



RE-384

***OVE's Proposal for Integrating
Evaluability Review into the
Bank's Quality Control
Function***

***Office of Evaluation and Oversight,
OVE***

Inter-American Development Bank
Washington, D.C.
January 2011

For Official Use Only

TABLE OF CONTENTS

A. Objectives 1
B. Background..... 1
C. Proposal 1

ANNEX A. EVALUABILITY GUIDELINES

ANNEX B. EVALUABILITY SCORING SHEET

A. Objectives

- 1.1 The objective of this document is to comply with BOD instructions to propose a revised set of reliable evaluability standards to be applied to the Bank's projects prior to their approval. This document consists of a main proposal, an annex with evaluability guidelines (Annex A), and an annex with an evaluability scoring sheet (Annex B).

B. Background

- 1.2 The IDB is the first Development Bank to have conducted systematic Evaluability reviews of approved projects. All projects approved in 2001, 2005 and 2009, have been reviewed for evaluability by the Bank's Office of Evaluation and Oversight (OVE). In 2009, the evaluability review, as part of the assessment of project quality, was incorporated within Management's proposed Development Effectiveness Matrix (DEM). The DEM was applied to all projects approved in 2009 and 2010. Likewise, in 2010, the Board of Governors (BOG) adopted the 9th General Increase in Capital (IDB9). In the Cancun Declaration the Governors endorsed the requirement that Bank-approved projects contain a minimum "development effectiveness threshold". The GCI document proposed to implement this "development threshold" in part by using an "evaluability threshold". The BOG also indicated that the DEM be reviewed and used as the instrument by which this "threshold" should be measured. In November of 2010, Management presented to the Board a set of proposed revisions to the DEM in order to comply with the IDB9 requirements. Following this meeting, the IDB Board of Directors identified areas of improvement in the DEM regarding the implementation of evaluability reviews. These concerns ranged from issues of content to those of process and method. The BOG instructed Management to consult with OVE in order to address these pending issues. Following the BOD instructions, OVE met with SPD and provided preliminary comments on how the treatment of evaluability should be modified, particularly with respect of SPD's role. These comments provide a comprehensive proposal on how evaluability can be adjusted, in terms of functions, process, and content.
- 1.3 This proposal is based on the premise that the revision of the means by which the Bank ensures minimum evaluability entails changes not only to the evaluability guidelines but also to the Bank's norms which govern the preparation and review of projects. It consists of six sections, each detailing the measures required to integrate evaluability review into the Bank's quality control function.

C. Proposal

- 1.4 **The Bank will adopt and implement Evaluability Guidelines.** Project teams must have useful and meaningful guidelines of what constitutes an evaluable project. These guidelines, presented as annex A of this document, are based on OVE's experience in the qualitative assessment of evaluability and have been revised to further incorporate areas of emphasis of the IDB Board of Directors

and valuable lessons learned from the Bank's recent experience with the DEM. These guidelines identify the main components of evaluability, as well as the questions that a loan proposal should answer, as well as the general content requirements in order to answer these questions. The guidelines are qualitative in nature, and focus on concepts and questions, rather than on specific content requirements. This affords project teams the flexibility required to prepare complex operations, and to avoid the common problem of "teaching to the test". These guidelines also serve as the terms of reference for the Evaluability Assessment Note.

- 1.5 **The Bank will introduce an Evaluability Assessment Note.** The assessment of project evaluability requires an analytical review of the project, and a thorough discussion of the problems found in the loan proposal. A peer-reviewed assessment is the only way to assess project evaluability in a credible and meaningful manner. The Bank will implement a narrative-based peer-review of project evaluability for each loan proposal prepared, in the form of an Evaluability Assessment Note (EAN). The peer review process requires a two-stage process based on the constitution of a review team for each project, and the revision by an Evaluability Panel, under the responsibility of SPD, which will sign off on the quality of the review process and the final EAN. The review will be done according to the attached evaluability guidelines, and will report on the degree to which projects comply with each of the standards. The review will also identify the specific areas in which projects have evaluability problems. The EAN will be necessarily integrated into the Bank's quality control process. The EAN will play an essential role in this quality control process by providing qualitative analytical information before the QRR, and the corresponding managerial documentation. The Bank unit responsible for the review will be SPD, which will be in charge of the preparation of the EAN prior to the project QRR. The review will report on SPD's main findings and provide a basis for discussion regarding project evaluability at the QRR. In order to perform this quality control function with an adequate degree of separation from the object of evaluation and independence, the responsible unit will have functional separation from the project preparation process and cannot participate in project teams. Although this review will also form the basis of the project's evaluability scoring by SPD later in the process, the review will be undertaken separately (and before) scoring. The EAN will be adjusted to the extent that the project team makes the required adjustments to its project. This last version of the EAN will be included with the POD package and presented to the OPC and to the Board.
- 1.6 **The Bank will adopt a clear, reliable and meaningful evaluability rating.** The Bank will adopt a clear, reliable and meaningful evaluability rating based on the attached guidelines. These guidelines were revised according to practice and the review of the DEM. These guidelines may be complemented with a set of binary ("checklist") questions from other parts of the DEM, as appropriate. The scoring guidelines identify a rating mechanism in which the numerical rating produced has an interpretation with respect to the adequacy of the project evaluability. A minimum acceptable score should be established, and this score should be part of

the minimum DEM score of “5” adopted by the BOG. The rating will be informed by the analysis of the Evaluability Assessment Note and the degree to which issues raised in the EAN were addressed, as reflected in the final note. This scoring must be calculated and reported separately, and can then be subsumed into the broader DEM. It is important that the focus of the evaluability review be on improving Bank operations, rather than on achieving a specific number on an evaluability rating. To this end, projects will be rated for evaluability only after the QRR meeting, so that the discussion and remedies applied revolve around the relevant issues rather than around the scoring exercise.

- 1.7 **The Bank will clarify the assignment of responsibility in the assessment of Evaluability.** The review of the Evaluability mandated by the Governors implies modifications in the processes and responsibilities laid out in the New Project Cycle (NPC). To put this new system in place and to ensure projects are evaluable, the Bank will rely on different units within the institution. VPS and VPP are responsible for the quality of projects, and will be responsible for applying evaluability guidelines and producing evaluable projects. To do so, VPS and VPP should internalize, disseminate, provide training, and ensure application of the guidelines by Sector Managers, and Project Teams. SPD will be responsible for the quality of the review of project evaluability, including the quality of an EAN and the quality of the Evaluability ratings (see below). To perform this function, SPD will be responsible for constituting and managing a peer-review mechanism by adopting and applying the review standards contained in this proposal. The EVP will be responsible for ensuring that projects sent forth to the Board have received at least a minimum acceptable evaluability score (or have explicitly requested a waiver). In order to do so, the EVP will ultimately be responsible for the quality of the projects and Evaluability reviews that are submitted to the BOD, and for sending back projects that do not meet minimum evaluability standards. The President of the Bank will be responsible for the request of waivers under special circumstances, as established in the GCI document. Lastly, OVE will be responsible for reporting to the Board annually on the Evaluability of Bank Projects as well as on the performance of Bank Management in the production of evaluable projects.
- 1.8 Beyond these general responsibilities, specific changes to the project cycle of every project under the new system include:
 - a. In order for project evaluability to improve, projects must be reviewed for evaluability at a formal review instance, the QRR. Management will revise guidelines for the QRRs, so that (i) the practice of virtual meetings is discontinued—meetings should be held live; (ii) Bank units involved in quality control must participate, and in the case of SPD, the EAN must be presented. EANs must be presented at the CRM for VPP.
 - b. After all changes are made and the final POD is scored, SPD will make a recommendation to the EVP as to the project meeting minimum evaluability standards. However, the EVP is ultimately responsible for

accepting or declining this recommendation, and for moving the project forward, sending it back, or sending it to the President for a waiver.

- c. A table summarizing activities and accountabilities for each project is presented below (Table 1.1):

Table 1.1

Step	Activity	Responsibility/Unit
1	A POD is prepared, according to published evaluability guidelines.	Project team/VPS or Project team/VPP
2	POD is assessed by preparing an <i>Evaluability Assessment Note and peer reviewed</i> , according to published evaluability guidelines	SPD Manager
3	<i>Evaluability Assessment Note</i> findings are discussed at a “live” QRR review of the POD. SPD mandatorily participates in the QRR (CRM). The QRR (CRM) minutes document VPS (VPP) instructions to improve evaluability.	VPS or VPP SPD Manager
4	<i>Evaluability Assessment Note</i> findings regarding deficiencies are addressed in the POD	Project team/VPS or Project team/VPP
5	POD is deemed ready for Evaluability rating and is forwarded for rating before submission to the OPC	VPS Manager or VPP Manager
6	POD is rated. Rating reflects how <i>Evaluability Assessment Note</i> findings have been addressed	SPD
7	Determination is made if minimum standards have been met, and recommendation is made to EVP	SPD Manager
8	OPC considers SPD’s recommendation, makes a determination as to the adequacy of the project’s evaluability, and the President of the Bank then requests a waiver, as appropriate.	EVP/PRE
9	Independent report on the evaluability of Bank projects is prepared and presented to the BOD annually.	OVE

- 1.9 **The Bank will revise the normative establishing the content of the POD.** The POD is the institution’s executive loan document; it is the main source of information available to top Management and the Board regarding loan approval decisions. As such, it must be sufficiently self-contained regarding evaluability content. The NPC normative will be revised in order to clarify content and organizational issues of the POD. In particular, these revisions will: (i) require the POD to be self-contained regarding evaluability—it should either contain or reflect (through proper referencing and documentation) the required information in order to establish project evaluability; (ii) the purpose, content requirement, and structure of all mandatory annexes will be specifically identified, and in particular specific requirements will be identified for the mandatory monitoring and evaluation annex; (iii) relevant information must be available at the time of approval, and not deferred—in circumstances in which information is deferred (“evaluability deferred”) a detailed plan for data collection and corrective

measures shall be included as an annex. The NPC will also be revised to reflect the attached evaluability guidelines and requirements.

- 1.10 **The Bank will implement Country-led evaluability in project preparation on an experimental basis.** The advent of increasingly sophisticated and competent country bureaucracies—particularly in middle- and upper-middle income LAC borrowing countries—can be a valuable asset for the Bank in the preparation of evaluable projects. It can also be a cost controller to the extent that it can promote increased efficiency within the institution. Management should experiment with an approval process in which project preparation can be undertaken directly by borrowing countries. In these instances the Bank would retain an oversight and quality control function. However, projects will be prepared and presented directly by the borrower for consideration by the Bank, subject to the same requirements (and evaluability standards) as those projects prepared and submitted by VPS and VPP. Management should then assess the results of this experimentation with a view to improve the evaluability of Bank-financed projects—both those designed by the institution and those designed directly by borrowing countries.

EVALUABILITY GUIDELINES

These guidelines are based on OVE's experience in the assessment of evaluability and have been revised to further incorporate areas of interest of the IDB Board of Directors and valuable lessons learned from the Bank's recent experience with the DEM.

The quality of the application of these guidelines in the review by SPD depends substantially on the thoroughness of the work made by the reviewing team, and in the rigor of the quality review carried out by the panel. These guidelines should also provide guidance to project teams under VPS and VPP responsibility since the beginning of the design process.

Given that Bank's projects are complex in nature and should necessarily differ substantially from one to another setting, the guidelines only specify fundamental content requirements for the projects, as a point of departure for a peer-review based qualitative assessment, which requires professional judgment and experience, as it cannot be applied mechanically.

A. Diagnosis

The diagnosis should include these basic characteristics:

- Development Problem. The problem or need that the project attempts to address is clearly identified and characterized.
- Underlying Causes. The underlying causes of the problem are identified: both proximate and root causes.
- Beneficiaries. The potential beneficiaries are clearly identified. This includes both direct and indirect beneficiaries.
- Evidence-based. Diagnostics are data-driven, evidence-based, and analytical (as opposed to merely descriptive).

The diagnosis of the problem or problems should be well defined, meaning that the diagnosis should incorporate accurate information to respond to the question of *why* a certain situation is perceived as problematic. An adequate response to this question implies that the diagnosis should incorporate the following:

- Clear, accurate, relevant, and current information and data regarding the problem and its sector should be provided. In instances where the Bank has had experience in the sector, the diagnostic will draw on data, evidence, and knowledge produced by the Bank;
- An analysis of the reasons why the situation constitutes a problem requiring an intervention, including descriptions of relevant markets and their operation, supply and demand characteristics, or other facets of the model being used to interpret the problem situation should be provided. In most cases this should

include an analysis of the policy environment and its relevance for the problems being addressed;

- The relationships and interactions that occur among the factors and elements that constitute the problem are given and analyzed;
- The historical, political, social, and economic context in which the problem arose, as well as those in which its solution will be attempted is provided;
- The conditions or implicit assumptions that need to occur for the problem to be solved and the risks implied by the solution itself.

The identification of the underlying relationships mentioned above and their analysis is crucial for a proper diagnostic and also impacts on the project's logic. This analysis should include both the proximate or immediate causes of observed phenomenon as well as analysis of root causes, which should ultimately be the object of an intervention. A diagnosis that does not identify the factors that make up the essential core of the problem, that does not define the relative weight of the elements of the problem in a precise manner, or that incorrectly identifies the interactions and relationships among them, does not provide an adequate framework for Bank action. The analysis of these relationships is what differentiates a "description of problems" from a diagnosis; it is this analysis which provides the distinction between a "problem statement" and a diagnostic in "solvable form". This should necessarily include an analysis of the policy environment. Operations are almost always implemented in the context of existing public policies.

These factors and elements do not interact in a vacuum, but rather in a certain historical, social, political, and economic context which in many cases will provide the origins of the problem situation. Therefore, an essential part of the diagnosis is the clear and accurate description of the relevant factors that, without being strictly a part of the problem, have a bearing on its occurrence and solution. If systematic impact evaluation reviews—or other evidence of effectiveness—of similar projects exist, their results should be analyzed. The review of previous assessments and studies should address the external validity of the interventions reviewed and their potential applicability for the country in which the project is being implemented. Furthermore, given the Bank's breadth of scope and history of project design and implementation in LAC, the Bank's own experience should be used, when available, in order to complement the empirical basis in the diagnostic on the effectiveness of public policy.

All diagnoses incorporate, implicitly or explicitly, the key elements related to the desired future situation (the current reality is analyzed from this), what level of deficiency is tolerable in the future, or what is inadmissible because it implies the permanence of the problem being diagnosed. The answers to these questions are the basis for determining the parameters for assessing the success or failure of the project, whether in terms of benchmarks, milestones or targets.

Finally, the identification of the beneficiaries is one of the key elements in determining the quality of the diagnosis. The appropriate identification of the beneficiaries is fundamental in the analysis of changes required in the status quo, present or future risks, and project assumptions and risks. The “beneficiaries” should constitute a basic point of reference for the diagnosis. Beneficiaries may include both direct beneficiaries of the intervention, as well as indirect beneficiaries. The identification of beneficiaries also serves to provide the implicit “weights” to the problem situation.

The diagnosis is an intellectual process in which a problematic situation is translated into a specific problem. This means that the diagnosis is a process whereby the general perception of something that does not work, or does not work well, is converted into a detailed analysis of the what and the why of the situation; of the elements, factors, and components the problem breaks down into technically; of how these interact; of what their relative weight is in the general situation; of where their roots are found and the causes of their presence; and of what are, definitively, the consequences and effects of this set of factors.

B. Definition of the Objectives

Project objectives define the project’s reason for being. They provide the link between the Bank and the development, and as such constitute the institution’s statement(s) of intent. These statements are what define in specific terms what one hopes will be accomplished. Models of change typically identify the relationship between changes implemented and results achieved as such:

Inputs→Activities→Outputs→Outcomes (and Impacts)

Here outcomes refer to the changes to the underlying constraints; although a limited interpretation permits that immediate or proximate constraints also be considered outcomes (or intermediate outcomes). In terms of the normative of the Bank’s Results Matrix, outcomes refer to both goal and purpose, whereas outputs refer to project components. The identification of two statements of intent for objectives at the outcome level (goal and purpose) is somewhat arbitrary, given that outcomes can be imagined to be distributed both temporally and in terms of “depth”. Nevertheless, it provides a useful taxonomy for distinguishing between short- and long-term and between proximal and distal results. Regardless of this differentiation, the definition of a change in the development situation that is attributable to the project’s intervention and is measurable in terms of outcomes and impacts is the centerpiece of evaluability, as it provides focus to the analysis as a whole, from the Diagnostic section to the Monitoring and Evaluation Section.

Since every project is a response to a problem situation, the top-level objective of a project is a description of the development situation once the diagnosed development problem is solved. The appropriate description of that objective is a key element to the project’s evaluability. A project with diffuse or poorly defined objective may be interpreted in many different ways, and it might be entirely possible that no consensus is reached regarding what the project has been attempting and under what conditions.

For its part, changes in the problem situation which can be addressed after project completion are typically referred to as the project purpose in the current normative. The objective should indicate the direct impact or immediate result that will result from the utilization of the product and/or services provided by the project. This implies that the definition of the objectives should incorporate temporal dimensions and qualitative and quantitative measures of success or failure.

Project objectives cannot be understood to be an “empty box”: as something that can be described in several ways without affecting essentially the content of the project. On the contrary, the objectives constitute the link between the diagnosis and the defined activities and components, therefore their proper understanding and interpretation is linked to the context in which they are presented.

Objectives should be carefully defined, and it is necessary that the greatest level of consensus possible be achieved in their definition insofar as it is in that definition that the criteria for evaluating the success or failure of the project are found. This means that the objectives should clearly define how the performance of the project would make the problem less severe than it would be without the project. Objectives thus clearly describe the expected outcomes that will accrue to the identified population of project beneficiaries.

The definition of the objectives should satisfy six basic criteria: they must be specific, measurable, agreed upon, realistic, and have temporal characterization (SMART). They should also clearly identify the beneficiaries and those responsible for their achievement. Note that these criteria apply to how objectives are measured also (see below).

Specific means that they should have sufficient clarity and precision so as to impart a common understanding of Bank intent, regardless of the reader’s background or priors. In instances where sector terminology and conventions are used, they should be adequately explained so as to impart the specific meaning of terminology. Specific also requires that objectives clearly delineate the universe of expected beneficiaries.

Measurable indicates that the objectives should be defined in such a manner that it is possible to determine the project’s degree of success or failure in terms of the level of achievement of the defined objective. This condition typically requires measurement of both the quantitative and qualitative dimensions of expected benefits.

Agreed upon implies that the objective is shared by both the Bank and other stakeholders, including the executing agencies, country counterparts, and generally those parties who are involved. This should be based on which objectives will be used for the final measurement of the success or failure of the activity.

Realistic means that the objectives should be defined in such terms that render their achievement possible.

Temporal characterization is understood to be the condition of the objectives by which these are subject to a clearly bound timeframe within which they are to be achieved, and to specific milestone to indicate progress.

C. Logic of the Project

The logic of the project constitutes one of the key elements for its evaluability. The analysis of project logic must be carried out taking into account two different elements: substantive logic and formal logic.

- Substantive logic refers to the analysis of the reasons that justify the project's existence and the selection of its objectives;
- Formal logic is understood to be the analysis of the relationships among the different elements of the project, or the project development model.

The project's substantive logic establishes the degree to which the project's diagnostic and context information provided justify the project at that point in time. In other words, is there an empirical basis for establishing the need for the project, where need is specified in terms of improvements for beneficiaries? Likewise, does the project provide the information necessary to establish that the selected approach would be preferable? And does the evidence given indicate that prevailing conditions would warrant such an approach on the timeline specified, and under the specific conditions identified in the project documentation, including specific risks?

From the point of view of the formal logic, the analysis has been based on two basic elements extracted from the results matrix:

- The vertical logic: the relationship of cause and effect between the different parts of a problem that correspond to the four levels of objectives. More specifically, the relationship between:

Inputs→Activities→Outputs→Outcomes (and impacts)
- The horizontal logic: the relationship or correspondence that links each level of the objectives with the measures of achievement (indicators and means of verification) and with the conditions that could affect its execution and subsequent performance (risks).

The vertical logic postulates that if certain inputs and activities are carried out, then certain components (or outputs) will be obtained. This will create a necessary and sufficient relationship between the proposed activities and their corresponding outputs, as long as the assumptions hold. The production of these outputs will contribute to the achievement of the identified outcomes. Therefore, the project's vertical logic defines the project's development model, also referred to in the literature as the project's theory of change. Note that the vertical logic assesses both the relationship between the components described and the magnitudes required for these relationships to hold. In other words, are, say, the magnitude of activities and outputs congruent with the magnitude of identified outcomes?

The above presentation supposes that intervention models are well identified, and with known parameters. However, in some instances the project may serve both a development purpose as well as an opportunity to test intervention models whose parameters are not well estimated—at least not in the context in which they are being fielded. In these instances, part of the project’s vertical logic should include the degree to which project components (and evaluation framework) will allow for the testing of specific hypothesis regarding the channels of the intervention model. To the extent that this experimental nature is recognized as part of the Bank’s intent, and that the project is structured in order to test these precepts, a project’s vertical logic can be partially defined and still be evaluable.

Horizontal logic is closely related to the indicators specified for measuring progress toward stated intent, and when applicable for monetizing these benefits in dollar terms. It thus reflects how stakeholders will be able to measure the achievement of the expected results at each level of the previously described ascending scale. In this regard, it is essential, along with the aforementioned adequate definition of the objectives, that the project incorporates appropriate indicators for each level of the scale, and that it use these indicators in cost-benefit assessment exercises, which translate a vector of indicators into a measure of relative effectiveness.

Cost-benefit assessments (CBA) are usually used as a mechanism for comparing benefits with costs across different public policies. However, CBAs can also serve the purpose of further clarifying and making explicit the intervention model. By providing parameter estimates for the causal channels by which project components can benefit specific populations, CBA can also be beneficial in specifying the project’s vertical logic.

However, this exercise is only as useful to the extent that the assumptions underlying the CBA are well-founded and, more broadly, the analysis is robust. In this context, the technical quality of the CBA, and the assumptions underlying the CBA have to be evidence-based, explicit, credible, and in accordance with good practices.

D. Assumptions and Risks

Given the diversity in the use of the terminology, it is necessary to clearly identify the concepts of “risk” and “assumption.” The term risk denotes two aspects: uncertainty about its occurrence and the impact generated by its occurrence. From this perspective, we can define risk as the probability of occurrence of an event that is exogenous to the program and that will in some way impact the achievement of its objectives. Risks are usually identified—and this is the Bank practice—in terms of events that hinder achievement of objectives, although the definition could also be applied to those that enhance achievement.

In contrast to risks, “assumptions” are those events or situations whose occurrence or satisfaction is necessary for the achievement of project objectives. The assumptions and risks are outside the executor’s control but the assumptions are formulated as positive conditions and are to be directly related to the project itself.

Assumptions and risks therefore identify the universe of events, exogenous to the project, that are either necessary conditions for effectiveness or which may impact effectiveness. Many of the risks can be identified ahead of time, and in these cases it is possible that mitigation measures can be implemented, even in instances where the magnitude of the risk cannot be specifically identified or its probability of occurrence precisely specified

The evaluability instrument identifies two key elements for the assessment of project evaluability in this dimension: the identification of the conditions needed for the execution of the project and the achievement of the objectives and the identification of the follow-up measures to demonstrate the validity of the assumptions. The elements that have been taken into account to assess the projects have been the following:

- The quality of the analysis of the identification of the assumptions and risks.
- The presence or not of risk evaluation, meaning the quantification and gradation of the risks.
- The adoption or not of follow-up measures and the measurement of the assumptions and risks.
- The adoption or not of risk mitigation or incentive measures for the identified assumptions.

Risk identification should focus on those elements with significant impact, not only on the execution of the project but also on the achievement of its objectives and goals. The identification of these risks therefore requires not only a careful analysis of the project and a review of past experiences, but also an analysis of the sociopolitical reality in which the project will be carried out. It is in this aspect where the risk analysis complements the diagnosis.

Risks can be grouped into three large categories:

- Known risks, which are those that are derived from the Bank's previous experience in the sector or country.
- Predictable risks, which are those that may be identified from the very definition of the project with regard to the content of the same or the sector in which the project is implemented.
- Unpredictable risks, which are those whose appearance depends on factors that are completely beyond the control of the project team, the executor, or the Bank, and the materialization of which is not assumable from the logical point of view.

Examples can illustrate the nature of the risks we are referring to. The institutional weaknesses of a country's public sector in addressing profound structural reforms constitute a type of "known risk". An example of predictable risk is resistance to changes brought about by a project when these changes imply an important shift in an organization's culture.

Of these categories, only the first two, known risks and predictable risks, should be the object of analysis in a project document. A great number of classifications could be made within these two categories, but for the effects of this exercise, one should be noted, namely that which distinguishes among:

- *Operational risks*, which are those that affect the normal development of the execution of the project. This category can include, among many others, political, social, economic, etc. risks.
- *Financial risks*, which affect repayment capacity.
- *Institutional risks*, whose occurrence can affect the Bank's institutional image or its perception by society or groups affected by the operation.
- *Development risks*, which are those that can affect the impact on the development sought by the operation.

The second element taken into account in the exercises was the quantification or gradation of the identified risks. This gradation should be produced at three levels:

- The level of probability of the occurrence.
- The degree of the impact on the project and on its objectives and goals.
- The ordering of the risks in terms of the need for intervention and mitigation. The previous gradation is a consequence of the two previous ones to the extent that those risks that present a greater index of probability and that can have a major impact will be deemed a priority.

The third element refers to the adoption of measures for monitoring and assessing risks. Once a risk has been identified and assessed, it is necessary to implement measures for monitoring changes in the probability of its occurrence or variations in its hypothetical impact. For this, it is necessary to establish a series of indicators and benchmarks that can alert us regarding its evolution.

The fourth and last of these elements is the adoption of mitigation measures. The selection of the specific measures should depend precisely on the results of the risk assessment performed: determining under which conditions to go ahead with the project, which mitigation measures to implement to prevent risks from occurring or to minimize their impact if they do materialize, and determining the conditions for the implementation of previously defined contingency plans.

The current Bank normative includes a risk review. This main results of this risk review should answer the questions above, and be reflected in the project document. The full result of the risk review should be available for inspection in the form of a standardized annex.

E. Outcome Indicators

The selection of appropriate indicators is key to the process of determining the attainment of project's objectives. A good indicator should specify information on three dimensions: quantity, quality, and time.

Other important characteristics of good indicators are: validity, reliability, accuracy, exhaustiveness, and mutual exclusiveness.

The condition of *validity* of an indicator refers to measuring what is really intended to be measured. The condition of *reliability* refers to the consistency or dependability of data and evaluation judgments when repeated observations using similar instruments under similar conditions produce similar results. The condition of *accuracy* implies that the indicator not be erroneous, meaning that it be based on the greatest number of distinctions possible, which requires a correct identification of the variables involved. The condition of *exhaustiveness* implies that the categories inferred from the same be sufficient so that the reality or situation to be measured not be fractured. Finally, the condition of *mutual exclusivity* implies that the indicator not overlap with other indicators in a way that affects the understanding of the event being measured.

Despite the indications in the preceding paragraph, it is possible that the construction of an indicator can present serious technical problems or be too costly. In such cases, it might be advisable to use *substitute or proxy indicators*. These are indicators that, without referring to the central issue, when measured, provide sufficient information to infer its evolution. The selection of these indicators should be made with great care and should be explicitly acknowledged.

The effort made in selecting and defining indicators can be hampered if baseline information is not available to serve as point of comparison. But, in turn, to the degree that measuring the changes in a particular situation is difficult or requires several indicators it may be necessary to predetermine the "value(s)" at which the objectives are deemed satisfied. It may also be convenient to define a point of reference or *benchmark* for each of the selected indicators on the basis of information derived from the experience of other countries.

Likewise, the indicators selected should reflect the changes in a situation within a given period of time and should establish "when" the objectives are to be fulfilled. *Milestones* should be defined to measure progress toward the expected target. The indicators selected should be measured at a reasonable cost and, if possible, by means of information generally available.

The quality of the indicators selected will affect greatly the efficiency and effectiveness of the monitoring and evaluation systems. In the cases where indicators are absent, evaluators are forced to reconstruct the past in order to compare it with the present. In these circumstances the assessment of project results is extremely difficult and margins of error are likely to be very wide.

The Results Matrix allows for the identification of different levels of outcome indicators. As mentioned before, the typology of purpose/goal can be useful in specifying Bank intent, and if this is the approach adopted, both should have contain indicators that measure both quantity and quality of the concept being measured. Furthermore, when a program contains a set of outcomes the assessment of the quality of indicators will be based by appropriately giving weight toward outcomes that are more central beneficiary welfare than those are less important. In this context a well-developed development model, can be useful in making value judgments regarding adequacy of indicators, as can a fully-fledged CBA. Furthermore, although intermediate outcome indicators should be specified, they will necessarily receive less weight than “higher order” outcome indicators.

F. Output Indicators

Output indicators present fewer complexities than outcome indicators. These indicators measure whether the proposed activities have been executed within the expected period of time, and if they have produced the desired outputs, in the desired manner and for the cost initially established and that these actions have produced the desired outputs in terms of quantity, quality and time.

Output indicators should be identified at the level of components. Output indicators should be brief, clear descriptions of each of the deliverables or the products that should be generated during the execution of the project. They should specify quantity, quality and the period of time expected for their delivery of the goods, works and services generated through the execution of the project.

As with the outcome indicators, the output indicators should be valid, reliable, and unequivocal. They should also be oriented toward the measurement of products, practical, verifiable, and independent.

G. Baselines for Outcomes

Baselines for outcomes provide *ex ante* information on conditions that are expected to change as the result of the project. Baselines establish where the project is starting from, and are essential for measuring progress and accomplishment at some later date. Baselines are essential for characterizing the program status *ex ante*—and therefore central to the diagnostic—and for measuring progress. Baselines for outcomes are not necessary for the estimation of program impacts, *strictu sensu*, in the case programs with randomized controlled trials (RCT). In these instances the *ex post* difference in the values of outcomes variables between treated and control groups is sufficient to identify program impacts. However, the practice of omitting baselines from RCTs should be

strongly discouraged, as they are useful for both longitudinal assessments of beneficiaries, and as a way of verifying the appropriateness of a RCT design.

The process for obtaining the baseline information has three phases: the first defines the problem or situation and identifies its characteristics; the second gathers the information that is pertinent to the theoretical formulation effected; and the third orders and analyzes the data according to pre-established criteria.

The phase of the definition of the problem and identification of the characteristics presented, in relation to the baseline information, are of crucial importance since this information cannot include all the variables that can be of interest, but rather must select those that are relevant.

In instances in which a full baseline characterization is not possible *ex ante*, an attempt should be made to identify key outcome variable values, and provisions for a more thorough and reliable collection of data should be identified and agreed upon with the borrowing country. Building such data-gathering capacity is both a contribution to the individual project and a contribution to the development of the borrowing country's own evaluation capacity.

H. Baselines for Outputs

Baselines for outputs define the goods and services being offered prior to the initiation of the project. With the exception of very specific circumstances (such as Greenfield projects), projects typically modify or add to an existing flow of benefits, both for SG or for NSG operations. Therefore, projects must identify the baseline of outputs being delivered as the delivery status *ex ante*. Likewise, for operations that identify as their components changes in normative or legislation, baselines for outputs should not be defined as the absence of the change in normative/policy/legislation, but rather the qualitative characterization of the existing normative/policy/legislation before the intervention.

I. Monitoring and Evaluation

Interventions should have both monitoring and evaluation components. Monitoring and evaluation systems serve different purposes, and require different institutional arrangements and data requirements. These requirements should be addressed fully.

Monitoring implies the development of four basic actions: gathering, describing, explaining, and interpreting information with the objective of transforming it into useful inputs for the execution of the project. Monitoring systems have the following objectives:

- To determine the progress of the program with regard to the established goals, and thereby help to determine the success or failure of the program with respect to obtaining output targets.

- To identify the problematic areas or obstacles encountered during its implementation in a timely manner, so that corrective measures can be taken.

In the formal aspects, an appropriate monitoring system implies the following:

- A specific entity, with the necessary bureaucratic and administrative arrangements, is responsible for conducting monitoring activities.
- The entity is integrated into the structures of the administration of the corresponding sector and helps to construct reinforce the institutional capacity for an adequate monitoring of the programs and projects, thereby guaranteeing future sustainability.
- A clear conceptual framework of the monitoring program has been defined, including appropriate baseline information collected.

A good monitoring system should ensure:

- That the established monitoring system guarantees feedback to the execution process in order to eliminate project execution obstacles or problems, adjusting the expected time frames, defining and implementing the systems and procedures that improve the efficiency of the execution process.
- That the identification of the outcome indicators and data sources avoid the duplication of information-gathering efforts, employing for such purpose, whenever possible, available sources of data.
- That milestones are set to allow for the detection of degrees of progress in terms of certain specific periods of time.
- That the factors, conditions, or elements that could introduce measurements errors or biases in terms of variations in the information-gathering systems, or as a cause of the variations in information-gathering criteria or standards, be identified.
- That the total costs implied by the defined monitoring system be rational and proportionate to the total cost of the project.

Evaluation is linked to, but separate from, project monitoring. Evaluation activities are designed to provide feedback on results achieved, and should be carried out by individuals or organizations not directly responsible for execution. Evaluations can be of two types: in-process and *ex post*. Evaluations can also be central in providing the mechanisms required for testing development hypothesis regarding the development model being implemented.

In process evaluations are developed during the execution of the project and analyzes the management, some of its activities or components, or the intermediate outcomes obtained

in terms of the purpose of the operation. Examples of such evaluations are the mid-term reviews held for some Bank projects in which executing agencies directly responsible for execution meet with Bank staff not directly involved to discuss progress and problems.

Ex post evaluations are developed once the execution is concluded, or at such a time that is consistent with the time required for the materialization of specified project outcomes. Likewise, while the first is directed toward introducing changes or improvements in the execution, or toward introducing changes in the programming of the same, the second attempts to assess results and transform them into appropriate knowledge applicable to future operations.

In the case of evaluations, the following questions are relevant and should be addressed in an evaluable proposal:

- Does the proposal specify what will be learned as a result of the evaluation and which questions will be answered? And is the evaluation's contribution to existing knowledge consistent with the evaluation effort expense? Existing knowledge regarding program effectiveness and the channels by which programs operate differs widely across projects and from one context to another. Evaluations should be structured so that resources employed and effort expended is commensurate with what one hopes to learn regarding effectiveness.
- Does the proposal identify a method and strategy for answering key questions? Is this done so in a clear and precise manner? Is the method appropriate for the questions being asked and for the nature of the intervention proposed? A central question in evaluation is effectiveness. In order to answer this question a comparison of with and without program—or at least a comparison of with program versus with an alternate program is usually required. This in turn will require different methodological approaches, depending on the nature of the intervention, the availability of data and costs associated with producing additional data, as well as ethical issues related with experimentation, among others.
- Is the operation structured so that project components design is consistent with the requirements of the evaluation strategy? In many cases, particularly in cases where the intervention model is being tested, project components and project execution will have to be integrated *ex ante* in order for an adequate *ex post* evaluation to be conducted. This is most evident in the case of randomized control trials (RCTs), where benefits need to be randomized, but it is also true in the cases where the data-generating process is otherwise connected to the execution of program components.
- Who performs the evaluation, what are the required basic conditions and qualities, and what is the budget assigned to the same and its source?

- What are the requirements of stakeholders, i.e., what are the time needs, the resource needs, or the information needs that must be reinforced to ensure a reasonable development of the evaluation in the country or sector where this will occur?
- In what form or manner is the planned evaluation connected to the evaluation systems established in a country or institution in order for it to generate aggregate value of the country's evaluative capacity?

The evaluability instrument incorporates five key aspects to be considered by the project teams in the area of monitoring and evaluation to ensure the evaluability of the program. These elements are:

- Bank and borrowers have defined a data gathering system to generate information on indicators.
- Resources have been identified and committed to ensure that predefined data will be collected and analyzed.
- Provisions have been made for using the information for project monitoring.
- Sources of information are specified for all indicators.
- An evaluation strategy is presented, with an adequate identification of method, data sources. Project components are designed in a manner consistent with the evaluation strategy.

EVALUABILITY SCORING SHEET

The following scoring sheet is intended as an aid for scoring project evaluability. The scoring sheet is based on OVE's guideline and practice in scoring evaluability, as well as on the questions contained in the DEM regarding evaluability. The scoring sheet adopts the principle that projects are scored according to evaluability dimensions based on the questions identified in the evaluability guidelines contained in Annex A. The sheet requires scores to be produced for each of the evaluability dimensions identified in Annex A. It does not score individual questions, but rather the collection of questions that deal with each of the evaluability dimensions. The scoring is intended to be done by peers or by a panel, so as to limit evaluator-specific error across projects. The scoring system will adopt four categories, with two being adequate and two being inadequate categories. Reviewers should assess the degree to which guidelines are complied with.

Evaluability Scoring Sheet
<p><u>A. Diagnosis (1-4)</u></p> <ul style="list-style-type: none"> • The problem or need that the project attempts to address is clearly identified and analyzed. • The proximate and root causes of the problems are identified and analyzed. • The potential beneficiaries are clearly identified. • The diagnostic is data-based and empirically-driven, including data from prior Bank experience.
<p><u>B. Definition of Objectives (1-4)</u></p> <ul style="list-style-type: none"> • Expected results of the intervention are clearly identified and explained, and objectives are <i>SMART</i>. • When the intervention will produce several flows of benefits and results, objectives will encompass all benefits. • Objectives should be specific to the beneficiary population identified.
<p><u>C. Project Logic (1-4)</u></p> <ul style="list-style-type: none"> • Project vertical logic holds, and in particular the project model of theory of change is clearly explained, and consistent with: Activities→Inputs→outputs→Outcomes • All Components contribute to the achievement of stated objectives. • Project substantive logic holds. In particular, proposed solution and objectives are consistent with the problematic presented in the diagnostic. The diagnostic is a sufficient condition to demonstrate the need for the project. • The magnitude of solutions is consistent with the magnitude of the problem identified and the scope of the objectives identified.

Evaluability Scoring Sheet
<ul style="list-style-type: none"> • In the case of projects which attempt to test development models, the program contains the required elements for these hypotheses to be tested. • Project contains a CBA, which is (i) is of adequate quality, (ii) consistent with the intervention model provided, (iii) uses reasonable assumptions regarding benefits and monetization, (iv) and is empirically-based.
<p><u>D. Assumptions and Risks (1-4)</u></p> <ul style="list-style-type: none"> • The conditions required for the execution of the project and for the achievement of objectives are identified, as are factors which can inhibit or impede execution. • Risk factors are characterized empirically, and their relevance to the achievement of objectives is clearly stated. • Mitigation measures are identified; and for risks which cannot be mitigated, these are stated as such.
<p><u>E. Output indicators (1-4)</u></p> <ul style="list-style-type: none"> • The output indicators are defined and are sufficient to track the evolution of project outcomes. All project components contain a set of indicators. • Indicators are specific and measurable. • Output indicators clearly specify expected target levels during and at the end of project.
<p><u>F. Outcome indicators (1-4)</u></p> <ul style="list-style-type: none"> • Outcome indicators (at the goal and purpose level, or other level of outcome, as defined by the project) identify quantitative and qualitative measures of expected results (outcomes) at end of project execution. Indicators match the specified beneficiary population. • Indicators are sufficient to measure the achievement of project outcomes, and all project outcomes have appropriate sets of indicators identified. • Indicators are specific and measurable. • Outcome indicators clearly specify expected target levels during and at end of project, or at a time-frame consistent with the nature of the intervention (e.g. goal indicators may be longer-term).
<p><u>G. Output indicator baselines (1-4)</u></p> <ul style="list-style-type: none"> ▪ Baseline data is provided for all output indicators, except for specific instances in which no flow of outputs existed prior to the project, in which case an implicit value of zero is appropriate. ▪ Baselines are presented for all indicators.
<p><u>H. Baseline measures for outcomes (1-4)</u></p> <ul style="list-style-type: none"> • Baselines for project outcomes are provided for all indicators.

Evaluability Scoring Sheet

- In instances in which these have not been collected, clear indication is available of when they will be collected, and how.

I. Monitoring and Evaluation (1-4)

- An agreed-upon data gathering system to generate monitoring information on outputs is identified, along with resources, institutional arrangements, and provisions for data collection and reporting.
- A system for incorporating monitoring information into project design and improvements is described (if it is already in existence) and any changes to this system are specified.
- A full set of questions and hypothesis to be tested are presented.
- An evaluation strategy along with a technical methodological approach is defined and is appropriate for the intervention and data environment.
- An agreed-upon data gathering system to generate information on outcomes is identified, along with resources, institutional arrangements, and provisions for data collection and reporting.
- Sources of information are specified for all indicators.
- Project components and execution provisions are consistent with the needs of the evaluation strategy.