a group of related projects (e.g., in the same sector or geographic area). See also guidance paper*:* *<u>A Stakeholder Framework for Assessing Development Impact.</u>*
	Scope of Measurement for projects of types A, B and C: The rating of economic sustainability is determined through the application of the following methods. The choice of method should be appropriate to the project type, and should use quantitative metrics wherever possible: <u>1. Quantitative Method:</u> The rating is based on the project's net quantifiable economic benefits and costs, as measured by the project's real economic rate of return (ERR) or by the economic return on invested capital (EROIC) i.e., by the time-adjusted internal rate of return on the economic costs and benefits on a before-vs-after basis.* The analysis should also consider other material, but unquantifiable, costs and benefits to key economic stakeholders. <u>2. Qualitative Stakeholder Analysis:</u> Where quantified estimates of the direct economic costs and benefits to all relevant economic stakeholders are not possible, each economic stakeholder group affected by the project should be identified and a judgment made broadly as to the magnitude and direction (positive or negative) of the impact on each.	*Methods*	* In general, an ERR should be calculated where the economic effects (cashflows) of the project can be differentiated from those of the company's other activities. An EROIC may be more appropriate in the case of corporate investments or expansion projects.

Private Sector Principles: Evaluation metrics and benchmarks

Evaluation Principle (Standards and Elements)	Standard Operational Practices	Linked to	Notes
	Binary Benchmark for projects of types A, B and C: As a minimum, for a positive rating of the project's economic sustainability, it should achieve the following benchmarks: 1. Quantitative Method: The ERR or EROIC is equal to or greater than the larger of either: (i) a multiple of 1.2 times the project company WACC*; or (ii) 10%. A positive rating may also be awarded if the ERR or EROIC falls short of the quantitative benchmark, but there are other material un-quantified net economic benefits that could be expected to raise the ERR or EROIC sufficiently. 2. Qualitative Stakeholder Analysis: Either: (a) the project meets the minimum standard for satisfactory **financial performance**** and there is evidence that it has generated a balance of benefits for its wider economic stakeholders (i.e., those other than the project company's owners and financiers); or (b) the project just fails to meet the minimum standard for satisfactory **financial performance**, but there is evidence that it has generated substantial net benefits for its wider economic stakeholders. In either case, the project should not rely on economic distortions to maintain its **financial performance**.	*Benchmarks*	See Annex B.5~~,~~ note OP 12.3 for guidance on an extended rating scale. * The project company WACC should be calculated using accepted principles per OP 11.3. ** The reference here to project financial performance does not imply overlap between the two indicators, although the two are by definition linked. Financial performance is a measure of the project's impact on its financial stakeholders, who represent a sub-set of all of the project's economic stakeholders. It is therefore a starting point for assessing the project's overall economic impact. However, the metrics here consider stakeholders beyond the project company's owners and financiers and, therefore, the wider economic contributions of the project beyond those measured by project financial performance alone.
	Scope of Measurement for projects of types D1 and D2: The rating of economic sustainability is determined through the application of either method D1 or D2 accordingly: D1. Economic Activities of Sub-Borrowers: The rating is based on the economic activities of the sub-borrowers as the principal stakeholder group (i.e., customers of the financial intermediary). If a quantitative assessment of the net economic benefits generated by sub-borrowers is not possible, the analysis should consider the markets supported specifically by the project and/or more generally by the financial intermediary along with evidence of increased or decreased economic activity in these markets. The existence of economic distortions in these markets should also be considered.* D2. Economic Viability of Fund Investees:** The rating is based on the economic viability of the fund's investee companies, a proxy for which is their individual and	*Methods*	* The criteria by which market sectors are deemed economically viable is left to the judgment of CED. Examples of the types of data that would support such an assessment would include country- and/or sector-level data on productivity, growth and competitiveness, and the existence of subsidy or other state support, and other macroeconomic factors that have affected the project's sustainability or could do so in the future. ** In rating the economic

Private Sector Principles: Evaluation metrics and benchmarks

Evaluation Principle (Standards and Elements)	Standard Operational Practices	Linked to	Notes
	combined contribution to the fund's gross return (before management fees). The extent to which the commercial performance of the fund and its investee companies is influenced by economic distortions should also be considered.		sustainability of a fund, the CED should look through the fund to the investee companies and assess their underlying economic viability. Economic rate of return calculations might be possible at the investee level. Otherwise it is possible to infer that if the investees have generated positive equity returns for the fund and are operating in competitive and non-subsidized markets, they are likely also to have generated positive economic contributions for their wider stakeholders.
	Binary Benchmark for projects of types D1 and D2: As a minimum, for a positive rating of the project's economic sustainability, it should achieve the following benchmarks: D1. Economic Activities of Sub-Borrowers:* Both (i) the project has succeeded in reaching targeted groups of sub-borrower; and (ii) there is direct evidence (from sub-portfolio data) that sub-borrowers are economically viable, or indirect evidence (from market data) that market sectors supported by the project and/or more generally by the financial intermediary are economically viable and do not rely on economic distortions to maintain their commercial viability. D2. Economic Viability of Fund Investees: Either (i) the gross equity fund portfolio return (before management fees) is equal to or greater than the FWACC x 1.2; or (ii) at least half of equity fund investees have positive equity returns yet the gross portfolio return (before management fees) is less than FWACC x 1.2 but not less than the FWACC x 0.8.** In either case, there is direct evidence (from sub-portfolio data) that investees are economically viable, or indirect evidence (from market data) that market sectors supported by the project are economically viable and do not rely on economic distortions to maintain their commercial viability.	*Benchmarks*	See Annex B.5, note OP 12.5 for guidance on an extended rating scale. * Since project type D1 concerns credit lines designed to target specific groups of sub-borrower (rather than a more general corporate investment in a financial intermediary – project type B), the IFI should make adequate provision to track at a minimum the broad sector groups reached through the intermediation. In the absence of such information, the CED has the option of assigning a rating of No Opinion Possible for the project's economic development. ** The FWACC should be determined in accordance with OP 11.5.

Private Sector Principles: Evaluation metrics and benchmarks

Evaluation Principle (Standards and Elements)	Standard Operational Practices	Linked to	Notes
Outcome Indicator 3 – Contribution to IFI mandate objectives: A. *Method:* The scope of the indicator and scope of measurement is tailored to reflect the mandate of the IFI. B. *Balanced:* The rating considers both positive and negative contributions. C. *Benchmark:* The rating considers the degree of attribution and the quality of the project's contribution to the IFI's mandate objectives.	Scope of Measurement: This indicator measures the project's contribution to the IFI's mandate objectives,* be they to stimulate development of the private sector, development of efficient financial / capital markets, or transition to a market economy. The scope of measurement should be adjusted to match the scope of the IFI's mandate. It should consider, for example, the positive and negative contributions of the project in the following areas:** Competition; market expansion; private ownership and entrepreneurship; frameworks for markets; transfer and dispersion of skills; demonstration effects; standards for corporate governance and business conduct; development of financial institutions and financial / capital markets; attracting FDI flows; and development of physical infrastructure.***	*Method* *Balanced*	* Mandate objectives as set out in the IFI's Articles of Association or equivalent document. ** It is conceivable that the benefits or costs of the project's effects in these areas have already been quantified in economic terms and reflected in the rating of the project's economic sustainability. Where this is the case, the assessment should avoid double-counting and instead cross-reference the earlier economic calculations. In reality, quantification of the project's attributable value in these areas is unlikely to be possible and so the project's impacts can be considered here in qualitative terms without risk of overlap. *** These components are defined in more detail in Annex B.5 , note OP 13.1.
	Binary Benchmark: As a minimum, for a positive rating of the project's contribution to the IFI's mandate objectives, the project should have: (a) demonstrable effects consistent with the IFI's mandate objectives (for example, in furthering the country's private sector development, development of efficient financial / capital markets, or transition to a market economy); and (b) a clear preponderance of sustainable positive impacts in this respect. A rating of "Neutral" is permitted for this indicator, to account for cases where a project has no observable or attributable impacts (positive or negative) of relevance to the IFI's mandate objectives.*	*Benchmark*	See Annex B.5, note OP 13. for guidance on an extended rating scale. * Note that that a "Neutral" rating is not a middle rating falling between satisfactory and partly (un)satisfactory. Rather, it signifies that this indicator should have no influence on the synthesis project outcome rating. Accordingly, it should also be discounted from both numerator and denominator in the calculation of success rate for this indicator.

Private Sector Principles: Evaluation metrics and benchmarks

Evaluation Principle (Standards and Elements)	Standard Operational Practices	Linked to	Notes
Outcome Indicator 4 – Environmental and social performance: A. *E&S Performance:* The indicator measures the project company or enterprise's overall environmental and social performance. B. *E&S Capacity:* The rating considers the environmental and social management capacity of financial intermediaries / fund managers. C. *Sub-Project Performance:* Where required, the rating considers the environmental and social performance of sub-projects / fund investee companies. D. *Benchmark:* The rating is based on the project company / enterprise / sub-projects achieving compliance with the IFI's specified standards at approval.	Scope of Measurement: The rating of environmental and social performance considers the project company's / enterprise's overall environmental and social performance in the area of influence of the project, as follows: All Project Types: based *primarily* on the IFI's specified standards in effect at approval, and *secondarily* on the IFI's standards prevailing at the time of the evaluation. The assessment is based on the project company's management of its environmental and social aspects, (i.e., the elements of the organization's activities, products or services that can interact with the environment and society) and, to the extent covered by IFI's policies, includes pollution loads, wastes, energy and resource efficiency, biodiversity conservation, workers' and communities' health and safety, public consultation and participation, land acquisition and cultural heritage. Project Types D1/D2: In addition, the assessment should consider the adequacy of the financial intermediary's or fund manager's Environmental & Social Management System (ESMS) and its implementation. If so required by the IFI's specified standards at approval, the environmental and social performance of sub-projects / fund investee companies should also be considered.	*E&S Performance* *E&S Capacity* *Sub-Project Performance*	See also the ISO 14031 standard "Environmental Performance Evaluation" for additional guidance on the scope of measurement. An optional supplementary indicator can be used to measure the extent of progress or regress in environmental and social performance since approval. See Annex Note EP 14 .
	Binary Benchmark: As a minimum, for a positive rating of environmental and social performance, by project type: Project Types A/B/C: The project company / project enterprise should be in material compliance with the IFI's at-approval requirements including implementation of any environmental and social action program (ESAP). Project Types D1/D2: The project company / enterprise has implemented an appropriate Environmental & Social Management System (ESMS), which has been functioning over the project life (as reflected also in environmental and social standards being applied to projects financed by the intermediary). If required by the IFI's specified standards at approval, the environmental and social performance of sub-projects / fund investees are in material compliance with the IFI's requirements.	*Benchmark*	See Annex B.5, note OP 14.2 for guidance on an extended rating scale.
	The project's environmental and social performance should be rated *Not Applicable* where, by virtue of the project's expected lack of environmental and social impacts, the IFI has not prescribed any at-approval environmental and social requirements, and the status of the project at evaluation remains the same. Evidence should be provided to support such a rating. Note, however, that should	*Benchmark*	

Private Sector Principles: Evaluation metrics and benchmarks

Evaluation Principle (Standards and Elements)	Standard Operational Practices	Linked to	Notes
	the project have subsequently changed in scope and given rise to environmental and social impacts, its performance should be rated accordingly against the standards that would have been prescribed had this been known at approval.		

Private Sector Principles: Evaluation metrics and benchmarks

Evaluation Principle (Standards and Elements)	Standard Operational Practices	Linked to	Notes
Rating the IFI's investment profitability: A. *Scope:* The indicator measures the profitability of the IFI's investment(s) in the project. They may be reported separately or synthesized into a single rating. B. *Net Method:* The rating is either based on each investment's net profit contribution… C. *Gross Proxy Method:* …or on the quality of each investment's gross profit contribution. D. *Benchmark:* The rating is based on the investment(s) yielding a return commensurate with the IFI's targeted profitability or return on capital objectives.	Scope of Measurement: The indicator measures the profitability of each of the IFI's investment(s) in the project company. The rating of the IFI's investment profitability is based on either: (a) the investment's net profit contribution (the gross income less financing costs, loss provisions / write-offs, transaction and administrative costs), measured in risk-adjusted, discounted cash flow terms, provided reliable cost data are readily available from management information systems; or (b) the quality of the investment's gross profit contribution (i.e., its likely profitability net of financing costs and loss provisions / write-offs but before deducting transaction and administrative costs).	*Scope* *Net Method* *Gross Proxy Method*	This EP is not relevant for MIGA. While the scope permits the use of either net or gross profit contributions, the net contribution method is the more rigorous and should be favoured if cost accounting data are available. Gross profit contribution is applied in a largely qualitative manner as a proxy for likely investment performance, based on the incidence (or not) of loan impairments, called guarantees, or equity gains / losses.
	Binary Benchmark: <u>As a minimum</u>, for a positive rating of investment profitability for loans, either: (a) the loan's net profit contribution is sufficient in relation to the IFI's target return on capital employed or overall profitability objectives; or (b) the loan is expected to be paid, or has been paid, as scheduled (or rescheduled) or prepaid, with no loss of capital. In other words, the loan's gross profit contribution quality meets at-appraisal expectations.	*Benchmark*	
	Binary Benchmark: <u>As a minimum</u>, for a positive rating of investment profitability for financial guarantees, either: (a) the guarantee's net profit contribution is sufficient in relation to the IFI's target return on capital employed or overall profitability objectives; or (b) all guarantee fees have been received or are expected to be received, and the guarantee is not called, or is called but expected to be fully repaid in accordance with the terms of the guarantee agreement. In other words, the guarantee's gross profit contribution quality meets at-appraisal expectations.	*Benchmark*	
	Binary Benchmark: <u>As a minimum</u>, for a positive rating of investment profitability for equity investments, either: (a) the expected or realized net profit contribution (or net RoE) is sufficient in relation to the IFI's overall profitability objectives or target return on capital employed; or	*Benchmark*	

Private Sector Principles: Evaluation metrics and benchmarks

Evaluation Principle (Standards and Elements)	Standard Operational Practices	Linked to	Notes
	(b) the expected or realized gross profit contribution (or gross RoE) reflects an appropriate spread over actual or notional loan yields for the same credit risk, in line with the policy-defined at-entry approval standard. In both cases, the valuations of active equity investments should be appropriately discounted to reflect the uncertainty of still-to-go dividend income or capital realization.		
	Where the IFI makes more than one type of investment in a single project, either: (a) one rating is assigned on the basis of the combined net profit contribution of the investments; and as a minimum, for a positive rating of investment profitability, the net profit contribution is sufficient in relation to the IFI's target return on capital employed or overall profitability objectives; or (b) ratings are assigned and reported for each investment instrument separately.	*Benchmark*	Since gross profit contribution quality is a predominantly qualitative concept within this EP, it can not be numerically aggregated for the purposes of a synthesis rating. Should the CED wish to report a synthesis investment outcome from gross profit contribution measures, it could use a lookup table based on the ratings for each underlying investment instrument, possibly applying a weighting in line with the size of each type of investment made in the project.
Rating IFI work quality / bank handling: A. *Scope:* The indicator measures the quality of the IFI's pre-commitment work and on-going monitoring and supervision. B. *Stand-Alone:* The rating is independent of – and so not directly influenced by – the project's results.	Scope of Measurement: The indicator considers both the IFI's pre-commitment work in at-entry screening, appraisal and structuring / underwriting, and its monitoring and supervision of the operation following commitment / guarantee issuance. These elements can be rated separately or in combination as IFI work quality / bank handling. The assessment should be made independently of the ratings assigned for the project's outcome and the IFI's investment profitability. It should reflect the quality of the IFI's contributions to good or bad outcomes, not the good or bad outcomes themselves.	*Scope* *Stand-Alone*	
C. *Pre-Commitment:* The rating considers all aspects of the IFI's work in screening, appraising and structuring the project and the IFI's associated investment. D. *Post-Commitment:* The rating considers all aspects of the IFI's portfolio responsibilities in monitoring and supervising the project and the IFI's	<u>Pre-commitment work</u> quality assesses how effectively the IFI carried out its work prior to approval and commitment of the investment. It should consider all factors relevant to the institution's processing of the investment, for example: - the quality of the IFI's assessment of the operation as being relevant to the IFI's corporate, country and sector strategies; - the assessment of sponsors, company, management, country conditions, market dynamics, project concept, configuration and cost;	*Pre-Commitment*	

Private Sector Principles: Evaluation metrics and benchmarks

Evaluation Principle (Standards and Elements)	Standard Operational Practices	Linked to	Notes
associated investment. E. *Benchmark:* The rating is assigned on the basis of the IFI having executed its responsibilities to an internally or externally recommended standard.	- the appraisal of the project financial plan, source of project funds, and assumptions used in the project's financial projections; - the effectiveness of the IFI's assessment of project and political risks, and steps taken to mitigate them; - the appraisal of project environmental and social risk, and inclusion of appropriate requirements in the legal agreement; - investment instrument selection (as applicable), structure, pricing, exit / repayment mechanism, security, covenants and other terms and conditions; and - the clients' satisfaction with the IFI's service quality up to commitment.		
	Monitoring and supervision work quality assesses to what extent the IFI has adequately executed its portfolio responsibilities for the operation following commitment of the investment. It should consider all factors relevant to the institution's administration of the investment, for example: - the completeness of supervision reports in documenting project status and risk; - the monitoring of the client company's compliance with the terms of the investment, including financial, information and performance covenants; - the monitoring of the client company's environmental and social performance, and adherence to relevant government regulations and IFI requirements; - the adequacy and timeliness of the IFI's response to emerging problems or opportunities; - the effectiveness of hand-over procedures should there be changes in IFI staff monitoring responsibilities; - the clients' satisfaction with the IFI's service quality following commitment; and - the conduct of and contribution made by a representative of the IFI (if applicable) on the client company's board.	*Post-Commitment*	
	Binary Benchmark: As a minimum, for a positive rating of IFI work quality / bank handling (or for its two components individually), the IFI should have materially met its prescribed operational procedures and quality standards throughout all stages	*Benchmark*	As a point of reference, this rating uses the IFI's internally documented standards as the

Private Sector Principles: Evaluation metrics and benchmarks

Evaluation Principle (Standards and Elements)	Standard Operational Practices	Linked to	Notes
	of the operation. The IFI should have kept itself sufficiently informed to react in a timely manner to any material change in the project and/or company's performance (or any event or circumstance that could be the basis for a claim under an IFI's guarantee), and have taken timely action where needed.		benchmark. However, the CED should check periodically that such standards are in line with any internationally recognised standards of good practice in commercial banking, investment or insurance institutions.

Private Sector Principles: Evaluation metrics and benchmarks

Evaluation Principle (Standards and Elements)	Standard Operational Practices	Linked to	Notes
Rating the IFI's additionality: A. *Counterfactual:* The indicator measures the IFI's additionality in supporting the project, based on the counterfactual of what would have happened without the IFI's support. B. *Financial Additionality:* The rating considers the IFI's financial additionality in providing funding and/or catalysing other funding. C. *Non-Financial Additionality:* The rating considers the IFI's non-financial additionality in improving the project's risk profile, design or functioning. D. *Benchmark:* The rating is assigned on the basis of the IFI having fulfilled its mandate-defined objectives as a financier of private sector projects.	Scope of Measurement: The rating of the IFI's additionality considers the IFI's value proposition in providing support to the project. It is based on the counterfactual assessment of how the project would have (or would not have) proceeded without IFI support. It should consider all factors relevant to the role and contribution of the IFI, for example:* <u>Financial Additionality:**</u> - Would the client have been able to obtain sufficient financing / insurance from private sources on appropriate terms? Judgments on this indicator consider pricing (including additional costs arising from IFI conditions that would not be imposed by a private investor), tenor, grace period, currency, and timeliness (i.e., the availability of financing without unduly delaying the project). - Was the IFI catalytic in mobilizing funds from other investors and lenders, or was it merely helping to complete the financing package? - Was the IFI (by virtue of its being an IFI) needed to reduce risks or provide comfort (i.e., improve the investors' perceptions of the risks involved) and, thus, to encourage other investors and lenders to proceed? <u>Non-Financial Additionality:**</u> - Was the IFI needed to bring about a fair, efficient allocation of risks and responsibilities e.g., between the public and the private investors? - Did the IFI improve the project's design (through contributing knowledge or innovation), help the client's functioning in business (including adoption of new or better standards), or otherwise contribute to the client's capacity-building objectives?***	*Counterfactual* *Financial Additionality* *Non-Financial Additionality*	* Depending on the IFI's mandate objectives or the scope of its engagement in project selection and structuring, some of the factors listed may not be relevant to the rating of additionality. ** For the purpose of the GPS, additionality factors are grouped into financial and non-financial types, though CED may choose different categories, for example: risk mitigation; policy setting; knowledge and innovation; and standard setting. *** These types of contribution could arise through a parallel or linked technical assistance project. The CED should determine whether to reflect this in the evaluation of the investment operation, or separately in an evaluation of the TA intervention.
	Binary Benchmark: <u>As a minimum</u>, for a positive rating of additionality, it is evident that, absent the IFI: (a) the project would not have gone ahead with financing on appropriate terms and/or without undue delays; or (b) the project would have entailed an unfair or inefficient allocation of risks and responsibilities; or (c) the project would have been weaker in design, business, developmental, transition, social or environmental terms.	*Benchmark*	

Annex IV.1: Benchmarking against the GPS Fourth Edition

1. Compliance with the GPS Fourth Edition is measured by the extent to which IFIs fulfil the Elements defining each of the EPs. The assessment is made by examining the IFI's practices and judging their consistency with the Elements of the EPs. An IFI can achieve full compliance with an EP by implementing all the associated standard OPs documented in the GPS. However, the IFI can also adopt alternative, non-prescribed practices that it believes (and the benchmarking consultant agrees) are consistent with the EP; these may in turn be incorporated in future revisions to the GPS. Note that compliance will rely on evidence of the IFI actually implementing appropriate practices; it is not sufficient for the IFI to have documented procedures and policies if these are not then carried out in practice.

2. Benchmarking scores will show the proportion of EP Elements (as relevant to each institution) with which the IFI's practices are determined to be consistent, along with the consultant's decisions on each Element individually. Compliance with the Elements of the Generic EPs will be reported separately from the Elements of the Private Sector EPs. Additionally, the benchmarking will highlight EPs where the IFI is found to be particularly strong or, conversely, falls significantly short of an acceptable level of compliance. The latter information is intended to assist the IFI in identifying its own internal strengths and weaknesses and aims to promote continuous improvement in evaluation practices. The determination of strong or weak EPs is not rigidly defined, but is left to the judgment of the benchmarking consultant based on those Elements of the EP that are complied or not complied with.

3. The following schematic illustrates the benchmarking process and presentation of the results for each institution.

Examine the IFI's Operational Practices	→	**Determine Compliance with Evaluation Principle Elements**	→	**Determine Areas of Strength or Weakness**
Standard Prescribed OPs: Fully Implemented Partly Implemented Not Implemented Alternative Non-Prescribed OPs:* Implemented	}	Score each EP Element: Materially Compliant (1 pt) Partly Compliant (0.5 pt) Not Compliant (0 pt) ↓ Report	}	EPs where compliant with virtually all Elements EPs where compliant with few or no Elements ↓ Report

Scorecard

Name of IFI	Compliance (%) =	$\frac{\sum \text{Scores for EP Elements}}{\sum \text{EP Elements**}}$	List of Strong EPs List of Weak EPs

* Alternative, non-prescribed practices should be considered for incorporation into future revisions of the GPS.
** As relevant to each institution.

Annex IV.2: Project Typology

Where appropriate, the GPS have been tailored to the characteristics of different types of project supported under the IFI's private sector mandate. These project types broadly fall into five groups. It is the responsibility and ultimate decision of the CED to assign the relevant project type when applying the framework. For projects exhibiting a mix of project types, it is recommended that CEDs use a range of metrics as appropriate.

Project Type:	**Description:**	**Financing Instrument typically used to support project:**
Group A: Capital expenditure projects involving direct investments in identifiable assets		
Greenfield or limited recourse	Investment in a new venture or in a stand-alone company created for the purpose of investing in new assets / undertaking a concession etc.	Loan, Quasi-Equity, Equity, Financial Guarantee, Partial Risk Guarantee, MIGA PRI
Expansion of existing operations	Investment made by an existing company to expand capacity / enter a new business or market. Investment is made on balance sheet.	Loan, Quasi-Equity, Equity, Financial Guarantee, Partial Risk Guarantee, MIGA PRI
Rehabilitation or modernization	Investment made by an existing company to upgrade existing assets. No new assets are created by the investment.	Loan, Quasi-Equity, Equity, Financial Guarantee, Partial Risk Guarantee, MIGA PRI
Group B: Institutional investments supporting broad corporate investment programs		
General Corporate Investment in a Non-Financial Institution	To support a broad corporate expansion program where individual investments are too numerous or too general to be identifiable.	Loan, Quasi-Equity, Equity, Financial Guarantee, Partial Risk Guarantee
Corporate Investment in a Financial Institution	To improve the FI's capital / liquidity / maturity profile. May include corporate governance reforms or other institutional improvements. Corporate FI investments should be treated under Group D1 where their effects are manifested in changes in the FI's portfolio.	Loan, Quasi-Equity, Equity, Financial Guarantee, Partial Risk Guarantee, MIGA PRI
Group C: Projects supporting financial diversification, refinancing or short-term funding requirements		
Corporate Financial Restructuring	Concerned with refinancing (right-side balance sheet) of existing debt / equity structure. No new assets created.	Loan, Quasi-Equity, Equity, Financial Guarantee, Partial Risk Guarantee, MIGA PRI
Working Capital Finance	To support short-term or permanent funding requirements arising out of the normal course of trade.	Loan, Financial Guarantee, Partial Risk Guarantee
Securitization / Credit Enhancement	Participation in, or credit enhancement of, new securities backed by a pool of income-generating assets.	Loan, Quasi-Equity, Equity, Financial Guarantee, Partial Risk Guarantee, MIGA PRI
Group D1: Investments made in multiple sub-projects via intermediation in a bank or other credit institution		
Intermediation via FI with / without attribution to sub-projects	Credit line provided with specific objectives to support the investment programs of sub-borrowers. May or may not have requirements to track and report sub-loans.	Loan, MIGA PRI
Trade Finance / Factoring	To support, via intermediation, the short-term funding requirements of multiple sub-borrowers arising out of the normal course of trade. Typically no attribution.	Loan, Financial Guarantee, Partial Risk Guarantee
Group D2: Investments made in multiple sub-projects via intermediation in a fund		
Investment in Private Equity Fund	Equity (sometimes loan) subscription to fund, where the underlying investments are not listed / traded on any exchange.	Loan, Quasi-Equity, Equity
Investment in Listed Equity Fund	Equity subscription to a fund where the underlying investments are listed and traded on local / international exchanges. May feature a debt instrument to provide leverage.	Loan, Quasi-Equity, Equity

Annex IV.3: Lookup Table for Determining Early Operating Maturity

IFI Support Instrument:	Project Type A	Project Type B	Project Type C	Project Type D1	Project Type D2
Loan / Equity / Quasi-Equity	(a) The project has been substantially completed; and (b) The project has generated at least 18 months of operating revenues for the company; and (c) The IFI has received at least one set of audited annual financial statements covering at least 12 months of operating revenues generated by the project.	(a) The IFI has made its final material disbursement (i.e., any further disbursements will be minor in comparison to the overall facility size and not critical to project implementation); and (b) Other parallel financing (if applicable) has also been disbursed; and (c) The IFI has received at least one set of audited annual financial statements covering at least 36 months of operating revenues post-disbursement.	(a) The IFI has made its final material disbursement (i.e., any further disbursements will be minor in comparison to the overall facility size and not critical to project implementation); and (b) Other parallel financing (if applicable) has also been disbursed; and (c) The IFI has received at least one set of audited annual financial statements covering at least 24 months of operating revenues post-disbursement.	(a) Where the principal objective is to assist capital expenditure projects in sub-borrowers, at least 30 months should have elapsed following the IFI's final material disbursement to the Financial Intermediary. (b) Where the principal objective is to support the short-term working capital or trade finance requirements of sub-borrowers, at least 24 months should have elapsed following project approval/commitment.	(a) For all funds, the substantial majority of sub-investments should have been exited; or (b) For private equity funds, at least 36 months should have elapsed following the IFI's final material disbursement to the fund (ignoring disbursements for small follow-up investments in existing client companies and disbursements to cover management fees or other expenses); or (c) For listed equity funds, at least 18 months have elapsed following the IFI's final material disbursement to the fund (ignoring disbursements to cover management fees or other expenses).
Financial Guarantee	(a) The project has been substantially completed; and (b) The project has generated at least 18 months of operating revenues for the company; and (c) The IFI has received at least one set of audited annual financial statements covering at least 12 months of operating revenues generated by the project.	(a) The IFI has issued the guarantee and is at or near its approved exposure limit; and (b) Other parallel financing (if applicable) has also been disbursed; and (c) The IFI has received at least one set of audited annual financial statements covering at least 36 months of operating revenues post-issuance and disbursement of parallel funding (if	(a) The IFI has issued the guarantee and is at or near its approved exposure limit; and (b) Other parallel financing (if applicable) has also been disbursed; and (c) The IFI has received at least one set of audited annual financial statements covering at least 24 months of operating revenues post-issuance and disbursement of parallel funding (if	(a) Where the principal objective is to support the short-term trade finance requirements of sub-borrowers, at least 24 months should have elapsed following project approval/ commitment.	Not Applicable.

		applicable).	applicable).		
MIGA Political Risk Insurance	(a) at least 36 months should have elapsed following the issuance of the PRI guarantee.	(a) at least 36 months should have elapsed following the issuance of the PRI guarantee.	(a) at least 36 months should have elapsed following the issuance of the PRI guarantee.	(a) at least 36 months should have elapsed following the issuance of the PRI guarantee.	(a) at least 36 months should have elapsed following the issuance of the PRI guarantee.

Annex IV.4: Comparative Names of CED in IFIs and Reports on Indirect and Direct Evaluation[47] of Private Sector Operations

	Terms Used in GPS Document and their equivalent in each IFI			
	Central Evaluation Department (CED)	**Expanded Annual Supervision Report (XASR)**	**XASR Assessment**	**Performance Evaluation Report (PER)**
African Development Bank (AfDB)	Operations Evaluation Department (OPEV)	Expanded Supervision Report (XSR)	XSR Review Note	Project Performance Evaluation Report (PPER)
Asian Development Bank (AsDB)	Independent Evaluation Department (IED)	Extended Annual Review Report (XARR)	XARR Validation Report	Project Performance Evaluation Report (PPER)
Council of Europe Development Bank (CEB)	Ex-Post Evaluation Department	Annual monitoring report on the preparation and follow up of projects	Not Applicable	Ex-Post Evaluation Report
European Bank for Reconstruction and Development (EBRD)	Evaluation Department (EvD)	Expanded Monitoring Report (XMR)	XMR Assessment (XMRA)	Operation Performance Evaluation Review (OPER)
European Investment Bank (EIB)	Operations Evaluation (EV)	Project Progress Report (PPR) / Project Completion Report (PCR)	Not Applicable	Project Evaluation Report (PER)
Global Environment Facility (GEF)	Evaluation Office	Not Applicable (undertaken by GEF agencies)	Not Applicable (undertaken by GEF agencies)	Annual Performance Report
Inter-American Development Bank (IaDB)	Office of Evaluation and Oversight (OVE)	Expanded Supervision Report (XPSR)	XPSR Addendum (XPSR-A)	Independent Evaluation Report of the Expanded Project Supervision Report Exercise
Inter-American Investment Corporation (IIC)	Office of Evaluation and Oversight (OVE)	Expanded Annual Supervision Report (XASR)	XASR Addendum (XASR-A)	Annual Independent Evaluation Report to the IIC Board of Executive Directors
International Finance Corporation (IFC)	Independent Evaluation Group (IEG)	Expanded Project Supervision Report (XPSR)	XPSR Evaluation Note (EvNote)	Project Evaluation Summary (PES)
International Fund for Agricultural Development (IFAD)	Independent Office of Evaluation (IOE)	Project Completion Report	Project Completion Report Validation	Project Performance Assessment
Islamic Development Bank (IsDB)	Group Operations Evaluation Department	Project Implementation Assessment and Support Report (PIASR)	PIASR Evaluation Note (PIASREN)	Project Performance Evaluation Report (PPER)
Multilateral Investment Fund (MIF)	Office of Evaluation and Oversight (OVE)	Expanded Supervision Report (ESR)	ESR Comments	Independent Evaluation Report to the Donors Committee of the MIF
Multilateral Investment Guarantee Agency (MIGA)	Independent Evaluation Group (IEG)	MIGA Project Evaluation Report (PER)	Validation Note (VN)	IEG Project Evaluation Report (PER)
Black Sea Trade and Development Bank (BSTDB)	Evaluation Office	Operation Completion Report (OCR)	OCR Validation Note	Operation Performance Evaluation Report (OPER)

[47] See footnote #44.

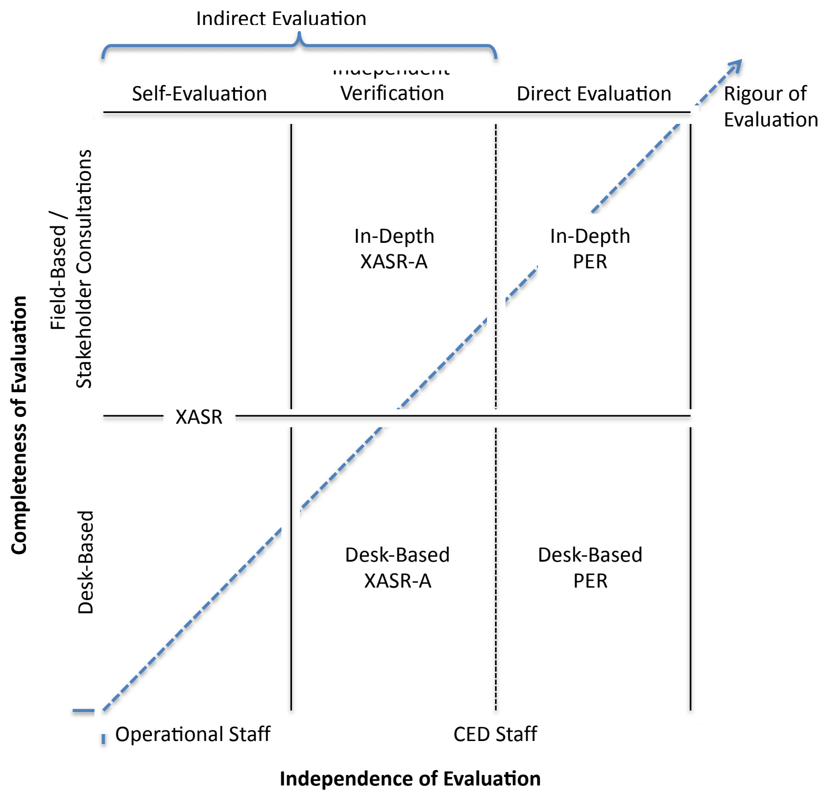
Annex IV.5: Additional Guidance Notes

1. These guidance notes are provided for the benefit of members as additional material to supplement the good practice standards. They do not constitute good practice standards in and of themselves, and so fall outside the scope of any benchmarking exercise. References are to the respective Evaluation Principle or Operational Practice.

EP 8/9: Comparison of Direct and Indirect Evaluation Methods [48]

EPs 8 and 9 cover the processes of direct evaluation by the CED, and indirect evaluation involving self-evaluation by operational staff and independent verification by the CED. The GPS covers evaluation and verification methods based on desk reviews and/or field-based stakeholder consultations. In general, the IFI should favour the more rigorous approaches as far as resources permit. The schematic (right) indicates the level of rigour typically associated with each approach and the related evaluation product.

Note that a desk-based PER should only be considered where the institution has collected regular monitoring data on the project.



OP 10.2: Project Outcome – Extended Rating Scale

2. The project's outcome is rated using benchmarks substantially consistent with the following:

Highly Successful:	A project with overwhelming positive results, and no flaws.
Successful:	A project with some strong results, and without material shortcomings.
Mostly Successful:	A project with a clear preponderance of positive results (i.e., it may exhibit some minor shortcomings though these should be clearly outweighed by positive aspects). The guiding principle should be that if all the IFI's projects exhibited this level of performance, the IFI should be able to demonstrate the successful execution of its institutional mandate.
Mostly Unsuccessful:	A project with either minor shortcomings across the board, or an egregious shortcoming in one area that outweighs other generally positive results.
Unsuccessful:	A project with largely negative results, clearly outweighing positive results.
Highly Unsuccessful:	A project with material negative results and with no material redeeming positive results.

[48] See footnote #44.

OP 11.2 : Outcome Indicator 1 – Financial Performance and Fulfilment of Project Objectives – Scope of Measurement for Project Types A, B and C

3. For project types A and B, where it should be possible to identify cashflows associated with the project assets, the evaluation is based primarily on an estimation of the project financial rate of return (FRR) or return on invested capital (ROIC). IFIs are expected to use the project company's Weighted Average Cost of Capital (WACC) as the benchmark, but the GPS does not prescribe how this should be calculated or what assumptions should be made. Given the diversity of projects supported by IFIs, the CED should have the flexibility to apply its own assumptions in relation to debt / equity ratios, cost of debt, tax rates, and equity premium, in estimating project WACCs. The onus falls upon the CED to validate the WACC calculation or make its own estimate according to accepted principles. Further guidance on the calculation of the WACC can be found in *Principles of Corporate Finance*; Brealey R. and Myers S.; McGraw-Hill.

4. Some members have based their evaluation of financial performance on a comparison of actual financial results against those projected at the time of Board approval. Past GPS have not supported such methodology, because it introduces possible bias depending on the efficacy of the benchmarks (i.e., two identically performing projects could be rated differently by virtue of differing levels of optimism in their respective appraisal projections). However, the comparison of actual financial results against appraisal projections can have a place in the GPS, provided that the CED verifies that the appraisal projections represent a valid benchmark. For example, at a minimum the appraisal projections should demonstrate that the project generates sufficient profit and cashflow to meet the company's obligations to lenders and creditors, and yields a net return to shareholders commensurate with the project risk. Provided that these checks are made, then the process of comparing actual results against appraisal projections is essentially the same as comparing the project FRR / ROIC against the WACC, since the WACC is the return necessary to satisfy all the project's financiers / shareholders. GPS4 therefore permits a methodology based on comparison of actual results against appraisal projections, provided that there is sufficient evidence (quantitative or qualitative) that the project has satisfied the return requirements of all financial stakeholders in the company.

OP 11.3: Outcome Indicator 1 – Financial Performance and Fulfilment of Project Objectives – Extended Rating Scale for Project Types A, B and C

5. The project's financial performance and fulfilment of project objectives is rated using the following benchmarks based on the methodology chosen as set out in OP 11.2 as follows: 1. Quantitative Method; 2. Achievement of Appraisal Projections; 3. Achievement of Objectives; 4. Analysis of Financial Statements; and 5. Business Prospects.

Excellent:	1. The project's FRR or ROIC is equal to or greater than 1.25 x WACC. 2. Actual performance exceeds appraisal projections such that the project has demonstrably met its obligations to lenders and creditors, and has yielded a premium return to its shareholders well in excess of that commensurate with the project risk. 3. The project's process and business goals articulated at approval are surpassed. 4. Performance indicators demonstrate clear outperformance against appraisal estimates. 5. The project company's overall profitability and prospects for sustainability and growth are strong, such that it is expected to retain or achieve market-leading status.

Satisfactory:

1. The project's FRR or ROIC is equal to or greater than the project company WACC.
2. Actual performance meets or exceeds appraisal projections such that the project has demonstrably met its obligations to lenders and creditors, and has yielded the minimally acceptable return to its shareholders commensurate with the project risk.
3. The project's process and business goals articulated at approval are broadly achieved or are deemed within reach albeit with some risk to their realisation.
4. Performance indicators are in line with appraisal estimates.
5. The project company's overall profitability and prospects for sustainability and growth are sound, such that it is expected to remain competitive in relation to the market and its sector peers.

Partly (Un)satisfactory:

1. The project's FRR or ROIC is equal to or greater than 0.7 x WACC.
2. Actual performance has lagged appraisal projections such that the project has demonstrably met its obligations to lenders and creditors, but the return to shareholders is less than that deemed minimally acceptable albeit at least equal to the cost of debt.
3. At least one of the project's process and business goals articulated at approval is not met.
4. Performance indicators have fallen short of appraisal estimates in one or more key areas.
5. The project company's prospects for sustainability and growth are weak, such that it is struggling to remain competitive in relation to the market and its sector peers.

Unsatisfactory:

1. The project's FRR or ROIC is lower than 0.7 x WACC.
2. Actual performance has lagged appraisal projections such that the project has failed to meet its obligations to lenders and creditors and/or has yielded a return to shareholders that is less than the cost of debt.
3. Most of the project's process and business goals articulated at approval are not met.
4. Performance indicators have fallen short of appraisal estimates in the majority of key areas.
5. The project company's prospects for sustainability and growth are weak or negative, such that it is clearly underperforming in relation to the market and its sector peers.

OP 11.5 : Outcome Indicator 1 – Financial Performance and Fulfilment of Project Objectives – Scope of Measurement for Project Type D2

GPS3 prescribed that the financial performance of a fund should be evaluated on a comparison of the return on equity (RoE) to investors with the return on the S&P index over the same period. However, this was at odds with the principle that projects should be judged as far as possible against absolute benchmarks, rather than relative to market indices. It introduced the possibility of anomalies such as where a fund's investee companies perform poorly, yet the overall business success is judged satisfactory by virtue of a fall in the S&P index and an artificially low benchmark. The rating would therefore fail to describe accurately the actual commercial performance of the sub-projects themselves. In contrast, had the IFI made direct investments in the sub-projects, their financial performance would be rated on more exacting criteria (for example on an FRR vs. WACC, or actual vs. expected performance basis).

Consequently, GPS4 has dropped the reference to the S&P index as a relevant benchmark for rating the financial performance of funds. Instead, it recommends a methodology similar to that proposed for project types A or B i.e., a comparison of the aggregate RoE to the fund's investors with the fund's effective cost of capital.

The fund's cost of capital is estimated by calculating the average cost of debt based on the country composition of the fund, and then levying a premium of 600 bpts for the combined equity instrument and project risk. The fund's weighted average cost of capital (FWACC) is therefore:

$$FWACC = \frac{[E_1(C_d + c_1) + E_2(C_d + c_2) + ... + E_n(C_d + c_n)]}{(E_1 + E_2 + ... + E_n)} + 600bpts$$

where: E_n is the amount of the fund actually invested in country n;

C_d is the 10 year fixed rate swap equivalent of 6 month LIBOR, as at the date of commitment; and

c_n is the spread applied by the IFI's pricing policy in respect of country n to reflect country macro risk, as at the date of commitment.

This formula assumes that the fund comprises only equity funding, and is not leveraged through debt.

OP 11.5 : Outcome Indicator 1 – Financial Performance and Fulfilment of Project Objectives – Extended Rating Scale for Project Types D1 and D2

6. The project's financial performance and fulfilment of project objectives is rated using the following benchmarks based on the methodology chosen as set out in OP 11.4 as follows: 1. Performance of Sub-Portfolio; 2. Performance of Fund Portfolio; 3. Achievement of Objectives; and 4. Performance of Intermediary.

Excellent:	1. There is strong evidence (quantitative or qualitative) that the sub-portfolio has substantially raised the financial intermediary's profitability, and substantially improved its viability. 2. The projected or realized net return on equity (RoE) or net IRR to the fund's investors is equal to or greater than the FWACC x 1.25 3. The project's business and process goals articulated at approval are surpassed. The intermediary has substantially increased its reach to sub-borrowers or investee groups that were specified as targets at approval. 4. The intermediary's overall profitability, adaptability and prospects for sustainability and growth are strong, such that it is expected to retain or achieve market-leading status.
Satisfactory:	1. There is adequate evidence (quantitative or qualitative) that the sub-portfolio has had a positive effect on the financial intermediary's profitability, and helped improve its viability. 2. The projected or realized net RoE or net IRR to the fund's investors is equal to or greater than the fund's weighted average cost of capital (FWACC). 3. The project's business and process goals articulated at approval are broadly achieved or are deemed within reach albeit with some risk to their realisation. The intermediary has succeeded in reaching sub-borrowers or investee groups that were specified as targets at approval. 4. The intermediary's overall profitability, adaptability and prospects for sustainability and growth are sound, such that it is expected to remain competitive in relation to the market and its sector peers.
Partly (Un)satisfactory:	1. There is evidence (quantitative or qualitative) that the sub-portfolio has had a negative effect on the financial intermediary's profitability and/or detracted from its viability. 2. The projected or realized net RoE or net IRR to the fund's investors is equal to or greater than the FWACC x 0.7. 3. At least one of the project's business and process goals articulated at approval is not met. The intermediary has failed to reach sub-borrowers or investee groups that were specified as targets at approval. 4. The intermediary's overall profitability, adaptability and prospects for sustainability and growth are weak, such that it is struggling to remain competitive in relation to the market and its sector peers.
Unsatisfactory:	1. There is evidence (quantitative or qualitative) that the sub-portfolio has had a substantial negative effect on the financial intermediary's profitability and/or harmed its viability. 2. The projected or realized net RoE or net IRR to the fund's investors is less than the FWACC x 0.7. 3. Most of the project's business and process goals articulated at approval are not met. The intermediary has failed to reach sub-borrowers or investee groups that were specified as targets at approval and/or has used funds to support undesirable sub-borrowers. 4. The intermediary's overall profitability, adaptability and prospects for sustainability and growth are negative, such that it is clearly underperforming in relation to the market and its sector peers.

OP 12.3 : Outcome Indicator 2 – Economic Sustainability – Extended Rating Scale for Project Types A, B and C

7. The project's economic sustainability is rated using the following benchmarks based on the methodology chosen as set out in OP 12.2, either: 1. Quantitative Method; or 2. Qualitative Stakeholder Analysis.

Excellent:	1. The ERR or EROIC is equal to or greater than the larger of either: (i) a multiple of 1.75 times the project company WACC; or (ii) 17.5%. 2. The project meets the minimum standard for satisfactory financial performance and there is evidence that: (i) it has generated substantial net economic benefits for its wider stakeholders (i.e., those other than the project company's owners and financiers); and (ii) it does not rely on economic distortions to maintain its commercial viability.
Satisfactory:	1. The ERR or EROIC is equal to or greater than the larger of either: (i) a multiple of 1.2 times the project company WACC; or (ii) 10%. A positive rating may also be awarded if the ERR or EROIC falls short of the quantitative benchmark, but there are other material un-quantified net economic benefits that could be expected to raise the ERR or EROIC sufficiently. 2. Either: (i) the project meets the minimum standard for satisfactory financial performance and there is evidence that it has generated a balance of benefits for its wider economic stakeholders (i.e., those other than the project company's owners and financiers); or (ii) the project just fails to meet the minimum standard for satisfactory financial performance, but there is evidence that it has generated substantial net benefits for its wider economic stakeholders. In either case, the project should not rely on economic distortions to maintain its financial performance.
Partly (Un)satisfactory:	1. The ERR or EROIC is equal to or greater than the larger of either: (i) a multiple of 0.8 times the project WACC; or (ii) 5%. 2. Either: (i) the project fails to meet the minimum standard for satisfactory financial performance and there is insufficient evidence of significant net economic benefits for its wider stakeholders (i.e., those other than the project company's owners and financiers); or (ii) the project relies on economic distortions to maintain its commercial viability.
Unsatisfactory:	1. The ERR or EROIC is less than the larger of either: (i) a multiple of 0.8 times the project WACC; or (ii) 5%. 2. The project fails to meet the minimum standard for satisfactory financial performance and has resulted in net economic costs for its wider stakeholders (i.e., those other than the project company's owners and financiers).

OP 12.5 : Outcome Indicator 2 – Economic Sustainability – Extended Rating Scale for Project Types D1 and D2

8. The project's economic sustainability is rated using the following benchmarks based on the methodology chosen as set out in OP12.4 , either: D1. Economic Activities of Sub-Borrowers; or D2. Economic Viability of Fund Investees.

Excellent:	D1. Both: (i) the project has succeeded in reaching targeted groups of sub-borrower; and (ii) there is direct evidence (from sub-portfolio data) that sub-borrowers have made strong economic contributions, or indirect evidence (from market data) that market sectors supported by the project and/or more generally by the financial intermediary are major economic contributors to society. D2. Both: (i) the gross equity fund portfolio return (before management fees) is equal to or greater than the FWACC x 1.75; and (ii) at least half of equity fund investees have positive equity returns. There is direct evidence (from sub-portfolio data) that investees are economically viable, or indirect evidence (from market data) that market sectors supported by the project are major economic contributors to society.
Satisfactory:	D1. Both: (i) the project has succeeded in reaching targeted groups of sub-borrower; and (ii) there is direct evidence (from sub-portfolio data) that sub-borrowers are economically viable, or indirect evidence (from market data) that market sectors supported by the project and/or more generally by the financial intermediary are economically viable and do not rely on economic distortions to maintain their commercial viability. D2. Either: (i) the gross equity fund portfolio return (before management fees) is equal to or greater than the FWACC x 1.2; or (ii) at least half of equity fund investees have positive equity returns yet the gross portfolio return (before management fees) is less than FWACC x 1.2 but not less than the FWACC x 0.8. In either case, there is direct evidence (from sub-portfolio data) that investees are economically viable, or indirect evidence (from market data) that market sectors supported by the project are economically viable and do not rely on economic distortions to maintain their commercial viability.
Partly (Un)satisfactory:	D1. Either: (i) the project has largely failed to reach targeted groups of sub-borrower; or (ii) there is direct evidence (from sub-portfolio data) that most sub-borrowers are not economically viable, or indirect evidence (from market data) that market sectors supported by the project and/or more generally by the financial intermediary are weak economic contributors to society. D2. Both: (i) the gross equity fund portfolio return (before management fees) is equal to or greater than the FWACC x 0.8; and (ii) more than half of the fund's investees have zero or negative equity returns. There is direct evidence (from sub-portfolio data) that most investees are not economically viable, or indirect evidence (from market data) that market sectors supported by the project are weak economic contributors to society and/or rely on economic distortions to maintain their commercial viability.
Unsatisfactory:	D1. Both: (i) the project has largely failed to reach targeted groups of sub-borrower; and (ii) there is direct evidence (from sub-portfolio data) that most sub-borrowers are not economically viable, or indirect evidence (from market data) that market sectors supported by the project and/or more generally by the financial intermediary are weak economic contributors to society. D2. The gross equity fund portfolio return (before management fees) is less than FWACC x 0.8; and/or nearly all of the fund's investees have zero or negative equity returns. There is direct evidence (from sub-portfolio data) that most investees are not economically viable, or indirect evidence (from market data) that market sectors supported by the project are weak economic contributors to society and/or rely on economic distortions to maintain their commercial viability.

OP 13.1: Outcome Indicator 3 – Contribution to IFI Mandate Objectives – Component Definitions

The scope of components assessed under this indicator is as follows:

Competition: Contributions to greater efficiency, quality, innovation or customer service of other suppliers through competitive pressures, or contributions to restrictions on competition.

Market expansion: Expansion of markets through the project company's interactions with suppliers (backward linkages) and customers (forward linkages) and through contributions to the integration of business activities within the national or international economy.

Private ownership and entrepreneurship: Significant increase or consolidation of private provision of goods and services and support for entrepreneurial initiative; or weakening of support for private ownership and entrepreneurship (e.g., due to allocation by a financial institution of project resources to purchase government securities or make loans to state-owned enterprises).

Frameworks for markets (institutions, laws and policies that promote market functioning and efficiency): Creation or strengthening of public and private institutions that support the efficiency of markets; improvements to the functioning of regulatory entities and practices; contributions to government policy formation and commitment, promoting competition, predictability and transparency; contributions to laws that strengthen the private sector and an open economy.

Transfer and dispersion of skills: Significant upgrading of technical and managerial skills beyond the project entity; introduction of new technology or know-how, including financial know-how.

Demonstration effects (spread of new behaviours and activities): Introduction of replicable products and processes that are new to the economy; new investments stimulated by the project; demonstration of ways of successfully restructuring companies and institutions; new ways and instruments to finance private sector activity.

Standards for corporate governance and business conduct: Improvements in accounting standards, disclosure standards, risk management standards, governance quality, reputation and/or business practices, which serve as a positive corporate role model.

Development of financial institutions and financial / capital markets: Development of sustainable financial institutions and the financial markets in which they operate (including creation of new fund management companies of subsequent investment funds); improved financial strength in sector (e.g., by improving asset-liability management); pioneering listing on stock exchange or significant broadening of listed value; greater resource mobilization; and improved allocation efficiency.

Development of physical infrastructure: used by other private parties.

OP 13.2: Outcome Indicator 3 – Contribution to IFI Mandate Objectives – Extended Rating Scale

9. The project's contribution to IFI mandate objectives is rated using benchmarks substantially consistent with the following:

Excellent:	Considering its size, the project had: (a) substantial positive effects consistent with the IFI's mandate objectives (for example, in furthering the country's private sector development, development of efficient financial / capital markets, or transition to a market economy); and (b) no negative impacts in this respect.
Satisfactory:	The project had: (a) demonstrable effects consistent with the IFI's mandate objectives (for example, in furthering the country's private sector development, development of efficient financial / capital markets, or transition to a market economy); and (b) a clear preponderance of sustainable positive impacts in this respect.
Partly (Un)satisfactory:	The project had mainly negative effects in respect of the IFI's mandate objectives (for example, in furthering the country's private sector development, development of efficient financial / capital markets, or transition to a market economy), but these negative effects are not expected to be of long duration or broad applicability.
Unsatisfactory:	The project had substantial negative effects in respect of the IFI's mandate objectives (for example, in furthering the country's private sector development, development of efficient financial / capital markets, or transition to a market economy), and these impacts are likely to be widespread, of long duration, or both.
*Neutral:**	The project made no discernable contribution, either positive or negative, to the IFI's mandate objectives (for example, in furthering the country's private sector development, development of efficient financial / capital markets, or transition to a market economy). This is distinct from a project with a balance of observed positive and negative impacts in which case a performance rating should be assigned.

*While most projects are expected to have some measurable effect in furthering the country's private sector development, development of efficient financial / capital markets, or transition to a market economy, there may be some projects that have no discernable impact. Examples might include:

- A mutual fund, which invests in listed equities, but due to its small overall size and lack of significant stake in any of its investee companies has no effect on market liquidity or influence on corporate governance quality. The fund's size and performance acts as neither an incentive nor deterrent to other investors in the country or index.
- A loan to support unspecified corporate expansion in a mature industry sector, which is repaid or prepaid having had no net impact on the firm's profitability, product range or market share.
- A credit line to a financial intermediary, which remains largely unutilized.

10. In such cases, it would be inappropriate to describe the project's contribution to IFI mandate objectives as either satisfactory or less than satisfactory. Moreover, it would introduce artificial bias to then base the overall synthesis rating of project outcome on such a rating. Hence, GPS4 permits the use of a "neutral" rating for this indicator where there is no discernable impact (as distinct from a balance between observed positive and negative impacts, in which case a rating should be assigned). It should be stressed that a "neutral" rating is not a

middle rating falling between satisfactory and partly (un)satisfactory. Rather, it signifies that the indicator should have no influence on the synthesis project outcome rating. It is different also to a rating of "No Opinion Possible", which could imply significant yet unknown positive or negative impacts.

EP 14: Optional Supplementary Indicator – Extent of Environmental and Social Change / Impact

11. The rating of extent of environmental and social change / impact considers both the ex-ante and ex-post conditions of the project compared with the IFI's specified requirements at approval and, therefore, the extent of progress or regress in the project's environmental and social performance. Whereas the rating of E&S performance is based on compliance with prescribed standards at the time of evaluation, this optional indicator measures whether such performance has improved or deteriorated over time (i.e., since approval).

12. Rating Scale: For this indicator, the CED should use a rating scale, which: (i) reflects the extent of environmental and social change delivered by the project (the largest positive change having occurred when the performance rating was the lowest at appraisal and the highest at evaluation); and (ii) captures wider E&S impacts to the industry sector, region, country, and supply chains (demonstration effect). A rating of Not Applicable should be used in cases where the project did not, and was not expected to, deliver any environmental or social impacts.

OP 14.2: Outcome Indicator 4 – Environmental and Social Performance – Extended Rating Scale for Project Types A, B and C

13. The company's overall environmental and social performance, in the area of influence of the project, is rated using benchmarks substantially consistent with the following:

Excellent:	The company meets both the IFI's at-approval requirements (including implementation of an ESAP, if any) and the IFI's at-evaluation requirements, and the extent of environmental and social change / impacts: (i) go beyond the expectations of the ESAP and key environmental and social requirements, or (ii) have materially improved overall environmental and social performance, or (iii) have contributed to a material improvement in the environmental and social performance of local companies (e.g., by raising industry standards, acting as a good practice example, etc.).
Satisfactory:	The company is in material compliance with the IFI's at-approval requirements (including implementation of an ESAP, if any).
Partly (Un)satisfactory:	Both: (a) the company is not in material compliance with the IFI's at-approval requirements (including implementation of an ESAP, if any), but is addressing deficiencies through on-going or planned actions; and (b) such non-compliance has not resulted in environmental damage.
Unsatisfactory:	Both: (a) the company is not in material compliance with the IFI's at-approval requirements (including implementation of an ESAP, if any); and (b) mitigation prospects are uncertain or unlikely, or non-compliance resulted in substantial and permanent environmental damage.
Not Applicable:	Where, by virtue of the project's expected lack of environmental and social impacts, the IFI had not prescribed any at-approval environmental and social requirements, a rating of Not Applicable may be assigned. However, should the project have subsequently changed in scope and given rise to environmental and social impacts, its performance should be rated accordingly against the standards that would have been prescribed had this been known at approval.

No Opinion Possible: Where, after best efforts, the relevant information to establish material compliance (or lack thereof) cannot be obtained, a rating of No Opinion Possible may be assigned. This rating should be a last resort, after reasonable effort has been made to obtain the necessary information. The company's failure to report should result in a partly unsatisfactory or unsatisfactory rating only if it has repeatedly refused to cooperate on this issue.

OP 14.2 : Outcome Indicator 4 – Environmental and Social Performance – Extended Rating Scale for Project Types D1 and D2

14. The company's overall environmental and social performance, in the area of influence of the project, is rated using benchmarks substantially consistent with the following:

Excellent:	The extent of environmental and social change / impacts: (i) go beyond the expectations of the ESAP and key environmental and social requirements, or (ii) have materially improved overall environmental and social performance, or (iii) have contributed to a material improvement in the environmental and social performance of local companies (e.g., by raising industry standards, acting as a good practice example, etc.). In addition, the company has provided transparent and timely reports, verifying that the project has consistently met the IFI's at-approval requirements, and as applicable, sub-projects have been appropriately appraised and supervised and their adverse environmental and social impacts have been mitigated, and that the environmental and social performance is deemed acceptable in view of the IFI's current requirements.
Satisfactory:	The company has implemented an appropriate Environmental & Social Management System (ESMS), which has been functioning over the project life (as reflected also in environmental and social standards being applied to projects financed by the intermediary). If required by the IFI's specified standards at approval, the environmental and social performance of sub-projects / fund investee companies are in material compliance with the IFI's requirements.
Partly (Un)satisfactory:	The company is not in material compliance with the IFI's at-approval requirements (including implementation of an ESMS), but is addressing deficiencies through on-going or planned actions and negative impacts are moderate. For example: the FI's ESMS is adequate, but some sub-projects have resulted in environmental damage that has not been corrected; or the sub-projects have acceptable environmental standards, but the ESMS is materially inadequate; or the company initially had no ESMS, but has recently introduced a functioning ESMS.
Unsatisfactory:	Both: (a) the company is not in material compliance with the IFI's at-approval requirements (including implementation of an ESMS); and (b) mitigation prospects are uncertain or unlikely, or sub-projects' non-compliance resulted in substantial and permanent environmental damage.
Not Applicable:	Where, by virtue of the project's expected lack of environmental and social impacts, the IFI has not prescribed any at-approval environmental and social requirements, a rating of Not Applicable may be assigned. However, should the project have subsequently changed in scope and given rise to environmental and social impacts, its performance should be rated accordingly against the standards that would have been prescribed had this been known at approval.
No Opinion Possible:	Where, after best efforts, the relevant information to establish material compliance (or lack thereof) cannot be obtained, a rating of no opinion possible may be assigned. This rating should be a last resort, after reasonable effort has been made to obtain the necessary information. The company's failure to report should result in a partly unsatisfactory or unsatisfactory rating only if it has repeatedly refused to cooperate on this issue.

V. GPS on Country Strategy and Program Evaluation

Contents

Background

32. **Rationale for MDB County-Level Evaluation.** Country strategy and program evaluations (CSPEs)[49] seek to describe and explain the performance of an MDB at the country level. They question whether the country program did the right things, in terms of whether the design and its implementation were right for the circumstances of the country. They ask not just "Did the country program work?" but "What made it work or fail?" and "How can we make it better"? Due to the fact that they usually evaluate both completed and ongoing operations, their forward-looking nature, and the controversy that they may generate, CSPEs tend to engage evaluation clients more than other forms of independent evaluation. Typically, they have been one of the more influential types of evaluation. Consequently, they play an increasingly important role in the work programs of the independent evaluation offices that conduct them.

33. CSPEs undertaken by MDBs are major and often costly evaluation exercises. They are classified as higher level evaluations because of their focus on strategic issues and because they build on the findings of evaluations of projects, programs, and sector or thematic issues of concern. A 1999 review of MDB evaluation experience describes the benefits of CSPEs:

> (i) CSPEs can identify and assess broad and long-term issues and concerns better than other forms of evaluation;
>
> (ii) they provide valuable information about the country strategy process, whether project selection was based on merit, impact of non-project forms of assistance, aggregating results of activities across all sectors and providing input into, and strengthening, subsequent country strategies;
>
> (iii) CSPEs are better able to identify overall program and project delivery weaknesses, institutional difficulties, capacity utilization constraints, borrower's acceptance, commitment and compliance to conditions and impact of other aid agencies and external factors;
>
> (iv) they provide a framework for rating overall performance in meeting development goals and objectives, and better assess impact and sustainability issues for long-term aid effectiveness; and
>
> (v) they provide a valuable instrument for improving aid co-ordination among institutions and bilateral agencies and for the broader participation goal of increasing the role of national and local governments, civil society and the private sector in the developmental process.[50]

34. As the locus of MDB assistance shifts from individual projects toward country-based strategies, programs, and interventions having economy-wide effects, the country becomes the most logical unit of aid management and accountability. Adoption of similar goals for development assistance (e.g., the Millennium Development Goals [MDGs]) and agreements to harmonize and align assistance with national poverty reduction strategies also make country-level evaluation of external assistance imperative.

---

[49] Use of the term "country" does not imply any judgment by the authors or the Asian Development Bank as to the legal or other status of any territorial entity (Source: p. ii of GPS on CSPE, 2008).

[50] OECD-DAC. 1999. *Evaluating Country Programmes. Report of the Vienna Workshop*. Paris. p. 115.

35. **CSPE Process.** CSPEs differ by purpose, by depth, and by the entity undertaking the evaluation. Within the MDBs, country assistance is typically evaluated as part of the preparation of new country strategies, both by the operational teams involved in preparing the country strategies and by the independent evaluation offices.[51] Most independent CSPEs undertaken by MDBs would be categorized as in-depth evaluation exercises or full-fledged CSPEs. These are most suitable and rewarding when there is something of value to learn through an in-depth assessment, plus an opportunity to make use of the findings. This would include cases in which

- a diverse portfolio of MDB assistance has been provided over an extended period,
- activities are sufficiently mature to be able to identify and/or anticipate results,
- government external assistance policies or aid agency assistance strategies are being formulated, and/or
- the lessons gleaned from the particular country case are expected to be of interest to other MDB member countries.

36. **GPS Formulation Process**. In October 2005, as part of ECG's ongoing effort to harmonize MDB evaluations, and consistent with the commitment to the OECD–DAC, ECG members declared their intention to prepare a set of GPS for the evaluation of country strategies and programs. The formulation has benefitted from consultations with MDB country evaluators and users and is also informed from experiences in specific country evaluations that provided an understanding of good practices from the perspective of country users in government, resident missions, development partners, and other stakeholders**.** The GPS was completed and presented to ECG in 2008.

37. **Purpose of GPS.** The standards aim to

- contribute to the ECG objectives of harmonization of evaluation criteria and processes,
- help MDBs link evaluation and operational standards in pursuit of corporate missions and objectives, and
- assist in learning from experience among MDBs for improved results.

38. **Guiding Framework.** The GPS have been developed within the general framework of the OECD–DAC evaluation principles, and they draw on the findings of an ECG review of CSPEs undertaken in 2007.[52] The GPS also build on the foundation of good evaluation practices that have already been identified and endorsed by the ECG in its GPS for public sector and GPS for private sector operations. More specifically, those GPS established for the organization and governance of the MDB independent evaluation process, as set forth in the 2002 *Good Practice Standards for Evaluation of MDB-Supported Public Sector Operations,* will likewise apply to CSPEs. Consequently, those GPS are not repeated in this Chapter.

39. The identified GPS on CSPEs are classified as "**Core**" GPS **[C-GPSs]** and "**optional**" GPS **[O-GPSs]**. A core GPS is defined as one that establishes the key principles for CSPEs and is necessary to permit comparability of evaluation results, to the extent possible, among MDBs. While the core GPS listed in this paper are currently in practice to some extent in all members, institutional differences may affect the pace at which harmonization can be achieved. An optional GPS is defined as one that is not strictly needed for comparability but is nonetheless designed to help improve accountability and learning within each institution. These GPS have a combined total of 86 C-GPS and O-GPS which are grouped into 16 standards consisting of

[51] Self-evaluation of country assistance is briefly discussed in Chapter VI.
[52] Tabor, Steven and Suganya Hutaserani. 2007. *Phase I Background Report for the Preparation of GPS for CSPEs*.

50elements. A summary of the standards, elements and number of C-GPS and O-GPS is presented in the Table below.

Summary of Standards and Elements on EPs and Number of OPs on CSPE

Evaluation Principles		**Number of C-GPS and O-GPS***	**Page**
Standards	**Elements**		
A. Process-Related GPS			
1. CSPE Goals, Objectives, Client Responsiveness, and Unit of Analysis	*A.* CSPE Goal	1	114
	B. Objectives	1	
	C. Client Responsiveness	1	
	D. Unit of Analysis	1	
2. Country Selection and Mutual Accountability	A. Country Selection	2	115
	B. Joint CSPs	1	
	C. Mutual Accountability	2	
3. Timing	A. Timely CSPEs	2	115
4. Advance preparations	A. Preparation Steps	1	116-117
	B. Sector/Thematic Studies	2	
5. Coverage	A. Time Period	2	117-118
	B. Product and Service Coverage	2	
	C. Second- or Third generation CSPEs	2	
	D. Limited Scope CSPEs	2	
	E. Validation Reports	1	
6. Approach paper for CSPEs	A. Specific Evaluation Approach	1	118
7. Preparation period	A. CSPE Implementation Period	1	119
8. Staffing	A. Evaluation Team	2	119
9. Guidelines	A. Uniform Guidelines, Quality Control, and Appropriateness	3	119
B. Methodology-Related GPS			
10. Methods and approaches for CSPEs	A. Overview	6	120-124
	B. Evaluation Questions	2	
	C. Counterfactuals	2	
	D. Attribution and Contribution	2	
	E. Evaluability	2	
	F. Multiple Evidence Sources	2	

Evaluation Principles		Number of C-GPS and O-GPS*	Page
Standards	**Elements**		
	G. Client Participation	1	
	H. Disclaimers	1	
11. Evaluation criteria for CSPEs	A. Relevance	6	124-130
	B. Positioning		
	C. Coherence		
	D. Efficiency	1	
	E. Effectiveness	3	
	F. Sustainability	1	
	G. Impact	1	
	H. Institutional Development	1	
	I. Borrower Performance	1	
	J. MDB Performance	1	
	K. Partnership and Harmonization	1	
12. Performance rating	A. Rating Principles and Comparability	5	130-131
	B. Rating Criteria	3	
	C. Rating Subcriteria	1	
	D. Weighting Criteria	1	
C. Reporting-Related GPS			
13. Findings, Lessons, and recommendation	A. Findings and Lessons	2	131
	B. CSPE Recommendations	1	
14. Reporting and review	A. Reporting	3	132
	B. CSPE Review	1	
15. Making findings available	A. Disclosure	2	132
	B. Dissemination	1	
16. Generalizing findings and tracking recommendations	A. Generalizing CSPE Findings	1	133
	B. Tracking recommendations	1	
Total no. of standards: 16	**Total no. of elements: 50**	**Total no. of C-GPS and O-GPS: 86**	

* Core GPS (C-GPS) is defined as one that establishes the key principles for CSPEs and is necessary to permit comparability of evaluation results, to the extent possible, among MDBs. Optional GPS (O-GPS) is defined as one that is not strictly needed for comparability but is nonetheless designed to help improve accountability and learning within each institution.

Source: GPS on CSPEs in next Section.

40. **Application.** The GPS pertain to the evaluation of country strategies and programs of both public and private sector-oriented MDBs, since they both strike a balance in their evaluation between "bottom-up" project-level evaluations and "top-down" assessments of business climate quality and the macroenvironment. It is also acknowledged that there are some differences between the CSPEs undertaken by public sector-oriented and by private sector-oriented MDBs. The private sector-oriented MDBs have financial return objectives that must pass the market test; they have far fewer assistance instruments aimed at having country-wide effects; their operations depend largely on market demand; and their corporate and country strategies tend to be illustrative of the range of activities in which their banks wish to engage. Consequently, their evaluations include more analysis of performance determinants, outcomes and impacts of projects, and technical cooperation activities. Moreover, private sector-oriented MDBs are very exposed to market fluctuations, and they maintain a frequent monitoring of the overall project portfolio for accounting and financial reporting purposes. The focus of private sector-oriented MDB CSPEs should therefore be more on lessons identified from strategy impact assessments such as environmental impacts, broader private sector development impacts, transition impacts, and economic/social impacts in the immediate area of the various projects.

41. **Updating GPS.** CSPE methods, approaches, rating criteria, and their application will continue to evolve over time. Adoption of results-based monitoring and evaluation systems in partner countries, and improvements in both the self-evaluation and independent evaluation of MDB operations, sector and thematic studies, special studies, and impact evaluations, will influence the nature of the evaluation data base upon which CSPEs are built. It is envisaged, therefore, that the GPS will require periodic stocktaking and updating. As members reach further agreements on CSPE methods, approaches, criteria, rating standards, and applications to special CSPE cases, they will document them in subsequent refinements of these GPS.

GPS for Independent Country Strategy and Program Evaluations

A. Process-Related GPS

GPS Category (Standards and Elements)	Core GPS	Optional GPS	Notes
A.1. CSPE Goals, Objectives, Client Responsiveness, and Unit of Analysis			
CSPE Goal	(a) The main goal of an MDB CSPE is to provide information on MDB performance at the country level that is credible and useful and enables the incorporation of lessons and recommendations that can be used to improve the development effectiveness of the MDB's overall strategy and program of country assistance.		
Objectives	(b) CSPEs are used for both accountability and lesson-learning purposes in the MDBs. They provide an accounting to the MDB's board of directors regarding the results achieved from MDB assistance in a country over an extended period of time. CSPEs also serve as an important learning experience by drawing on evaluation results to engage in a constructive dialogue on what could be done to improve the effectiveness of an MDB's assistance program in the future.		
Client Responsiveness	(c) CSPEs are designed to meet the information requirements of the main target clients, which would generally be the board, senior management, and relevant operations personnel within the country department. Identifying the government as the main target client is also a good practice, because the government will need information on past assistance performance if it is to demand better service from the MDB.		
Unit of Analysis	(d) CSPEs focus on evaluating the results of MDB assistance. They take the country as the unit of analysis and attempt to evaluate MDB assistance to the country using already prepared country strategy(ies) as a point of reference.		CSPEs do not evaluate the performance of a government or the progress of a country, although a CSPE may draw on country progress indicators to assess the performance of the assistance program.

GPS Category (Standards and Elements)	Core GPS	Optional GPS	Notes
A.2. Country Selection and Mutual Accountability			
Country Selection	(a) In practice, certain strategies and programs in some countries warrant more attention than others. Faced with limited evaluation resources, it is best to select those countries and programs for CSPEs where the findings and lessons will be most beneficial to the MDB and to the country.	(a) Factors such as portfolio size, country development characteristics, and the likely relevance of the evaluation findings to similar issues in other member countries should be considered in making the selection of countries for which a CSPE is to be undertaken. It is desirable to treat each borrowing member equally, and hence to make an effort to undertake CSPEs for all countries to which an institution provides assistance.	
Joint CSPEs	(b) Increasingly, evaluation on a broader scale than the traditional project, sector, or thematic levels will be required, not only to assess results at the country level but also to look more closely at the role of the different institutions in the process. Joint or multi-aid agency CSPEs can provide broader perspective while fostering cross-agency learning and reducing evaluation transaction costs for in-country stakeholders. While the situation varies in each case, MDBs should endeavor to reduce potential bottlenecks by undertaking joint CSPEs within each institution.		As of 2008, the vast majority of CSPEs have been undertaken by individual MDBs. Only a handful has been undertaken jointly by two MDBs, or by MDBs and other development partners. In many cases, joint CSPEs between MDBs have been conducted as parallel exercises, with separate reports. The main benefit of such a joint activity is the reduction in the burden and cost for the recipients.
Mutual Accountability	(c) While some bottlenecks are outside of the control of the evaluators (e.g., different reporting requirements or different country strategy timing), the broader efforts to foster MDB harmonization (e.g., joint MDB country strategies or pooled funding arrangements) are likely to make it more feasible to undertake multipartner CSPEs in the medium term. While multipartner CSPEs are recommended, the decision on whether or not to join forces with partners in a CSPE is best made on a case-by-case basis.	(b) There is also a need for multipartner evaluations of country assistance extending beyond MDBs to include all sources of external assistance, for which the evaluation challenges are significantly greater. Multipartner evaluations of the totality of country assistance should be encouraged. To the extent possible, the GPSs set forth in this report will be applied in such joint evaluation exercises.	

GPS Category (Standards and Elements)	Core GPS	Optional GPS	Notes
A.3. Timing			
Timely CSPEs	(a) A CSPE should be timed to permit the results, recommendations, and lessons identified to feed into the preparation of the next MDB country strategy and to be available to management and the executive board in time for reviewing or approving the new strategy.	(a) Optionally, the results of a CSPE could be provided at a time when the government is willing to make strategic decisions about the use of external assistance.	
A.4. Advance Preparations			

GPS Category (Standards and Elements)	Core GPS	Optional GPS	Notes
Preparatory Steps	(a) CSPEs build on the existing stock of MDB self- and independent evaluations..		Operational personnel should also be encouraged to prepare self-evaluations in a timely manner.
Sector/Thematic Studies		(a) At the discretion of each evaluation unit, sector or thematic studies, special evaluations, or impact assessments may be undertaken to prepare for a CSPE. If sector or thematic evaluations are undertaken in advance of a CSPE, then it is advisable to issue these as separate reports and to discuss them with the government agencies responsible for the particular sectors or thematic areas.	
		(b) Application of the same evaluation criteria and ratings systems at the sector/thematic level as those to be used for the CSPE facilitates the aggregation of sector / thematic assessments at the country level.	
A.5. Coverage			
Time Period	(a) CSPEs should cover a period of assistance that is long enough to witness development results, while providing more emphasis on evaluating recent performance during the current strategy period to ensure that the findings are operationally relevant.		
	(b) Newly initiated, completed, and ongoing assistance activities will be covered in an MDB CSPE.		
Product and Service Coverage	(c) A CSPE will cover the full content of the MDB's program of engagement with the country over the relevant time period. It will cover a series of MDB strategies and assistance in projects, programs, technical assistance, economic and sector work, and knowledge products as well as nonfinancial services—including the role that MDB assistance plays in policy dialogue; processes used in addressing issues in the execution of the program; as well as those used in coordinating, harmonizing, and catalyzing assistance from other development partners, the private sector, and civil society.	(a) In large country cases in which there are too many interventions to cover all of them, a CSPE will draw its inferences from a purposeful sample of an MDB's assistance activities that is representative of the main thrusts of the MDB's strategy and program of assistance.	

GPS Category (Standards and Elements)	Core GPS	Optional GPS	Notes
	(d) By necessity, some areas will be covered in more depth than others. Those areas of focus should be determined based on client needs and on the areas of past programs that can evoke the most important lessons for future strategy.		
Second- or Third Generation CSPEs	(e) Second- and third-generation CSPEs summarize the findings from previous CSPEs and take stock of the extent to which the lessons and recommendations of the earlier CSPEs were utilized.		
	(f) Coverage of the second- (or third-) generation CSPE will overlap with the previous CSPE by a period of a few years to validate end-of-period assessments and to provide continuity with the previous evaluation.		
Limited Scope CSPEs	(g) A limited-scope CSPE may be warranted when an MDB’s role in the country is quite minor, when there are likely to have been few results achieved during the CSPE period, or when there is little likelihood of findings and lessons from the CSPE going beyond what is already known from existing project and program evaluations.	(b) A limited-scope CSPE may also be needed to deliver evaluation findings to meet tight time-sensitive demands.	While recognizing that a full performance assessment of a complex assistance program should not be undertaken in a superficial manner, in special cases a limited-scope CSPE may be appropriate.
Validation Reports	(h) A validation report of a self-evaluation report can be treated as a special category of a limited-scope CSPE. In addition, validation of a country-level self-evaluation can serve to assess whether or not a full CSPE is required to investigate more deeply issues raised in the completion report. Independent validation of the completion reports should be undertaken to encourage internal consistency in the evaluations (often between indicators and evaluative judgments) and can be used to assess the adequacy of the documentation and performance ratings.		

GPS Category (Standards and Elements)	Core GPS	Optional GPS	Notes
A.6. Approach Paper for CSPE			
Specific Evaluation Approach	(a) A CSPE approach (or position) paper will be prepared to define the country-specific evaluation approach, to set out the main evaluation parameters, and to brief the evaluation team and stakeholders within the MDB and the government.		
A.7. Preparation Period			
CSPEs Implementation Period	(a) After the CSPE approach/position paper is approved, an in-depth CSPE will generally be implemented over a period of 6 – 12 months for data collection, analysis, reporting, and review. This should provide sufficient time for an in-depth review of secondary materials and for field visits, while ensuring that findings are delivered in a timely manner.		
A.8. Staffing			
Evaluation Team	(a) An MDB CSPE will generally be led by an experienced evaluator with sufficient experience in MDB operations to understand well the processes involved in formulating country strategies and assistance programs.	(a) To the extent that resources permit, a multidisciplinary team will be employed to undertake the CSPE.	
A.9. Guidelines			
Uniform Guidelines, Quality Control, and Appropriateness	(a) CSPE guidelines will be prepared by each MDB. While the guidelines should provide some latitude to tailor CSPE methods, coverage, and approach to diverse country circumstances, a uniform set of guidelines will be used to explain the CSPE, as an evaluation instrument, to stakeholders in the MDB, the country, and elsewhere. The guidelines will serve to establish a core set of CSPE goals and objectives, methods, evaluative criteria, evaluation questions, procedures, reporting formats, quality control processes, and outreach and dissemination arrangements.		
	(b) If a formal rating is included, then the guidelines should clearly specify the rating criteria and performance assessment methodology. Quality control processes should ensure that the principles set out in the guidelines are strictly adhered to so that performance assessments and other findings will be comparable across CSPEs.		
	(c) While the principles set out in the CSPE guidelines		

GPS Category (Standards and Elements)	Core GPS	Optional GPS	Notes
	should be strictly adhered to, the detailed scope, methods, and approach may need to be tailored to diverse country circumstances and to equally diverse assistance roles that the MDBs play.		

B. Methodology-Related GPS

GPS Category (Standards and Elements)	Core GPS	Optional GPS	Notes
B.1. Methods and Approaches for CSPE			
Overview	(a) A CSPE is premised on the assumption that a series of MDB country strategies and programs can be disaggregated into a contextual diagnosis, strategic and programmatic objectives, and an intervention logic that is amenable to formal evaluation. A typical MDB CSPE exercise begins with an effort to make explicit the causal model implicit in the design of the assistance program. It includes a contextual analysis to identify program objectives; assess the validity of the MDB’s diagnosis (in terms of the relevance of the objectives); and examine the relevance of the MDB’s strategy toward meeting the objectives, including the definition and delivery of the lending and nonlending assistance program.		
	(b) Top-down, bottom-up, and attribution-cum-MDB contribution assessments will be used to garner evidence on the extent to which strategic objectives were achieved and to test the consistency of evaluation findings.		
	(c) The evidence base will then be analyzed, using various techniques, to identify performance determinants and to examine the contribution made by the MDB to the achievement of development results.		
	(d) A set of evaluative criteria is applied to the evidence base to rate or otherwise reach an evaluative judgment about the performance of the country assistance in meeting its goals and objectives (see “B.2 CSPE Evaluation Criteria” below).		

GPS Category (Standards and Elements)	Core GPS	Optional GPS	Notes
	(e) Key findings and lessons are drawn from the performance assessment and provide the foundation for future-oriented recommendations.		
	(f) In MDB CSPE reports, the methodology used is clearly explained to ensure common understanding and to avoid disputes.		
Evaluation Questions	(a) A number of fundamental evaluation questions are defined to guide the assessment of country strategy and program performance. These will include both questions that are standard to all CSPEs, as well as those defined for the specific country case.		
	(b) The CSPE is expected to provide evidence- based answers to these questions. At the discretion of each evaluation unit, standard questions may be similar to the following: • Were the MDB's strategy and program relevant to the development challenges facing the country? • Were suitable instruments of assistance selected to achieve strategic priorities? • Did the MDB assistance achieve its desired objectives? If so, were they achieved efficiently? • Are these achievements sustainable over time? • Was the MDB's assistance effective in producing results, both at the level of individual interventions and at the level of the program as a whole? • What is the overall impact of the MDB's assistance, for example on the economy, on poverty reduction, and on the MDGs? • Did the MDB's assistance contribute to outcomes that will improve the country's capacity to manage the economy, combat poverty, and foster sustainable socioeconomic development? • Was there a suitable division of labor, and were there effective coordination arrangements with other development partners?		

GPS Category (Standards and Elements)	Core GPS	Optional GPS	Notes
	Both the general and the evaluation-specific questions that are asked will be documented in the CSPE report for the readers to be able to judge whether the evaluation team has sufficiently assessed them.		
Counterfactuals		(a) The most accurate measure of an MDB’s contribution is a comparison of the situation prevailing with and without its assistance. In practice, such counterfactuals are difficult to derive and defend for a country program as a whole. These should be used only when they are possible and defensible.	Separate impact evaluations are generally not conducted as part of a CSPE because of the cost, time required, and the limited extent to which the findings can be generalized.
		(b) In some instances, comparison with similar countries can be used as a counterfactual, although these tend to compare performance across countries and not across assistance program outcomes. It may, however, be possible to derive reasonable counterfactuals for specific components of an assistance program, such as cases in which one region was assisted and others were not, or when formal impact evaluations have been undertaken in advance of the CSPE.	
Attribution and Contribution	(a) Formal attribution (i.e., separating the MDB’s role from that of other internal or external players) is extremely difficult in a CSPE because of the multiplicity of factors that affect development outcomes and impacts at the country level. Therefore, the assessment of program results will focus on determining whether the MDB has made a contribution to key results or outcomes that is both plausible and meaningful, and identifying the main drivers of the outcomes. A plausible association of MDB assistance with development results can be assessed by: • characterizing the role played by the MDB in the sector or thematic domain (i.e., lead MDB, main policy interlocutor), • examining the policies and actions of other	(a) In addition, CSPEs will attempt to characterize the nature of the MDB’s contribution to results by assessing the extent to which MDB assistance delivered additional value beyond the financing provided.	

GPS Category (Standards and Elements)	Core GPS	Optional GPS	Notes
	major development partners for consistency with those of the MDB, and • examining evidence that the main outcomes were not achieved primarily due to exogenous events.		
Evaluability	(a) Evaluability, at the country level, is a measure of how well a proposed strategy or program sets out criteria and metrics to be used in its subsequent evaluation. A CSPE will include an assessment of the evaluability of the country strategy(ies) and program(s) of assistance.		Various factors influence the evaluability of country assistance, including the quality of the country diagnostic; the link between that diagnostic and the intervention logic; and the degree to which targets and indicators were specified *ex-ante*, baseline information was collected, outcomes were monitored, and results were reported.
	(b) Evaluability of country strategies and assistance programs can be a serious problem, especially if country strategies are very broad and have goals and indicators far removed from an MDB's contribution; if the intervention logic is not well defined; or if there are large backlogs of projects that should, but do not, have project completion reports. Evaluability constraints can be overcome by: • reviewing strategy, program, and project documents to reconstruct program objectives, indicators, and/or baselines; • retrofitting results frameworks from the reconstructed program logic; • undertaking sector reviews to assess performance of completed and ongoing operations; • collecting before-and-after performance evidence from executing agencies; project files; and, in selected cases, beneficiary surveys; and • concentrating the analysis on key trends in assistance performance for which data exist.		
Multiple Evidence Sources	(a) CSPEs examine quantitative and qualitative evidence from a wide range of both primary and secondary data sources. Differences in the	(a) Formal sample surveys, while less common, can also be used to assess project performance, to solicit feedback	Secondary data include documentation from the MDB and other development partners,

GPS Category (Standards and Elements)	Core GPS	Optional GPS	Notes
	evidence base need to be carefully reconciled and explained. The aim should be to obtain the widest possible breadth of information, to analyze the evidence carefully, and to base findings on information that has been successfully validated from multiple sources.	on the responsiveness of the MDB to key government agencies, and to assess the quality of the MDB's performance as a development partner. Client perception surveys can also be used to provide valuable evidence about MDB performance	government, research institutions, and other outside sources. Primary data are drawn from various sources, including: • interviews with key stakeholders, which are used to validate the key findings and reveal the reasons for particular patterns of performance; • focus group discussions, which are used to address specific issues or obtain beneficiary views; and • field visits to project sites, which are sometimes included to crosscheck information obtained from project files and government reports.
Client Participation	(a) Client participation in the CSPE process encourages respect for the fairness and objectivity of the CSPE, and contributes to early buy-in of the key results and recommendations. MDB CSPEs will endeavor to involve key stakeholders in the CSPE process from the design of the evaluation through its execution to the discussion of its key findings.		However, MDB CSPEs are independent evaluations, so they are not conducted jointly with the country.
Disclaimers	(a) Given the breadth and complexity of the task, and the possible weaknesses in the evidence base, there is only so much that any CSPE can conclusively evaluate. Therefore, the limitations of the CSPE methodology, and its application, should be frankly acknowledged in the evaluation report. This would include factors impinging on the accuracy of the performance assessment and the breadth and depth of the evidence base upon which performance assessments are drawn. This also makes it possible for evaluation clients to establish the degree of precision with which CSPE findings can be interpreted.		

GPS Category (Standards and Elements)	Core GPS	Optional GPS	Notes
B.2. Evaluation Criteria for CSPEs	(a) The performance of a country assistance strategy and program of assistance should be formally assessed using a set of well-defined evaluation criteria. The standard evaluation criteria that are applied to projects and programs can be interpreted and applied to the evaluation of country assistance. For harmonization purposes, relevance, efficiency, effectiveness, sustainability, and impact are considered mandatory criteria.	(a) Positioning, coherence, institutional development, borrower performance, an MDB's performance, and partner coordination are optional criteria.	
Relevance, Positioning, and Coherence	(a) A diagnosis of the evolving country context is used to assess the extent to which an MDB's strategic objectives and assistance program were relevant to the critical constraints affecting the country's long-term socioeconomic development and to the government's policies and strategic priorities, in light of other development partners' strategies, and to assess the consistency of its program with its strategy.	(a) The processes used to maintain relevance, such as an MDB's research and policy dialogue, may also be assessed.	Relevance refers to the degree to which the design and objectives of an MDB's strategy and program of assistance were consistent with the needs of the country and with the government's development plans and priorities.
		(b.1) Positioning may be used to evaluate the design of the country assistance strategy and program. Several subcriteria have been used to assess the extent to which an MDB's assistance was positioned appropriately, including the extent to which assistance: • was concentrated in areas of an MDB's evolved comparative advantage; • built on lessons of past experience; and • was selective/focused on a few sectors to reduce transaction costs and provided a sufficient quantum of assistance in any one area.	Positioning is a measure of how well an MDB responded to (or even anticipated) the evolving development challenges and priorities of the government, built on its comparative advantage, and designed the country strategies and programs in a manner that took into consideration the support available from other development partners.
		(b.2) Coherence may be used to evaluate the design of the country assistance strategy and program. Coherence may be examined along three dimensions:	Coherence refers to the extent to which there were measures aimed at fostering internal and external synergies within an MDB's program. This can

GPS Category (Standards and Elements)	Core GPS	Optional GPS	Notes
		• definition of programmatic focus in terms of anticipated results, • integration across an MDB's instruments in support of program objectives, and • specification of the division of labor with other development partners.	include complementarity between different program elements, the extent to which policies of an MDB are self-reinforcing, and the extent to which external partnerships promote an efficient and effective division of labor in providing assistance that allows for complementarities and synergies with other development partners' programs.
Efficiency	(a) Measuring efficiency is difficult at the overall country program level because of the difficulty of estimating the combined benefit flows of various categories of an MDB's assistance (i.e., policy support, capacity building, or aid coordination). Instead, CSPEs typically draw on proxy indicators of the efficiency of an MDB's support in comparison to cost. This may include indicators related to project/program implementation, for example, of planned versus actual commitments, disbursement patterns, project supervision, projects at risk, design and supervision coefficients, monitoring and evaluation arrangements, implementation problems and their resolution, and other factors affecting program implementation. Ratings accorded to projects, programs, and technical assistance are also used as a proxy for returns-on-investment and timely delivery of services, while economic internal rates of return for major investments may also be reviewed. Various proxies for transaction costs to the government may be assembled and analyzed, including the number of missions per year; the proportion of time that senior government officials devoted to servicing an MDB's missions; and the average amount of time that executing agencies have allocated to the design, implementation, monitoring, and evaluation of MDB-supported assistance activities. Factors affecting the efficiency with which resources are used are identified in an MDB's		Efficiency refers to the extent to which the design and delivery of assistance were most cost effective.

GPS Category (Standards and Elements)	Core GPS	Optional GPS	Notes
	CSPEs.		
Effectiveness	(a) Outcomes are assessed in a CSPE with respect to program objectives at different levels; across similar lending and nonlending projects; within key sectors and/or thematic thrusts; and at broader institutional, macroeconomic, and socioeconomic levels. Drawing primarily on a (bottom-up) analysis of cumulative program performance, CSPEs assess achievement of results both in terms of the extent to which strategic outcomes were achieved, and the extent to which sufficient development progress was made. Results are generally compared in three ways: • before and after the country assistance period being reviewed; • between the country and similar countries (within the same region or at a similar level of development), as appropriate; or • benchmarking against any absolute standards (e.g., the MDGs, costs of capital, rates-of-return).	(a) CSPEs are also uniquely suited to assess the suitability of an MDB's policies in different country contexts, such as compliance and results of safeguard policies, financial management policies, decentralization, human resource policies, relations with civil society, cofinancing policies, adequacy of an MDB's instruments, and responsiveness of an MDB's services to country-specific assistance requirements. Not all MDB's policies can be assessed in all country cases. In an MDB's CSPEs, a distinction will be drawn between those policies whose coverage is mandatory and those whose coverage is optional.	Effectiveness refers to the extent to which the assistance instruments achieved the intentions and objectives set.
	(b) The determinants of an MDB's performance in attaining strategic objectives are identified in the CSPE report		
Sustainability	(a) The degree to which the results of an MDB's assistance are likely to be sustained after the conclusion of the program will be covered by examining the degree to which past interventions have been sustained, identifying risks that could affect benefit flows, and assessing the extent to which policies are in place to mitigate such risks. In assessing the sustainability of benefit flows, a key issue is the extent to which adequate institutional arrangements have been established to further the implementation of program-supported measures. Similarly, factors that negatively affect sustainability, such as fiscal distress or insufficient attention to recurrent financing, may also be assessed.		Sustainability refers to the likelihood that actual and anticipated results will be resilient to risks beyond the program period.

GPS Category (Standards and Elements)	Core GPS	Optional GPS	Notes
Impact	(a) Impact is generally assessed with reference to an MDB’s contribution to the attainment of specified development goals (i.e., macroeconomic balance, socioeconomic conditions, transition impact, MDGs, and other specified national poverty reduction goals and objectives) and to the contribution of an MDB’s assistance individually to the national and/or sector-specific impact objectives established during the programming process. Program impacts will most often be assessed using before-and-after comparisons, and to a lesser extent by comparing performance with similar countries or with internationally accepted standards (e.g., MDGs). Factors exogenous to the program will be examined to distinguish those impacts that can reasonably be associated with the assistance program from those whose proximate determinants lie elsewhere.		Impact refers to an MDB’s contribution to long-term changes in development conditions.
Institutional Development		(a) The extent to which an MDB’s support has helped to develop institutional capacity may be separately assessed (if not part of impact assessment) by examining changes in the performance and governance of public institutions, nongovernment organizations, the private sector, and civil society. Institutional development is more frequently assessed as part of an overall assessment of effectiveness and impact, since capacity building has come to be treated as an integral crosscutting objective of most MDB programs.	Institutional development refers to the extent to which an MDB’s assistance improved or weakened the ability of the country to make more efficient, equitable, and sustainable use of its human, financial, and natural resources, for example through better definition, stability, transparency, enforceability, and predictability of institutional arrangements; and/or better alignment of missions and capacities of organizations with their respective mandates.
Borrower Performance		(a) Borrower performance may be assessed by examining the degree of client ownership of international development priorities, such as MDGs and an MDB’s corporate advocacy priorities; the quality of policy dialogue; and the extent to which the government provided consistent support for MDB-assisted programs. However, borrower	Borrower performance focuses on the processes that underlie the borrower’s effectiveness in discharging its responsibilities, with specific focus on the extent to which the government exhibited ownership of the assistance strategy and program.

GPS Category (Standards and Elements)	Core GPS	Optional GPS	Notes
		performance should not be formally rated.	
MDB Performance		(a) An assessment of an MDB's performance typically considers • the relevance and implementation of the strategy and the design and supervision of its lending interventions; • the scope, quality, and follow-up of diagnostic work and other analytical activities; • the consistency of its lending with its nonlending work and with its safeguard policies; and • its partnership activities. It may also include the extent to which the MDB was sensitive and responsive to client needs and fostered client ownership. The views of operational personnel, the borrower, executing agencies, and other development partners are also typically considered in assessing the MDB's performance.	An MDB's performance focuses on the processes that underlie its effectiveness in discharging its responsibilities as a development partner, including compliance with basic corporate operating principles; consistency with furtherance of its corporate, country, and sector strategies; and its client service satisfaction.
Partnership and Harmonization		(a) Robust partnerships are required to address complex development challenges. In recognition of this, CSPEs examine the extent to which an MDB has been an effective partner in a multistakeholder development assistance effort. This may include an assessment, but not a formal rating, of the MDB's participation in aid agency/partner groups, the extent to which its activities were well coordinated with those of other aid agencies, the degree to which it helped improve the government's capacity for mobilizing and utilizing external assistance, and the manner in which it fostered involvement of all stakeholders (e.g., government, private sector, civil society, nongovernment	Partner coordination refers to the contribution made by an MDB to coordinating external assistance and to building government and country ownership of external assistance processes.

GPS Category (Standards and Elements)	Core GPS	Optional GPS	Notes
		organizations, and other development partners) in the development process. The degree to which the Paris Declaration on Aid Effectiveness principles (i.e., government ownership, alignment with government strategies, results orientation, program approaches, use of country systems, tracking results, and mutual accountability) have been promoted should be covered in the assessment of the MDB's contribution to building robust development partnerships.	
B.3. Performance Rating			
Ratings Principles and Comparability	(a) If a quantitative rating is undertaken, then the rating system should use well-defined criteria and be kept as simple as possible, because ratings that are too numerous or too detailed may confuse the user. Moreover, discussion of the ratings should not distract from the main messages.	(a) A quantitative rating system is generally viewed as a useful component to a CSPE, because it can help to organize and discipline the evaluation and can make the assessment process transparent and uniform across countries.	While there will always be some element of evaluator judgment, strict adherence to CSPE rating guidelines and careful quality control can help to promote ratings that are comparable across CSPEs in those evaluations that include a quantitative rating.
	(b) For MDBs that wish to include ratings, the manner in which the ratings are derived should be clearly stated in MDB CSPE reports, and the summary evidence upon which they were made should be presented along with the rating itself.		
	(c) The limitations of the CSPE rating system should also be frankly acknowledged.		
	(d) Ensuring that CSPE ratings are comparable across CSPEs implies the need for a rating system that is uniform, both in its definitions and in its application in different country cases.		
Rating Criteria	(a) If a quantitative rating is undertaken, the ratings of the mandatory criteria (relevance, efficiency, effectiveness, sustainability, and impact) are considered to be C-GPS.	(a) The ratings of the additional criteria (positioning, coherence, institutional development, borrower performance, an MDB's performance, and partner coordination) are considered to be O-GPS.	
	(b) The ratings for each criterion that is employed should be presented separately so that the results		

GPS Category (Standards and Elements)	Core GPS	Optional GPS	Notes
	of the performance assessment are fully transparent to the evaluation users.		
Rating Subcriteria		(a) For MDBs that quantitatively rate performance, defining subcriteria, if any, in a way that is applicable to specific country cases can help to provide an evaluative framework for more uniform, systematic, and comparable assessment.	MDB evaluators have drawn on a decade of experience in undertaking CSPEs to evolve a set of evaluative subcriteria suitable for assessing country assistance performance in different country settings. A list of CSPE-specific subcriteria for each of the criteria indicated above is provided in Appendix 2. This list is not meant to be either exhaustive or minimal; it reflects many of the factors found to be important determinants of country assistance performance, a subset of which is likely to be suitable in varied settings. An evaluative judgment is required to assess the degree to which chosen subcriteria have been achieved in a particular evaluation.
Weighting Criteria		(a) If overall performance ratings (or headline ratings) are generated—as an optional good practice—then more emphasis should be accorded in the weighting to the results (i.e., effectiveness and impact) of the assistance program and to the sustainability of the net benefits.	

C. Reporting-Related GPS

GPS Category (Standards and Elements)	Core GPS	Optional GPS	Notes
C.1. Findings, Lessons, and Recommendations			
Findings and Lessons	(a) CSPE reports will include evaluation findings that are country-specific, follow logically from the main evaluation questions and analysis of data, and show a clear line of evidence to support the conclusions drawn.		
	(b) CSPEs will identify a few number of lessons that are unambiguously rooted in the evaluation evidence, and have clear operational implications.		
Recommendations	(c) CSPE recommendations will be conveyed constructively in the form of proposals that are actionable within the responsibilities of the users, few in number, country-specific, strategic, operational, and (ideally) not obvious.		
C.2. Reporting and Review			
Reporting	(a) Standard CSPE reporting formats will be used to foster uniformity in coverage and presentation while providing sufficient latitude to tailor the reports to the needs of a particular country case.		
	(b) The report should include coverage of the country context, country strategy and program, program implementation, program outcomes and impacts, partnerships, thematic issues, lessons, and recommendations.		
	(c) The CSPE report will be presented in plain language. It will be evidence- and analysis-based, and will focus on those key issues that could be evaluated conclusively, rather than on all issues that have been examined.		
CSPE Review	(d) For quality control purposes, the draft CSPE will be rigorously reviewed internally by the staff and management of the independent evaluation office, and externally by MDB operations personnel; government stakeholders; and, optionally, by external reviewers. The CSPE review process should also extend to parallel or supporting studies to ensure that they are contextually correct and consistent with the CSPE process.	(a) The revised CSPE report will reflect these comments and acknowledge any substantive disagreements. In cases in which there are such disagreements, the formal views of management, government, external reviewers, and/or the board will be reflected in the final CSPE report.	
C.3. Making Findings Accessible			
Disclosure	(a) It is recommended to publish the findings of		

GPS Category (Standards and Elements)	Core GPS	Optional GPS	Notes
	CSPEs. Publishing the CSPE findings helps to foster learning beyond the immediate client groups and also helps to promote transparency in the evaluation process.		
	(b) To spotlight the diversity with which CSPE findings can be interpreted, CSPE publications will generally include the formal views of management, government, and the board		
Dissemination		(a) It often requires considerable effort to ensure that the CSPE findings are disseminated beyond a small group of senior MDB and government officials. Presentations to parliament, public seminars, consultation workshops, and press briefings are some of the ways in which CSPE findings can be more widely disseminated.	
		(b) Summarizing the CSPE in a readily accessible form (such as an evaluation précis) and translation of CSPE findings into the local language can contribute to wider dissemination of findings and results.	
C.4. Generalizing Findings and Tracking Recommendations			
Generalizing CSPE Findings:	(a) The findings from CSPEs will be summarized and used for comparative purposes in the annual and/or biannual reviews of evaluation findings prepared by the independent evaluation offices. Using CSPEs for comparative purposes helps foster a more general understanding of the factors that influence country assistance performance.		
Tracking Recommendations		(a) Tracking and reporting on the progress by which CSPE findings, lessons, and recommendations are actually utilized by the MDB helps to facilitate institutional learning practices. This can be accomplished through either recommendation tracking systems or periodic reviews of the utilization of CSPE findings and recommendations	

Annex V.1: Evaluation Cooperation Group Progress Benchmarking Template

ECG members agree that periodic assessments will be undertaken to assess the extent to which the GPS are being applied. The GPS have been summarized in tabular form below to assist each participating MDBs in progress benchmarking.[53] Some time will be required to adjust member practices to GPS, and thereafter to conduct at least one CSPE under the new GPS' regime. Benchmarking of MDBs against the CSPE GPS is under consideration.

(GPS Category	Core GPS Description	Degree of MDB Alignment[a]	Optional GPS Description	Degree of MDB Alignment	Remarks[b]
A. Process-Related GPSs					
A.1. CSPE Goals, Objectives, Client Responsiveness, and Unit of Analysis	(a) Provide credible and useful information on the MDB's performance at the country level (b) Used for both accountability and lesson-learning purposes (c) Designed to meet information requirements of main target clients (d) Focus on evaluating the results of the MDB's assistance, with the country strategy(ies) as the main reference point				
A.2. Country Selection and Mutual Accountability	(a) Countries selected are those in which the findings and lessons will be most beneficial to the MDB and the country. (b) Efforts made to reduce potential bottlenecks in undertaking joint MDB CSPEs (c) Decision to pursue a multipartner CSPE made on a case-by-case basis		(a) Covering all countries and treating all borrowers equally (b) Multipartner CSPEs extending beyond the MDBs to include all sources of external assistance to a country encouraged		
A.3. Timing	(a) CSPE timed to feed into the preparation and review of the MDB's new country strategy		(a) Could also be timed to contribute to strategic decision making of the government		

[53] A baseline of CSPE practices for AfDB, AsDB, IADB, and WBG is included in the self-assessment questionnaires provided in the report cited in footnote 8.

(GPS Category	Core GPS Description	Degree of MDB Alignment[a]	Optional GPS Description	Degree of MDB Alignment	Remarks[b]
A.4. Preparatory Steps	(a) Evaluations of key projects, programs, and technical assistance scheduled to precede the CSPE		(a) Sector/thematic studies or impact assessments scheduled to precede a CSPE. (b) Application of the same evaluation criteria in sector/thematic studies as in the CSPE facilitates their use.		
A.5. Coverage	(a) Coverage long enough to see results, but more emphasis put on the current strategy period (b) Newly initiated, completed, and ongoing operations covered (c) Full content of the MDB's assistance covered (d) Depth of coverage depends on client needs and those areas most likely to evoke lessons for future strategy (e) For second- or third-generation CSPEs, previous CSPE findings summarized, and use of previous CSPE lessons and recommendations assessed (f) Subsequent CSPEs will have an overlap in the period covered of a few years. (g) CSPEs may have limited scope if the MDB's role is minor, if there were few results, or if there is little likelihood of findings and lessons of broader impact. (h) Completion reports of country strategies independently validated; if the completion and validation reports are comprehensive and apply CSPE criteria, they may serve as a limited-scope CSPE.		(a) In large country cases, a representative sample of assistance activities assessed (b) A limited-scope CSPE may also be needed to deliver evaluation findings to meet tight time-sensitive demands.		
A.6. CSPE Approach Paper	(a) A CSPE approach (or position) paper prepared for each CSPE				
A.7. CSPE Preparation Period	(a) A full CSPE implemented over 6–12 months				

(GPS Category	Core GPS Description	Degree of MDB Alignment[a]	Optional GPS Description	Degree of MDB Alignment	Remarks[b]
A.8. Staffing	(a) CSPE teams headed by an experienced evaluator with sufficient experience in MDB operations		(a) A multidisciplinary team engaged to undertake the CSPE		
A.9. Guidelines	(a) Each MDB will have CSPE guidelines that set out CSPE goals and objectives, methods, evaluative criteria, evaluation questions, procedures, reporting formats, quality control processes, and outreach and dissemination arrangements. (b) Quality control procedures will ensure that guidelines are followed. (c) While guidelines will be adhered to, the actual methods, scope, and approach may be tailored to the country setting.				
B. Methodology-Related GPS					
B.1. CSPE Methods and Approaches					
Overview	(a) CSPE methods include steps to make the causal model explicit in the country strategy, analysis of country context, assessment of the validity of the MDB's diagnosis, and analysis of the strategy and program relevance in design and delivery. (b) Top-down, bottom-up, and attribution-cum-contribution assessments used to assemble information on performance in achieving strategic objectives (c) Evidence base analyzed to identify performance determinants (d) Evaluation criteria applied to assess performance in multiple dimensions (e) Findings and lessons drawn, and future-oriented recommendations provided (f) Methods explained in the CSPE report				
Evaluation Questions	(a) General and country-specific evaluation questions posed to guide the assessment (b) Evaluation questions documented in the CSPE report				

(GPS Category	Core GPS Description	Degree of MDB Alignment[a]	Optional GPS Description	Degree of MDB Alignment	Remarks[b]
Counterfactuals			(a) Counterfactuals should be used only when they are possible and defensible. (b) Counterfactuals can be proxied through comparisons with similar countries, examination of those parts of the program for which a counterfactual can be more clearly identified, or for those parts of the program for which prior impact evaluations have been conducted.		
Attribution and Contribution	Since formal attribution is difficult to determine, assessment of program results will focus on determining whether the MDB has made a contribution to key results or outcomes, and identifying the main drivers of the outcomes.		(a) To characterize the nature of the MDB's contribution to results, the extent to which its assistance delivered additional value beyond the financing provided will be assessed.		
Evaluability	(a) CSPE includes an assessment of the evaluability of the MDB's strategy and program of assistance. (b) Evaluability constraints overcome by reconstructing the program logic, retrofitting results frameworks, drawing on available information sources, and collecting performance information				
Multiple Evidence Sources	(a) CSPE draws on the widest possible breadth of primary and secondary sources of information, and bases findings on information that has been successfully validated from multiple sources.		(a) Use of client perception surveys can provide evidence about the MDB's performance.		

(GPS Category	Core GPS Description	Degree of MDB Alignment[a]	Optional GPS Description	Degree of MDB Alignment	Remarks[b]
Client Participation	(a) Participation of key stakeholders in the CSPE process encouraged				
Disclaimers	(a) Limitations of the methodology and its application frankly acknowledged in the CSPE report				
B.2. CSPE Evaluation Criteria	(a) Relevance, efficiency, effectiveness, sustainability, and impact considered mandatory criteria		(a) Positioning, coherence, institutional development, borrower and MDB performance, and partner coordination considered optional criteria		
Relevance, Coherence, and Positioning	(a) Relevance examined by assessing if the MDB's strategy and assistance program were consistent with the country context and the government's strategic priorities		(a) The MDB's processes used to maintain relevance assessed (b) Criteria such as positioning and coherence used to assess the degree to which the design of the strategy and program harnesses positive synergies and builds on the MDB's core competence		
Efficiency	(a) Efficiency assessed using indicators affecting cost-effectiveness, transaction costs, portfolio performance, monitoring and evaluation arrangements, and other project/program implementation				
Effectiveness	(a) Extent to which strategic outcomes were achieved and sufficient development progress was made used to assess program effectiveness (b) Determinants of performance in achieving the MDB's objectives identified		(a) The MDB's contribution to broader corporate objectives assessed, but distinction drawn between those		

(GPS Category	Core GPS Description	Degree of MDB Alignment[a]	Optional GPS Description	Degree of MDB Alignment	Remarks[b]
			thematic issues whose coverage is mandatory and those whose coverage is optional		
Sustainability	(a) The degree to which the results of the MDB's assistance are likely to be sustained after the conclusion of the program assessed				
Impact and Institutional Development	(a) Impact assessed relative to national goals and to program-specific goals and targets		(a) Extent to which the MDB has helped to develop institutional capacity separately assessed if not part of impact assessment		
Borrower Performance			(a) Borrower performance, particularly the degree of program ownership, assessed but not formally rated		
MDB Performance			(a) The MDB's performance formally assessed, including its responsiveness to client needs		
Partnership and Harmonization			(a) CSPE examines the extent to which the MDB has been an effective partner in a multistakeholder development assistance effort.		
B.3. Performance Rating *Ratings Principles and Comparability*	(a) If quantitative rating is undertaken, the rating system should use well-defined criteria and be as simple as possible. (b) The manner in which ratings are derived is stated in the report. (c) Limitations of the rating system are		(a) A quantitative rating system is used to make the assessment process transparent and uniform across		

(GPS Category	Core GPS Description	Degree of MDB Alignment[a]	Optional GPS Description	Degree of MDB Alignment	Remarks[b]
	acknowledged. (d) Rating system is uniform.		countries.		
Rating Criteria	(a) If a quantitative rating is undertaken, ratings of the mandatory evaluation criteria (relevance, efficiency, effectiveness, sustainability, and impact) needed (b) If a quantitative rating is undertaken, ratings accorded for each criterion presented separately to make the performance assessment transparent		(a) If a quantitative rating is undertaken, ratings of the additional evaluation criteria (positioning, coherence, institutional development, borrower and MDB performance, and partner coordination) considered optional		
Rating Subcriteria			(a) Defining subcriteria, if any, in a way that is applicable to specific country cases can help to provide an evaluative framework for more uniform, systematic, and comparable assessments.		
Weighting Criteria			(a) If an overall rating is		
C. Reporting-Related GPSs					
C.1. Findings, Lessons, and Recommendations	(a) Evaluation findings are country-specific, evidence-based, and follow from the evaluation questions. (b) Lessons are few in number and evidence-rooted, and have operational implications. (c) Recommendations are few in number, constructive, actionable, strategic, operational, and not obvious.				

(GPS Category	Core GPS Description	Degree of MDB Alignment[a]	Optional GPS Description	Degree of MDB Alignment	Remarks[b]
C.2. Reporting and Review	(a) Uniform formats followed with latitude to tailor to the country case (b) Report covers country context, country strategy, program implementation, program outcomes and impacts, partnerships, thematic issues, lessons, and recommendations. (c) Report presented in plain language and covers those issues that could be conclusively evaluated. (d) Draft report and supporting studies rigorously reviewed internally and externally.		(a) Where there are substantive disagreements during the review process, these will be reflected in the final CSPE report.		
C.3. Making Findings Accessible	(a) CSPE findings published (b) To spotlight the diversity with which CSPE findings can be interpreted, CSPE publications will include formal views of management, government, and the board.		(a) Outreach events may be held to boost the dissemination of CSPE findings. (b) A précis or other summary publication may be issued and findings translated into the local language to make CSPE findings more accessible.		
C.4. Generalizing Findings and Tracking Recommendations	(a) Annual and/or biannual reviews of evaluation findings summarize and compare CSPE findings		(a) Recommendation tracking systems or periodic reviews of the utilization of CSPE findings and recommendations prepared to track CSPE use		

CSPE = country strategy and program evaluation, GPS = good practice standard, MDB = multilateral development bank.

[a] Alignment refers to the extent to which the MDB evaluation practice is fully, partly, or not harmonized with the relevant GPSs.

[b] The remarks section may be used to explain the reasons for divergence between the GPSs and MDB practice.

Source: GPS on CSPE, 2008. Manila. Appendix 1.

Annex V.2: Subcriteria for Evaluating Country Strategies and Programs

1. What follows is a suggested list of possible subcriteria that multilateral development bank (MDB) evaluators can draw from in tailoring the interpretation of evaluation criteria to the circumstances merited by each particular country case. This is neither a comprehensive nor a minimal checklist. The subcriteria listed here have been found to be important determinants of country assistance performance in MDB evaluations. They can be used to select and define the subcriteria employed in evaluating specific country cases. This is aimed at providing the flexibility required in a country evaluation so that the evaluative criteria are interpreted in a way that is most suitable, given varying country contexts, assistance roles, and data availability.

2. These subcriteria are divided into two groups. The first group belongs to standard evaluation criteria that can be applied to the program as a whole, or to particular components (e.g., sectors or themes). The second group belongs to additional evaluation criteria. For each, an evaluative judgment is required to assess the degree to which each chosen subcriterion has been achieved.

Standard Evaluation Criteria

A. Relevance: *the degree to which the design and objectives of the MDB's strategy and program of assistance were consistent with the needs of the country and with the government's development plans and priorities.*

- Based on a valid diagnosis of the context for external assistance
  - → development context thoroughly reviewed
  - → adequate assessments of key sectors and thematic areas of the MDB's proposed intervention
  - → candid review and assessment of government policies and strategies
  - → robust consultative process to identify and validate priorities
  - → careful assessment of feasibility of using country systems
  - → careful review of lessons of past experience
  - → informed understanding of factors driving aid effectiveness
- Consistency with country's long-term development requirements (for each major objective)
- Consistency (i.e., alignment) with government's development (or poverty) strategy and priorities (for each major objective)
- Designed in a manner consistent with government's institutional capacity to absorb external assistance
- Consistency with global agreements (e.g., MDGs, Paris Declaration commitments, or World Trade Organization (WTO) regulations) for each major strategic objective
- Consistency with the MDB's corporate policy and strategy (for each major objective)
- Importance of program objectives addressed to meet critical development constraints (by category, such as macroeconomic management, structural reform, sector reform, private sector development, institutional development, human development, environmental reform, and infrastructure development)
- Any important objectives that, in hindsight, should have been pursued, but, in the end, were not (i.e., were any important development issues omitted or ignored in the diagnosis?)
- Program formulation and design were relevant to achieving objectives
  - → adequacy of external financing for program operation

- → extent and appropriateness of medium-term framework
- → consistency and coherence of the program logic (e.g., identification of goals to be achieved; specific purpose[s] of the MDB's assistance; and program measures, their expected outputs, outcomes, and impacts, together with key assumptions and risks to performance all identified)
- → appropriate assistance instruments selected (e.g., assistance properly sequenced to reach targets, internally consistent, realistic/feasible, manageable, and with clearly defined targets and objectives)
- → social consequences assessed, and suitable mitigation measures incorporated in overall program design
- → performance risks (both internal and external) adequately identified, and suitable strategies for managing risk incorporated
- → realistic time frame for results to be delivered, given institutional and other constraints

- Extent to which sector and thematic objectives were sufficient to achieve a level of critical mass, balanced among objectives, selective, and focused
- Extent to which dialogue and consultation ensured effective ownership of the program by government and by society at large
- Degree to which the MDB's program was built on lessons from past experience, was sufficiently focused and selective, and drew on areas of its core competency
- The MDB's program took into consideration, and was harmonized with, assistance provided by other development partners
- The MDB's responsiveness in designing and then adapting the assistance strategy to fundamental changes in client circumstances throughout the implementation period
- Extent to which the assistance strategy and program maintained relevance to the client's development constraints and priorities over time
- Assistance strategy and program could be readily evaluated
    - → targets well defined, links traced, baseline values provided, and performance targets specified
    - → reporting, monitoring, and evaluation responsibilities assigned, and funding provided
    - → knowledge gaps identified, and actions identified for securing information needed for decision making included

B. Efficiency: *the extent to which the design and delivery of assistance were most cost-effective.*

- Readiness for implementation of all products and services was secured
- Products and services were delivered in a timely manner
- Extent to which strategic objectives were achieved on time
- Were benefits gained from early completion of assistance (or costs incurred from late completion)?
- Benefits of major interventions are, or are expected to be, substantial, as demonstrated by
    - → positive economic rates of return for major investments
    - → positive financial rates of return and/or return on equity for MDB-supported private investments
    - → major policy or institutional reforms were undertaken that did ease critical constraints to improved socioeconomic performance and poverty reduction

  → unambiguous evidence that benefits reached the poor
- Debt assumed and adjustment costs from MDB-supported reforms were relatively low compared with value arising from achievement of socioeconomic objectives (i.e., social benefits likely to exceed social costs)
- Overall program financing was provided in a timely manner through
  → financing provided in sync with external financing requirements
  → reasonable time for project design, negotiation, and effectiveness
  → disbursements took place according to plan
- Costs of providing assistance were similar or less than those in comparator country programs and were kept in line with the MDB's norms
- Unit costs were reasonable for major investments
- Transaction costs of providing assistance were modest (in terms of time spent preparing projects, number of missions undertaken, extent to which efforts were made to combine or hold joint missions with other development partners, and time spent by key government officials in design and oversight of the MDB's program)
- Public expenditures made adequate provision to meet government's portion of program counterpart costs, and sufficient financing was provided for future recurrent cost requirements

C. Effectiveness: *the extent to which the assistance instruments achieved the intentions and objectives set.*

- Degree to which activities anticipated in strategy and program were actually undertaken
- Sufficient interventions were undertaken to generate outputs and outcomes identified in country strategy and/or program
- Performance of portfolio as a whole was satisfactory in comparison with MDB-wide averages
- Extent to which major issues arose during execution and were (or were not) resolved
- Extent to which main assistance program objectives achieved progress toward each of their stated objectives
- Extent to which results defined under country assistance program were actually achieved
- Extent to which there were major shortcomings, such as unintended social costs or environmental damage, in achieving program objectives
- Performance as assessed by rating of the MDB's projects (both self- and independent ratings) in terms of achievement of major objectives
- Project evaluation judgments regarding achievement of development objectives have verifiable claims
- Extent to which achievement of program objectives demonstrated best practices in some areas
- Extent to which factors beyond government's control influenced the outcome of program (including changes in world markets, natural calamities, war/civil disturbance)
- Were other performance assessments reviewed and presented for major components of the MDB's assistance (including those whose findings contradict the evaluation)?
- Extent to which actual performance met or surpassed benchmarks for financial performance of similar categories of private investment (for private sector operations)

D. Sustainability: *the likelihood that actual and anticipated results will be resilient to risks beyond the program period.*

- Absence of major policy reversals
- Continued borrower commitment to assistance program objectives demonstrated through postprogram implementation of related measures
- Sociopolitical support for main objectives of assistance program
- Adequacy of institutional arrangements for implementing agreed upon reforms and program measures
- Conducive macroeconomic and political setting (i.e., stable and supportive)
- Continued need for (i.e., ongoing relevance and value of) the results and benefits
- Ownership by government and other key stakeholders
- Financial capacity to address recurrent costs
- Degree of resilience to risk of development benefits of country assistance program over time, taking into account:
  → technical resilience
  → financial resilience (including policies on cost recovery)
  → economic resilience
  → social support (including conditions subject to safeguard policies)
  → environmental resilience
  → ownership by government and other key stakeholders
  → institutional support (including a supportive legal/regulatory framework and organizational and management effectiveness)
  → resilience to exogenous effects such as international economic shocks or changes in political and security environments

E. Impacts: *the MDB's contribution to long-term changes in development conditions.*

- Anticipated and unanticipated (positive and negative) impacts identified and adjusted to take into consideration unexpected shocks or other factors beyond program's control, such as
  → country's macroeconomic balance
  → country's economic performance
  → poverty reduction
  → social development
  → governance
  → environmental sustainability
  → gender equality
  → regional cooperation
  → transition from central planning to market economy
  → other major social, political, or institutional changes in context
- Extent to which the program has improved the government's capacity, in key sectors and thematic areas, to make effective and efficient use of its human, financial, and natural resources
- Commercial performance of the MDB's private sector operations, i.e., degree to which these have had wider impacts on private sector development and extent to which these have catalyzed private sector investment in the country
- Anticipated and unanticipated impacts from major projects or programs identified for illustration of magnitude and pattern of intervention effects (e.g., from impact studies or beneficiary surveys)

- Evidence that impacts attributable to country program have been, to the extent feasible, isolated from those caused by other factors
- The program's additional contribution to development impacts (e.g., delivering relevant knowledge or advice, catalyzing change, and fostering more effective use of external resources)
- Degree to which the MDB's assistance makes a meaningful contribution to the government's efforts to foster achievement of the following MDGs:
  → reduce the proportion of people living in extreme poverty by half between 1990 and 2015
  → enroll all children in primary school by 2015
  → make progress toward gender equality and empowering women by eliminating gender disparities in primary and secondary education by 2015
  → reduce infant and child mortality rates by two thirds between 1990 and 2015
  → reduce maternal mortality ratios by three quarters between 1990 and 2015
  → provide access for all who need reproductive health services by 2015
  → implement national strategies for sustainable development by 2005 so as to reverse the loss of environmental resources by 2015

Additional Criteria

A. Positioning: *a measure of how well the MDB responded to (or even anticipated) the evolving development challenges and priorities of the government, built on its comparative advantage, and designed the country strategies and programs in a manner that took into consideration the support available from other development partners.*

- Country priorities and the MDB's corporate priorities were aligned
  → country goals and the MDB's corporate goals were aligned
  → strategic pillars were aligned to contribute to country strategic objectives
  → strategic gaps and risks were identified and agreed upon with the government
- Timing and scope of the MDB's engagement were in what turned out to be major development priorities of country
- Program was results-oriented, coherent, and translated strategy into appropriate operations, which collectively addressed critical development constraints
- Strategic focus was appropriate by sector, target group, and geographic area
- Program provided critical mass of assistance, sufficient to generate sustained results
- Mix of lending and nonlending services, as well as operational approaches, were tailored to the particular conditions of the country
- Productive relationships were forged with other development partners within the wider framework of development cooperation in the country
- The MDB was well positioned to respond effectively to country priorities
  → the MDB was structured, staffed, and managed to respond effectively to client requests
  → institutional arrangements fostered the generation and use of new knowledge to spur innovation
  → assistance was managed for delivery of development results
  → corporate safeguards were adhered to
- Results were delivered, and the most strategic opportunities for assistance were exploited effectively

B. Coherence:[1] *the extent to which there were measures aimed at fostering internal and external synergies within the MDB's program; this can include complementarity between different program elements, the extent to which the MDB's policies are self-reinforcing, and the extent to which external partnerships promote an efficient and effective division of labor in providing assistance that allows for complementarities and synergies with other development partners' programs.*

- Country priorities served to establish main development objectives
- Country strategies were realistic for forging progress toward selected development objectives and were aligned with, and supportive of, implementation of national development strategies and policies
- Country assistance program was designed to make a substantial contribution to achievement of defined objectives
- Choice of assistance, across objective area, included measures that would be innovative and have positive synergies and demonstration spillovers and foster complementary activities, so that value of program as a whole would be greater than sum of its individual parts
- Strategies and assistance choices were aligned with and supportive of assistance provided by other development partners in an effective division of labor
- Assistance instruments were chosen, and effectively integrated, to ensure that response to development challenges was sufficient, complete, and cohesive
- Choice of sectors, regions, and target groups was consistent with needs identified to meet program objectives

C. Institutional Development: *the extent to which the MDB's assistance improved or weakened the ability of the country to make more efficient, equitable, and sustainable use of its human, financial, and natural resources, for example through better definition, stability, transparency, enforceability, and predictability of institutional arrangements; and/or better alignment of missions and capacities of organizations with their respective mandates.*

- Contribution toward improving/strengthening capacity of public institutions to ensure stable, transparent, enforceable, and predictable execution of their mandates:
  - → soundness of economic management
  - → structure of public sector, and, in particular, civil service
  - → institutional soundness of financial sector
  - → soundness of legal, regulatory, and judicial systems
  - → extent of monitoring and evaluation systems
  - → effectiveness of aid coordination
  - → degree of financial accountability
  - → informal norms and practices that govern social and economic interactions
  - → extent of building nongovernment organization capacity
  - → level of social and environmental capital
- Contribution toward improving organizational capacity (in planning, policy analysis, skills upgrading, public awareness building and consultation, management, restructuring, decentralization, management of information systems, financial controls, financial restructuring, regulatory enforcement, and agency governance)

[1] Coherence is a separate evaluative criterion used by the Office of Evaluation and Oversight of the Inter-American Development Bank. It is assessed as part of positioning by the Operations Evaluation Department of the Asian Development Bank.

- Contribution toward improving private sector capacity (i.e., improving rules of the game for efficient, broad-based private sector development)
- Contribution to improving stability, diversity, and growth potential of financial sector services
- Contribution to improving nongovernment organization and civil society capacity
- Contribution toward improving governance of public sector (i.e., transparency, checks and balances, public participation, improved fiduciary policies and practices, and accountability in discharge of public duties)
- Extent to which capacity has been developed within the government to manage formulation and implementation of suitable public policies and programs

D. **Borrower Performance:** *focuses on the processes that underlie the borrower's effectiveness in discharging its responsibilities, with specific focus on the extent to which the government exhibited ownership of the assistance strategy and program.*

- Shared ownership of the MDB's country strategy and program
- Maintained high-level dialogue with the MDB's personnel and management
- Consulted with civil society and other stakeholders on program implementation
- Supported high-quality preparation of MDB-assisted projects:
  → degree of ownership and involvement in identification and design
  → political support for project-related reforms secured
  → adequate institutional arrangements for program implementation
- Provided sufficient counterpart funds and project personnel
- Followed procurement and safeguard (i.e., resettlement, environmental, indigenous peoples, and fiduciary) guidelines
- Carefully supervised project implementation
- Engaged in high-quality dialogue on policy matters with the MDB
- Implemented policy reforms, agreed upon between the MDB and government, in a timely manner
- Provided policy framework supportive of effective aid utilization (i.e., supportive macroeconomic policies and complementary structural and sector policies)
- Fostered public outreach, disclosure, and awareness building throughout program implementation
- Provided results-based monitoring, evaluation, and reporting

E. **MDB Performance:** *focuses on the processes that underlie the MDB's effectiveness in discharging its responsibilities as a development partner, including compliance with basic corporate operating principles; consistency with furtherance of the MDB's corporate, country, and sector strategies; and its client service satisfaction.*

- Quality of strategy and program at entry:
  → appropriate degree of selectivity
  → grounding in recent economic and sector work
  → adequate economic and financial rationale
  → adequate risk assessment
  → realistic assessment of financial requirements and borrowing capacity
  → incorporation of lessons identified in past evaluations
  → adequate institutional analysis
  → adequate poverty, social (including gender), environmental, and stakeholder analysis

- → incorporation of monitoring and evaluation indicators and reporting procedures
- → focus on areas of MDB comparative advantage
- → appropriate mix of assistance instruments selected
- → assistance strategy and program was suitable, given country context and institutional capacity of the government

- Quality of the MDB's supervision:
  - → degree to which supervision focused on achieving objectives
  - → degree to which civil society participation was fostered in program implementation
  - → problems identified during implementation were expeditiously assessed and resolved
  - → adequate resources devoted by the MDB to supervision
  - → attention paid to monitoring and evaluation data and processes
  - → quality and timeliness of self-assessment (i.e., country strategy completion reporting)
- Quality of other services:
  - → built client ownership of the assistance program
  - → built strong links between strategy and analytical and advisory services
  - → provided high-quality knowledge products
  - → maintained high-quality dialogue with government and civil society
  - → maintained high quality at entry for new projects
  - → explained and provided training in its policies, safeguards, and procedures
  - → provided personnel with appropriate skills mix to develop strategy and program
  - → strengthened the government's capacity for financial management and accountability
  - → enforced compliance with procurement guidelines, audit requirements, and other project cost controls
  - → managed portfolio effectively
  - → provided timely notice to the board of fundamental changes in its strategy
  - → solicited feedback on, and was responsive to, requests for ways of improving its performance
  - → provision of necessary long-term financing
  - → provision of suitable risk mitigation services

F. **Partner Coordination:** *the contribution made by the MDB to coordinating external assistance and to building government and country ownership of external assistance processes.*

- Degree to which assistance fostered government leadership of aid coordination
- Degree to which assistance built the government's capacity to plan its public investment and to mobilize and manage external assistance (including debt management) effectively
- Degree to which the MDB played a role in catalyzing or otherwise inspiring other stakeholders to cooperate toward achieving common development results
- Degree to which policies and strategies pursued by other partners were consistent with those pursued by the MDB (i.e., if there were major conflicts or inconsistencies, were steps taken to resolve these?)
- Extent to which assistance effort played catalytic role in resource mobilization
- Degree to which the MDB coordinated and mobilized aid resources effectively

- Degree to which the MDB served as an effective aid partner in terms of knowledge sharing, support for and participation in multipartner initiatives, design of complementary assistance initiatives, assistance provided to other partners to resolve problems of wider concern, and active participation in aid coordination arrangements

Sources: African Development Bank. 2004. *Guidelines for Country Assistance Evaluation* (prepared by O. Ojo). Tunis; Asian Development Bank. 2006. *Guidelines for the Preparation of Country Assistance Program Evaluation Reports*. Manila; Inter-American Development Bank. 2003. *Protocol for the Conduct of Country Program Evaluations*, Revised Version, Report RE-271-1. Washington, DC; Organisation for Economic Co-operation and Development–Development Assistance Committee. 1999. *Evaluating Country Programmes. Report of the Vienna Workshop*. Paris; World Bank. 2003. *Country Evaluation Guidelines* (Internet version) and *Country Questionnaire.* Washington, DC; and 2005. *Country Assistance Evaluation Retrospective: An OED Self-Evaluation*. Washington, DC.

B. Self-Evaluation Principles

VI. GPS on Self-Evaluation

Contents

Background

42. The three approved GPS on evaluation methods[55] all contained similar discussions of self-evaluation. Accordingly, for this GPS Big Book this topic has been addressed in a separate GPS. According self-evaluation full weight as an important part of the overall evaluation system is consistent with the expressed view and initiatives of members on encouraging management towards greater harmonization of self-evaluation processes that is expected to broaden the coverage and quality of self-evaluation activities.

43. Self-evaluation of public sector operations is manifested through the preparation of completion reports (CRs) while in private sector operations it is through the expanded annual supervision reports (XASRs). For country strategy and program (CSP), self-evaluation generally takes three main forms: (i) brief summaries of lessons identified, which are included in a new country strategy; (ii) a country strategy completion report; or (iii) a country strategy progress report or midterm review. Self-evaluations are generally validated by the independent evaluation offices to ensure consistency and to encourage candid and critical evaluation by the operations departments.

44. **Rationale for GPS on self-evaluation.** The inclusion of self evaluation in the three approved methodological GPS stemmed from the discussion[56] on the post-Monterrey environment wherein MDBs' management adopted the results paradigm in organizing their administrative activities, including planning, results measurement, monitoring and self-evaluation. This was seen as an opportunity for fine-tuning the relationship between self- and independent evaluation. In 2003, members agreed that harmonization in self-evaluation practices, operational policies and processes facilitate the harmonization of independent evaluation. After more than a decade, members continue to recognize the importance of self evaluation to independent evaluation.[57]

45. **Formulation Process.** The GPS on self evaluation formed part of the formulation of the three approved GPS and did not have separate formulation processes. Brief description of the formulation processes for each category of operations (i.e., public and private sector operations and CSP) is provided in the preceding chapters.

46. **Organization.** The GPS for self-evaluation are presented by category of operations. Standards for self-evaluation of public sector operations cover ensuring evaluability, preparation of CRs, involvement of CED in self-evaluation, and harmonization of self- and independent evaluation. For private sector operations, the standards define the scope of self- or indirect evaluation[58] which includes the executor of the evaluation, and report preparation. And lastly, standards for self-evaluation of CSP are limited to those most critical for the quality of independent evaluation, excluding, for example, topics such as the processing and review of self-evaluation reports.

---

[55] GPS on Evaluation of Public and Private Sector Operations, and Country Strategy and Program Evaluation (CSPE).
[56] Minutes of ECG meeting in Fall 2002.
[57] Minutes of ECG meeting in Spring 2012.
[58] GPS on Evaluation of Private Sector Operations has adopted some flexibility in using the terms "indirect and direct evaluation" to mean self- and independent evaluation respectively, to incorporate the terminology used in the different IFIs (This is based on clarification of the consultant for GPS on Private Sector Operations).

47. The GPS for self-evaluation of public and private operations are grouped by standards defined by a number of elements that correspond to the evaluation principles (EPs). Each EP is supported by one or more operational practices (OPs) that describe the policies and procedures that would normally need to be adopted in order to be deemed consistent with the respective EP. The standards for self-evaluation of CSPs are also defined by the corresponding elements which are defined by core GPS. A total of 6 Standards and 10 Elements form the GPS on self evaluation. The summary of the standards and elements are in the Table below while details of EPs and OPs are presented in the next section.

Summary of Standards and Elements on EPs and Number of OPs on Self-Evaluation

Evaluation Principles		Number of OPs	Page
Standards	**Elements**		
A. Public Sector Operations			
1. Self-Evaluation	A. Ensuring Evaluability	11	151
	B. Completion Reports	9	152
	C. CED Involvement in Self-Evaluation	1	152
	D. Harmonization	1	153
No. of Standards: 1	No. of Elements: 4	No. of OPs: 22	
B. Private Sector Operations			
Planning and Executing a Project Evaluation Program			
1. Scope of indirect evaluation	A. Self-Evaluation B. IFI Reporting	1	153
	C. XASR Research	1	153
	D. Transparency	1	153
No. of Standards: 1	No. of Elements: 4	No. of OPs: 3	
C. Country Strategy and Program (CSP)			
Process-related GPS			
1. Advance preparations	A. Preparatory Step	1	154
2. Coverage	B. Self-Evaluation Reports	1	154
No. of Standards: 2	No. of Elements: 2	No. of OPs: 2	
Total No. of Standards: 4	**Total No. of Elements: 10**	**Total No. of OPs: 27**	

Source: Detailed matrices in next section of this Chapter.

GPS on Self-Evaluation

A. Public Sector Operations

Evaluation Principles (Standards and Elements)	Operational Practices	Notes
1. Self-Evaluation *A. Ensuring evaluability:* IFI policy requires that project design include a minimum set of elements to ensure evaluability.	1.1 IFI policy requires that project design include a statement of objectives that is specific, realistic, has temporal characterization, is measurable, agreed upon, and clearly identifies the beneficiaries and those responsible for their achievement.	The outcomes in the objective statement are called "impacts" by some IFIs.
	1.2 The statement of objectives is based on a problem or diagnostic statement.	
	1.3 The objectives statement focuses on outcomes for which the project can reasonably be held accountable. It does not encompass objectives beyond the purview of the project, nor is it a restatement of the project's components or outputs.	
	1.4 To the extent that higher-level social and economic objectives and corporate goals (such as the achievement of Millennium Development Goals) are included, they are targeted at segments of the population that can reasonably be expected to be affected by the project, directly or indirectly.	
	1.5 IFI policy requires that project design include a results chain that represents the underlying logic to achieve the objective(s). It shows the links between project activities, outputs, and intended outcomes, and describes the assumptions and risks in causal relationships.	The final level of the results chain is called "impact" by some IFIs.
	1.6 IFI policy requires that project design include a reasonable number of key performance indicators that: are relevant to the project's objectives and span the range of the results chain from inputs to outcomes.	
	1.7 IFI policy requires that performance indicators be clearly defined and measurable, and that they use available data where possible.	
	1.8 IFI policy requires that any performance indicator target values be achievable .within the time-frame of the project.	
	1.9 IFI policy requires that project design include a plan for continuous self-evaluation of key activities throughout the life of the project.	
	1.10 IFI policy requires that data on the project's output and outcome indicators be collected, at a minimum, during project preparation/appraisal, and at project completion.	

Evaluation Principles (Standards and Elements)	Operational Practices	Notes
	1.11 IFI policy requires that output indicators be collected and monitored continuously throughout project implementation.	
B. Completion Reports (CRs): Operational departments execute CRs in accordance with the IFI's self-evaluation guidelines, and ensure CR quality and timely delivery.	2.1 Operational departments prepare a CR for every completed operation.	
	The CR summarizes the project's contribution to the intended outcomes contained in the project's statement of objectives. It assesses the following (2.2-2.6): 2.2 relevance of project objectives and design;	For IFIs that define "impact" as the final level in the results chain, the CR summarizes the project's contribution to the intended impacts contained in the project's statement of objectives.
	2.3 the degree to which the project achieved its objectives and delivered outputs as set out in the appraisal report;	
	2.4 the efficiency of the project in converting resources into results;	
	2.5 prospects for the project's sustainability; and	
	2.6 IFI and Borrower performance.	
	2.7 To the extent possible, the CR provides quantitative data to substantiate these assessments.	
	2.8 The CR also identifies key lessons learned related to the achievement of outcomes.	
	2.9 CRs are due within a fixed time (e.g., 6-12 months) after project closure. A longer time frame may be used in cases where outcomes are not observable within that time frame (e.g., some PBLs).	
C. CED involvement in Self-Evaluation: The CED is involved in the IFI self-evaluation system to support project evaluability and CR quality, but CED involvement is limited to activities that do not compromise the CED's independence.	3.1 The CED provides training to improve the evaluation capacity of Operations staff.	In addition, the CED may conduct evaluability assessments on projects at entry. This may include (i) preparing *ex post* evaluability assessments for a sample of projects that have been approved by the Board; and (iii) validating evaluability assessments conducted by Operations units.
D. Harmonization. The IFI's self-evaluation and independent evaluation systems are harmonized.	4.1 The CED coordinates with IFI units responsible for self-evaluation to ensure consistency in evaluation scope, criteria, and rating scales between self-evaluation and independent evaluation.	

Source: ECG Working Group on Public Sector Operations. GPS on Evaluation of Public Sector Operations. Revised Edition, 2012.

B. Private Sector Operations

Evaluation Principles (Standards and Elements)	Standard Operational Practices	Element Link	Notes
Private Sector Principles: Planning and Executing a Project Evaluation Program			
1. Scope of indirect evaluation A. *Self-Evaluation:* Indirect evaluations are undertaken by operational staff in line with GPS guidance. B. *IFI Reporting:* Findings from the indirect evaluations are reported in an Expanded Annual Supervision Report (XASR), which is signed off by operations' management. C. *XASR Research:* The XASR is based on internal IFI data, staff consultations, market research and stakeholder meetings. D. *Transparency:* The basis for findings are fully transparent in the XASR, including financial / economic calculations and environmental and social effects.	1.1 In an indirect evaluation, the project is evaluated by the IFI's operational staff. The scope of evaluation and indicator ratings should be consistent with the GPS. Staff report their findings in an Expanded Annual Supervision Report (XASR). The XASR is issued only after it has received the approval of the responsible operations department manager.	*Self-Evaluation* *IFI Reporting*	Note that the XASR is a once-only addendum to, or "expanded", Annual Supervision Report. The Annual Supervision Report is the regular supervision report prepared by the IFI's portfolio staff or equivalent.
	1.2 The research for XASRs draws from a file review, discussions with other operational staff involved with the operation since its inception, and external market research. The XASR should reflect consultations (in the field as necessary) with stakeholders who are knowledgeable about the country, company and project.*	*XASR Research*	* Such stakeholders could include: IFI specialists, the company's management, employees, auditors, suppliers, customers, competitors, bankers, any relevant government officials, industry associations, community representatives and local NGOs.
	1.3 The basis for findings and ratings are made fully transparent in the XASR. The XASR should also cite which stakeholder groups were consulted as part of the process. Where ex-post financial and/or economic rates of return for the project are cited, the document includes an attachment providing details supporting these calculations such as the key assumptions and underlying financial / economic time-series data. The XASR should include a summary of environmental, worker health and safety, and social performance information, for each of the IFI's environmental and social safeguards that apply to the project. Evidence from on-the-ground observations and/or client reporting should be included to support the assigned outcome and IFI work quality ratings. The information can be incorporated as an attachment to the XASR if preferred.	*Transparency*	

Source: ECG Working Group on Private Sector Evaluation. GPS on Evaluation of Private Sector Operations. 4th Edition, 2011.

C. Country Strategy and Program (CSP)

GPS Category (Standards and Elements)	Core GPS Description	Optional GPS Description	Notes
Process-related GPS			
1. Advance Preparations *A. Preparatory Step*	1.1 Evaluations of key projects, programs, and technical assistance operations should, if at all possible, be scheduled sufficiently in advance of the preparation of a CSPE. Operations personnel should also be encouraged to prepare self-evaluations in a timely manner		
2. Coverage *A. Self-evaluation Reports*	2.1 If self evaluation reports (i.e., country strategy completion reports) are properly done (and independently validated), this may reduce the need for in-depth independent CSPEs, particularly for small borrowers.		It can be difficult, however, for operations personnel to prepare candid and critical evaluations of country assistance performance, particularly in countries whose development results lag far behind what was expected.

Source: ECG. GPS on Country Strategy and Program Evaluations. 2008. Manila

BIBLIOGRAPHY

Development Committee Task Force on MDBs. 1996. *Serving a Changing World*. Washington, D.C.

Evaluation Cooperation Group. Fredrik Korfker. Annual Progress Report on Activities and Achievements for 2010. 19 September 2012. [https://wpqr1.adb.org/LotusQuickr/ecg/PageLibrary482576F500282648.nsf/269B1D2EDD6B3E6D4825798900283044/4224EB01BB5AB1124825798900283827/?OpenDocument]

_____________. Vinod Thomas. Annual Progress Report on Activities and Achievements for 2011. 19 September 2012. [https://wpqr1.adb.org/LotusQuickr/ecg/PageLibrary482576F500282648.nsf/269B1D2EDD6B3E6D4825798900283044/DA55DD7575B4295048257989002888 59/?OpenDocument]

_____________. *Peer Review of IFAD’s Office of Evaluation and Evaluation Function: Final Report*, February 2010. [https://wpqr1.adb.org/LotusQuickr/ecg/Main.nsf/7e6e83714d63fac348257731002a960f/3f296435ce2c24c648257731002a9639/?OpenDocument]

_____________. *A Proposal for an Evaluation Peer Review System for Multilateral Development Banks.* 2007 (A presentation in ECG Meeting in 2007). [https://wpqr1.adb.org/LotusQuickr/ecg/PageLibrary482576F500282648.nsf/h_Toc/73A018434637D8D0482577060058E5E2/?OpenDocument]

_____________. Good Practice Standards for the Evaluation of Private Sector Operations. Fourth Edition. 2011.

_____________. Good Practice Standards for the Evaluation of Public Sector Operations. Revised Edition. 2012.

_____________. Good Practice Standards: Country Strategy and Program Evaluation. 2008

_____________. *Report of Task Force on GPS on CSPE*. (A presentation in ECG 2007 Fall Meeting). [https://wpqr1.adb.org/LotusQuickr/ecg/PageLibrary482576F500282648.nsf/h_Toc/F0DA7EBD9503D60482577060058E4CC/?OpenDocument]

_____________. Good Practice Standards on Independence of International Finance Institutions' Central Evaluation Departments. June 2010.

_____________. Minutes of Meetings. Various years. Unpublished (ECG website).