

**Evaluation Cooperation Group
Working Group on Public Sector Evaluation**

**Good Practice Standards for the Evaluation of Public Sector
Operations**

2012 Revised Edition

February, 2012

Acronyms

| | |
|----------|---|
| APPI | Aggregate Project Performance Indicator |
| CED | Central Evaluation Department |
| CR | Completion Report |
| ECG | Evaluation Cooperation Group |
| WG PUB | Public Sector Working Group of the Evaluation Cooperation Group |
| EP | Evaluation Principle |
| ERR | Economic Rate of Return |
| FRR | Financial Rate of Return |
| GPS | Good Practice Standards |
| IFI | International Financial Institution |
| MDB | Multilateral Development Bank |
| OECD/DAC | Development Assistance Committee of the Organization for Economic Cooperation and Development |
| OP | Operational Practice |
| PBL | Policy-Based Lending or Policy-Based Loan |
| PER | Performance Evaluation Report |
| TA | Technical Assistance |

Table of Contents

| | |
|--|----|
| Background | 1 |
| Objectives and Organization..... | 2 |
| Benchmarking Against the GPS..... | 3 |
| Summary of Evaluation Principles | 6 |
| Good Practice Standards..... | 9 |
| Self-Evaluation | 9 |
| Report Preparation and Processes | 12 |
| Evaluation Approach and Methodology | 14 |
| Dissemination and Utilization | 24 |
| Annex: A Note on “Impact” and “Impact Evaluation” in the GPS | 26 |
| | |
| Guidance Note 1: Attributing Outcomes to the Project | 30 |
| Guidance Note 2: Efficiency Analysis | 35 |
| Guidance Note 3: Special Considerations for Policy-Based Lending | 38 |

Background

1. The Evaluation Cooperation Group (ECG) was established in 1996 among the major Multilateral Development Banks (MDBs) expressly “to develop... methodology and criteria for assessing and rating the MDB’s operational performance and development effectiveness.”¹ Subsequently, a number of additional International Financial Institutions (IFIs) have become members of the ECG.
2. In carrying out its mandate, the ECG has developed a series of Good Practice Standards (GPS) covering public and private sector operations, country programs, independence of evaluation, and other areas of development evaluation work. GPS on public sector operations first were adopted in 2002.² 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.³ 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.⁴
3. The process of revising the 2002 GPS began with a stocktaking of current IFI practices in evaluation of public sector operations, conducted in 2010.⁵ At its November 2010 meeting in London, the ECG Public Sector Evaluation Working Group (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 Good Practice Standards (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.
4. 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 private sector Working Group 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 Public Sector Working Group.
5. A draft revision of the ECG public sector evaluation Good Practice Standards 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

¹ Development Committee, Task Force on Multilateral Development Banks, “Serving a Changing World-- Report of the Task Force on Multilateral Development Banks,” March 15, 1996, p. 18.

² ECG, *Good Practice Standards for Evaluation of MDB Supported Public Sector Operations*, 2002. Standards covering policy-based lending later were added as an annex.

³ 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.

⁴ V. V. Desai, “Benchmarking of MDB Evaluation Systems Against the GPS for Public Sector Operations,” 2007.

⁵ Kris Hallberg, *Multilateral Development Bank Practices in Public Sector Evaluation*. Final Report, March 3, 2011.

GPS was updated in its third draft, dated May 2011.⁶ 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.⁷ Based on that exercise, members agreed on final revisions to these GPS.

Objectives and Organization

6. 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 GPS also attempt to define more effective linkages between independent and self-evaluation; 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.
7. 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.⁸
8. The GPS are organized into four sections, dealing with self-evaluation, report preparation and processes, evaluation approach and methodology, and dissemination and utilization. Standards for self-evaluation 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. 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 Central Evaluation Department (CED) reporting and disclosure standards.
9. 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"). 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.

⁶ Kris Hallberg, "Good Practice Standards for the Evaluation of Public Sector Operations: 2011 Revised Edition." Third Draft, May 22, 2011.

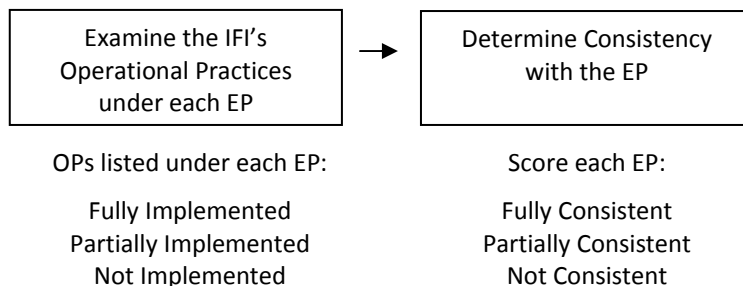
⁷ Patrick G. Grasso, "Benchmarking Pilot for the Draft Public Sector GPS," 2011. The pilot was based on a review of evaluation guidelines of the two IFIs, as opposed to reviewing performance.

⁸ ECG, *Review Framework for the Evaluation Function in Multilateral Development Banks*, 2009.

In addition to the EPs and OPs, the GPS contains 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).

Benchmarking Against the GPS

10. 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 (#15) 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.
11. 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.



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. |
| Borrower Performance | The adequacy of the Borrower's assumption of ownership and responsibility during all project phases, including government and implementing agency 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. |
| Completion Report | A record of an 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 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 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 impact, EU policy, European social cohesion, etc. |
| Cost-benefit analysis | 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 | 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. |
| Economic Rate of Return | 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 <i>per se</i> . |
| 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. |
| Financial Rate of Return | The internal rate of return of a time series of cash flows describing the project's financial investments and returns. |
| 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. |

| Term | Definition |
|-------------------------------------|---|
| Impact | Given the multiple meanings associated with this term, the 2012 GPS does not use “impact” as the name of an evaluation criterion nor of a level in the results chain. (See the annexed note on 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.) |
| International Financial Institution | A financial institution created by a group of countries that provides financing and advisory services for projects and programs in member countries. The term includes the World Bank, regional development banks, and other regional financial institutions. It does not include internationally-operating commercial banks. |
| Net present value | The sum of the present values of the time series of project costs and benefits. |
| Outcome | The final level in the results chain, reflecting the objectives of the project. If necessary, a distinction can be made between “intermediate outcomes” (the uptake of project outputs by beneficiaries, behavioral changes on the part of beneficiaries, results achieved once beneficiaries use project outputs, and/or results achieved in the short-to-medium term) and “final outcomes” (the ultimate outcomes or project goals arising from intermediate outcomes). “Final outcomes” are called “impacts” by some IFIs. |
| Output | The tangible goods and services that the project activities produce, generally under the direct control of the implementing agency. |
| Performance Evaluation Report | An independent project evaluation conducted by a Central Evaluation Department, which normally includes field work. |
| Project | 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”. |
| Relevance | Consistency of the project’s 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)”. |
| 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. |

Summary of Evaluation Principles

| No. | Evaluation Principle | Number of OPs | Page |
|--|--|---------------|------|
| Self-Evaluation Standards | | | |
| 1 | IFI policy requires that project design include a minimum set of elements to ensure evaluability. | 10 | 8 |
| 2 | Operational departments execute CRs in accordance with the IFI's self-evaluation guidelines, and ensure CR quality and timely delivery. | 9 | 92 |
| 3 | 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. | 1 | 9 |
| 4 | The IFI's self-evaluation and independent evaluation systems are harmonized. | 1 | 10 |
| Report Preparation and Processes | | | |
| A. Timing | | | |
| 5 | 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. | 2 | 11 |
| B. Coverage and Selection | | | |
| 6 | The CED has a strategy for its mix of evaluation products that balances the two evaluation functions of accountability and learning. | 1 | 11 |
| 7 | 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. | 2 | 11 |
| 8 | If an additional purposive sample of projects is selected for learning purposes, it is not used by itself for corporate reporting. | 1 | 11 |
| 9 | The sampling methodology and significance of trends are reported. | 3 | 12 |
| C. Consultation and Review | | | |
| 10 | Stakeholders are consulted in the preparation of evaluations. | 3 | 12 |
| 11 | Draft evaluations are reviewed to ensure quality and usefulness. | 3 | 12 |
| Evaluation Approach and Methodology | | | |
| A. Basis of the Evaluation | | | |
| 12 | Evaluations are primarily objectives-based. | 8 | 13 |
| 13 | If project objectives were revised during implementation, the project is assessed against both the original and the revised objectives. | 1 | 14 |
| 14 | The evaluation includes consideration of unanticipated outcomes. | 3 | 15 |
| 15 | Evaluations of PBLs assess the performance of the reform program as a whole. | 2 | 15 |

| No. | Evaluation Principle | Number of OPs | Page |
|--------------------------------------|---|---------------|------|
| B. Criteria | | | |
| 16 | Evaluations encompass all performance attributes and dimensions that bear on the operation's success. | 2 | 15 |
| 17 | The assessment of Relevance covers both the relevance of objectives and the relevance of project design to achieve those objectives. | 7 | 16 |
| 18 | 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. | 3 | 17 |
| 19 | Subject to information and CED resource constraints, the assessment of Effectiveness uses appropriate methods to determine the contribution of the project to observed outcomes in a causal manner. | 4 | 17 |
| 20 | The Efficiency assessment attempts to address two main 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? | 6 | 17 |
| 21 | 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. | 4 | 19 |
| 22 | The assessment of IFI Performance covers the quality of services provided by the IFI during all project phases. | 10 | 19 |
| 23 | Borrower Performance assesses the adequacy of the Borrower's assumption of ownership and responsibility during all project phases. | 6 | 20 |
| C. Ratings | | | |
| 24 | Each of the six criteria (five for PBLs) is assigned a rating. | 2 | 21 |
| 25 | Rules for assigning criteria ratings are reported. | 2 | 21 |
| 26 | An Aggregate Project Performance Indicator (APPI) is constructed from the core criteria. | 6 | 21 |
| Dissemination and Utilization | | | |
| 27 | The CED prepares a periodic synthesis report. | 5 | 23 |
| 28 | The CED makes evaluation findings and lessons easily available to IFI staff. | 3 | 23 |
| 29 | Within the guidelines of the IFI's overall disclosure policy, the CED discloses all evaluation products. | 2 | 23 |
| 30 | The CED pro-actively reaches the public with its evaluation results. | 1 | 24 |
| 31 | The CED follows up on IFI Management's implementation of recommendations made by the CED. | 3 | 24 |
| Total No. OPs | | 124 | |

Good Practice Standards

Self-Evaluation

| Evaluation Principle | Operational Practices | Notes |
|---|--|---|
| <p>1. IFI policy requires that project design include a minimum set of elements to ensure evaluability.</p> | <p>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.</p> | <p>The outcomes in the objective statement are called “impacts” by some IFIs.</p> |
| | <p>1.2 The statement of objectives is based on a problem or diagnostic statement.</p> | |
| | <p>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.</p> | |
| | <p>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.</p> | |
| | <p>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.</p> | <p>The final level of the results chain is called “impact” by some IFIs.</p> |
| | <p>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.</p> | |
| | <p>1.7 IFI policy requires that performance indicators be clearly defined and measurable, and that they use available data where possible.</p> | |
| | <p>1.8 IFI policy requires that any performance indicator target values be achievable .within the time-frame of the project.</p> | |

| Evaluation Principle | Operational Practices | Notes |
|--|---|---|
| | 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. | |
| | 1.11 IFI policy requires that output indicators be collected and monitored continuously throughout project implementation. | |
| 2. 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): | 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. |
| | 2.2 relevance of project objectives and design; | |
| | 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). | |
| 3. The CED is involved in the IFI self-evaluation system to support project evaluability and CR quality, but CED involvement | 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 <i>ex post</i> evaluability assessments for a sample |

| Evaluation Principle | Operational Practices | Notes |
|---|---|--|
| is limited to activities that do not compromise the CED’s independence. | | of projects that have been approved by the Board; and (iii) validating evaluability assessments conducted by Operations units. |
| 4. 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. | |

Report Preparation and Processes

| Evaluation Principle | Operational Practices | Notes |
|---|---|--|
| A. Timing | | |
| 5. 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. | 5.1 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. | |
| | 5.2 PBLs in a series are evaluated at the end of the series. | Relevant for IFIs that provide PBLs. |
| B. Coverage and Selection | | |
| 6. The CED has a strategy for its mix of evaluation products that balances the two evaluation functions of accountability and learning. | 6.1 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). |
| 7. 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. | 7.1 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. | |
| | 7.2 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. |
| 8. If an additional purposive sample of projects is selected for learning purposes, it is not used by itself for corporate | 8.1 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; |

| Evaluation Principle | Operational Practices | Notes |
|--|---|--|
| reporting. | | potential for planned or ongoing country, sector, thematic, or corporate evaluations; to verify CR validation ratings; and areas of special interest to the Board. |
| 9. The sampling methodology and significance of trends are reported. | 9.1 The CR validation sample and the PER sample are set in the CED's annual work program. Ratios and selection criteria are clearly stated. | |
| | 9.2 In corporate reporting the confidence intervals and sampling errors are reported. | |
| | 9.3 The significance of changes in aggregate project performance and how to interpret trends are reported. | |
| C. Consultation and Review | | |
| 10. Stakeholders are consulted in the preparation of evaluations. | 10.1 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. | |
| | 10.2 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. | |
| | 10.3 The CED invites comments from the Borrower on draft PERs. Their comments are taken into account when finalizing the report. | |
| 11. Draft evaluations are reviewed to ensure quality and usefulness. | 11.1 To improve the quality of PERs, draft PERs are peer reviewed using reviewers inside and/or outside the CED. | |
| | 11.2 To ensure factual accuracy and the application of lessons learned, draft PERs are submitted for IFI Management comments. | |
| | 11.3 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 | Operational Practices | Notes |
|---|---|--|
| A. Basis of the Evaluation | | |
| 12. Evaluations are primarily objectives-based. | 12.1 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 #26 and OP #26.1 and #26.2. |
| | 12.2 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. | |
| | 12.3 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). | |
| | 12.4 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 | |
| | 12.5 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 |

| Evaluation Principle | Operational Practices | Notes |
|---|--|--|
| | | need to be constructed from sources of information other than the project documents, including interview evidence from government officials and IFI staff. |
| | 12.6 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. | |
| | 12.7 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. |
| | 12.8 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. |
| 13. If project objectives were revised during implementation, the project is assessed against both the original and the revised objectives. | 13.1 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. | <p>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.</p> <p>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.</p> |

| Evaluation Principle | Operational Practices | Notes |
|---|---|--|
| <p>14. The evaluation includes consideration of unanticipated outcomes.</p> | <p>14.1 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.</p> | <p>Unanticipated outcomes are called “unanticipated impacts” by some IFIs.</p> <p>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.</p> <p>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 objectives.</p> |
| | <p>14.2 Unanticipated outcomes are taken into account in the assessment of Efficiency. The calculation of the project’s <i>ex post</i> 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).</p> | |
| | <p>14.3 Unanticipated outcomes, both positive and negative, are discussed and documented in a separate section of the evaluation.</p> | |
| <p>15. Evaluations of PBLs assess the performance of the reform program as a whole.</p> | <p>15.1 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.</p> | <p>Relevant for IFIs that provide PBLs.</p> |
| | <p>15.2 PBL evaluations assess the results of the overall program, regardless of the sources of financing.</p> | <p>Relevant for IFIs that provide PBLs.</p> |
| <p>B. Criteria</p> | | |
| <p>16. Evaluations encompass all performance attributes and dimensions that bear on the</p> | <p>16.1 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)</p> | <p>Definitions of the criteria are given in the Terms and Definitions section of this document.</p> |

| Evaluation Principle | Operational Practices | Notes |
|---|---|---|
| <p>operation's success.</p> | <p>along with IFI Performance and Borrower Performance. This applies to both CR validations and PERs.</p> | <p>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.</p> |
| | <p>16.2 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.</p> | <p>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.</p> |
| <p>17. The assessment of Relevance covers both the relevance of objectives and the relevance of project design to achieve those objectives.</p> | <p>17.1 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.</p> | <p>For further guidance on assessing Relevance for PBLs, see Guidance Note 3.</p> |
| | <p>17.2 The assessment also considers the extent to which the project's objectives are clearly stated and focused on outcomes as opposed to outputs.</p> | |
| | <p>17.3 The realism of intended outcomes in the country's current circumstances also is assessed.</p> | |
| | <p>17.4 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.</p> | |
| | <p>17.5 The assessment also considers the relevance of modifications to project design.</p> | |
| | <p>17.6 Whether the project's financial instrument was appropriate to meet project objectives and country needs also is assessed.</p> | |

| Evaluation Principle | Operational Practices | Notes |
|---|--|---|
| | 17.7 The relevance of objectives and design is assessed against circumstances prevailing at the time of the evaluation. | |
| 18. 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. | 18.1 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. |
| | 18.2 Both the actual and the expected results of an operation are included in the assessment of Effectiveness. | |
| | 18.3 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. |
| 19. 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. | 19.1 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 for a menu of quantitative and qualitative approaches to attributing outcomes to the project. |
| | 19.2 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 affected outcomes. | |
| | 19.3 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. | |
| | 19.4 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. |
| 20. The Efficiency assessment | 20.1 To address the first question (Did the benefits of the project, achieved or | See Guidance Note 2 for further detail on |

| Evaluation Principle | Operational Practices | Notes |
|--|--|---|
| <p>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?</p> | <p>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.</p> | <p>options for assessing Efficiency. Note that Efficiency is assessed for investment and TA loans but not for PBLs (see OPs #16.1 and #16.2).</p> |
| | <p>20.2 The methodology and assumptions underlying the calculation of an economic rate of return or net present value are clearly explained and transparent. <i>Ex post</i> estimates are compared with the <i>ex ante</i> estimates in the project documents.</p> | <p>Relevant when ERRs/NPVs are estimated.</p> |
| | <p>20.3 Sensitivity tests on ERRs based on possible changes in key assumptions are carried out. These assumptions reflect any concerns in the assessment of Sustainability.</p> | <p>Relevant when ERRs/NPVs are estimated.</p> |
| | <p>20.4 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.</p> | |
| | <p>20.5 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 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.</p> | |
| | <p>20.6 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.</p> | <p>Relevant for IFIs that provide lending for TA.</p> |

| Evaluation Principle | Operational Practices | Notes |
|--|--|---|
| <p>21. 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.</p> | <p>21.1 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.</p> | |
| | <p>21.2 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.</p> | |
| | <p>21.3 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 <i>per se</i>.</p> | <p>Intended outcomes are called “impacts” by some IFIs.</p> |
| | <p>21.4 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.</p> | <p>For PBLs, see also Guidance Note 3.</p> |
| <p>22. The assessment of IFI Performance covers the quality of services provided by the IFI during all project phases.</p> | <p>22.1 The assessment of IFI Performance at project entry covers the IFI’s role in ensuring project quality and in ensuring that effective arrangements were made for satisfactory implementation and future operation of the project. This includes:</p> | |

| Evaluation Principle | Operational Practices | Notes |
|--|--|-------|
| | the quality of the analysis conducted to identify problems and possible solutions; | |
| | 22.2 the consideration of alternative responses to identified problems; | |
| | 22.3 the degree of participation of key stakeholders; the use of lessons learned from previous operations; | |
| | 22.4 the quality of risk analysis and the adequacy of proposed risk mitigation measures; | |
| | 22.5 the adequacy of institutional arrangements for project implementation; | |
| | 22.6 the identification of safeguards relevant to the project; and | |
| | 22.7 the IFI's efforts to ensure the quality of the monitoring and evaluation framework. | |
| | 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: | |
| | 22.8 modifying project objectives and/or design as necessary to respond to changing circumstances; | |
| | 22.9 enforcing safeguard and fiduciary requirements; and | |
| 22.10 ensuring that the monitoring and evaluation system was implemented. | | |
| 23. Borrower Performance assesses the adequacy of the Borrower's assumption of ownership and responsibility during all project phases. | 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: | |
| | 23.1 government and implementing agency performance in ensuring quality preparation and implementation; | |
| | 23.2 compliance with covenants, agreements, and safeguards; | |
| | 23.3 provision of timely counterpart funding; | |

| Evaluation Principle | Operational Practices | Notes |
|--|---|--|
| | 23.4 provision of timely counterpart funding; and | |
| | 23.5 measures taken by the Borrower to establish the basis for project sustainability, particularly by fostering participation by the project's stakeholders. | |
| | 23.6 The assessment covers the performance of the government as well as the performance of implementing agencies. | |
| C. Ratings | | |
| 24. Each of the six criteria (five for PBLs) is assigned a rating. | 24.1 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. |
| | 24.2 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. |
| 25. Rules for assigning criteria ratings are clearly spelled out. | 25.1 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. |
| | 25.2 In evaluation reports, evaluators provide a justification for each rating. | |
| 26. An Aggregate Project Performance Indicator (APPI) is | 26.1 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 |

| Evaluation Principle | Operational Practices | Notes |
|--|--|--|
| <p>constructed from the core criteria.</p> | | <p>#16.1 above), their ratings are not used in the calculation of the APPI (i.e., they are “below the line”).</p> <p>A second aggregate indicator, including these additional criteria, may be constructed.</p> |
| | <p>26.2 For PBLs, the APPI is constructed from the three core criteria: Relevance, Effectiveness, and Sustainability.</p> | <p>Relevant for IFIs that provide PBLs.</p> |
| | <p>26.3 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.</p> | <p>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 project completion would be unusual.</p> |
| | <p>26.4 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.</p> | |
| | <p>26.5 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.</p> | |
| | <p>26.6 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.</p> | <p>Relevant for IFIs that construct a second aggregate indicator.</p> |

Dissemination and Utilization

| Evaluation Principle | Operational Practices | Notes |
|--|---|-------|
| 27. The CED prepares a periodic synthesis report. | 27.1 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. | |
| | 27.2 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. | |
| | 27.3 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. | |
| | 27.4 The CED's synthesis ratings are included in integrated corporate performance reporting. | |
| | 27.5 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. | |
| 28. The CED makes evaluation findings and lessons easily available to IFI staff. | 28.1 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. | |
| | 28.2 The CED relies primarily on its intranet website for document posting and notifies staff of new items through the corporate website. | |
| | 28.3 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. | |

| | | |
|--|--|---|
| 29. Within the guidelines of the IFI's overall disclosure policy, the CED discloses all evaluation products. | 29.1 The CED's disclosure policy for evaluation products is explicit and consistent with the IFI's general disclosure policy. | |
| | 29.2 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. |
| 30. The CED pro-actively reaches the public with its evaluation results. | 30.1 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. |
| 31. The CED follows up on IFI Management's implementation of recommendations made by the CED. | 31.1 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. | |
| | 31.2 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. | |
| | 31.3 The CED reports to the Board on IFI Management follow-up to its recommendations. | |

Annex: 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.”⁹

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.

⁹ Organization for Economic Cooperation 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*.

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 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 (OP #1.1).
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.”¹⁰

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.¹¹

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

¹⁰ International Initiative for Impact Evaluation. *3ie Principles for Impact Evaluation*. See www.3ieimpact.org.

¹¹ See Leeuw and Vaessen (*op. cit.*).

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.¹²

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 #19 and Guidance Note 1). 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.

¹² See Guidance Note 1 on attributing outcomes to a project.

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.”¹³ 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.¹⁴ 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.¹⁵ 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

¹³ 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”).

¹⁴ See White, Howard (2007). *Evaluating Aid Impact*. World Institute for Development Economics Research, Research Paper No. 2007/75, November.

¹⁵ Example from White, Howard (2009). *Some Reflections on Current Debates in Impact Evaluation*. International Initiative for Impact Evaluation Working Paper No. 1, April.

feasible or practical. Preference is to use mixed-methods designs. Use existing research relevant to the results of the intervention.¹⁶

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:¹⁷
- **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.

¹⁶ Leeuw, Frans and Jos Vaessen (2009). *Impact Evaluations and Development: NONIE Guidance on Impact Evaluation*. Network of Networks on Impact Evaluation.

¹⁷ 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.

- **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 (such as institutional reforms). Random assignment also may not be possible for political or ethical reasons.¹⁸
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.

¹⁸ See Soares (op. cit.) for the limitations of randomized control trials and of impact evaluation in general.

- Observes the project in operation, focusing on interactions that were expected to produce the intended outcomes.¹⁹
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.

Policy-Based Lending²⁰

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.

¹⁹ Morra-Imas, Linda, and Ray C. Rist (2009). *The Road to Results: Designing and Conducting Effective Development Evaluations*. World Bank.

²⁰ 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.

- 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 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.

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.²¹ 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.

²¹ 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.

- 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 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.

Sources:

African Development Bank. 2001. *Revised Guidelines on Project Completion Report (PCR) Evaluation Note and Project Performance Evaluation Report (PPER)*. OPEV, January (p. 29).

Asian Development Bank. 2006. *Guidelines for Preparing Performance Evaluation Reports for Public Sector Evaluations*. IED, January (pp. 12-14).

Independent Evaluation Group. 2010. *Cost-Benefit Analysis in World Bank Projects*. Washington, DC: World Bank.

International Fund for Agricultural Development. 2009. *Evaluation Manual: Methodology and Processes*. IOE, April (p. 34).

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,²² the background paper for the PBL GPS prepared by the Asian Development Bank,²³ 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).²⁴
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.²⁵

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

²² ECG (2004), *Good Practices for the Evaluation of Policy-Based Lending*. Addendum to the 2002 GPS.

²³ 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.

²⁴ 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.

²⁵ ECG 2004, p. 2.

completion. Where there is a series of overlapping and related PBLs, the timing question is more 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:
 - Macroeconomic stability: 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.
 - Political economy: 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.

- Complementary reforms: 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.