

# Guidance Manual for Independent Evaluation Group Validators

Implementation Completion  
and Results Report Reviews for  
Investment Project Financing



**IEG**  
INDEPENDENT  
EVALUATION GROUP

**WORLD BANK GROUP**  
World Bank • IFC • MIGA



**IEG**  
INDEPENDENT  
EVALUATION GROUP

**WORLD BANK GROUP**  
World Bank • IFC • MIGA

# **Guidance Manual for Independent Evaluation Group Validators**

## **Implementation Completion and Results Report Reviews for Investment Project Financing**

Last Revision: May 2024



# Contents

Abbreviations .....	v
Introduction .....	vii
What Is an Implementation Completion and Results Report? .....	vii
What Is an Implementation Completion and Results Report Review? .....	viii
On What Basis Does Independent Evaluation Group Assess Projects? .....	viii
What Are the Main Ratings? .....	viii
Structure of this Manual .....	ix
1. Procedures for the Implementation Completion and Results Report Review .....	1
Responsibilities of the Independent Evaluation Group Implementation Completion and Results Report Reviewer .....	1
Independent Evaluation Group’s Implementation Completion and Results Report Review Process .....	1
2. Guidance Manual .....	7
Section 1: Project Data .....	7
Section 2: Project Objectives and Components .....	7
Section 3: Relevance of Objectives .....	12
Section 4: Achievement of the Objectives (Efficacy) .....	14
Section 5: Efficiency .....	24
Section 6: Project Outcome .....	27
Section 7: Risk to Development Outcome .....	34
Section 8: Bank Performance .....	36
Section 9: Quality of Monitoring and Evaluation .....	41
Section 10: Other Issues—Safeguards, Fiduciary Compliance, and Unanticipated Impacts .....	45
Section 11: Ratings Summary .....	50
Section 12: Deriving Lessons .....	50
Section 13: Assessment Recommended .....	54
Section 14: Quality of the ICR .....	54
3. Other Considerations for the Implementation Completion and Results Report .....	58
4. Guidance Specific to Fragility, Conflict, and Violence Context .....	61
Project Context and Development Objectives .....	61

## Contents

Outcome.....	62
Bank Performance, Compliance Issues, and Risk to Development Outcome.....	66
Lessons and Recommendations .....	68
5. Note on Canceled Operations .....	70
What Is a Note on Canceled Operation?.....	70
Which Sections of the ICRR Review Should Be Completed, and What Ratings Assigned? .....	70
Rating the Quality of the Note on Canceled Operation .....	70
References.....	73

## Figure

Figure 2.1. Results Chain.....	18
--------------------------------	----

## Tables

Table 2.1. Deriving the Overall Outcome Rating for a Project, Tree View .....	29
Table 2.2. Deriving the Overall Outcome Rating for a Project, Table View.....	31
Table 2.3. Overall Outcome Ratings.....	34
Table 2.4. Relation among Quality at Entry Quality of Supervision, and Overall Bank Performance Ratings .....	41
Table 2.5. The Difference among Facts, Findings, Lessons, and Recommendations .....	52
Table 3.1. Red Flags” or Examples of Ratings Patterns to Check .....	60

# Abbreviations

BP	Bank Procedure ( <i>World Bank</i> )
CPF	Country Partnership Framework
DLI	disbursement-linked indicator
DRRs	drivers of fragility, sources of resilience, and risk factors
E&S	environmental and social
ERR	economic rate of return
ESCP	Environmental and Social Commitment Plan
FCV	fragility, conflict, and violence
ICR	Implementation Completion and Results Report
ICRR	Implementation Completion and Results Report Review
IEG	Independent Evaluation Group
IPF	Investment Project Financing
IRR	internal rate of return
M&E	monitoring and evaluation
NCD	noncommunicable disease
NCO	Note on Canceled Operation
OP	Operational Policy ( <i>World Bank</i> )
OPCS	Operations Policy and Country Services
PAD	Project Appraisal Document
PDO	project development objective
TOC	theory of change
TTL	task team leader



# Introduction

## What Is an Implementation Completion and Results Report?

The Implementation Completion and Results Report (ICR) is one of the World Bank's main instruments for self-evaluation. It is prepared by the World Bank at the close of every project funded by the International Development Association or the International Bank for Reconstruction and Development or, in the case of a series of programmatic policy operations, at the end of a series of projects. See the Operations Policy and Country Services ICR Guidelines for additional information and exceptions.

According to the guidelines for World Bank staff for preparing ICRs, they are intended to do the following:

- Provide a complete and systematic account of the performance and results of each project
- Capture and disseminate experience from the design and implementation of a project, to (i) improve the selection of interventions to achieve the goals of the Country Partnership Framework (or, previously, the Country Assistance Strategy); (ii) improve the design and implementation of interventions through lessons learned; and (iii) help ensure greater development impact and sustainability of projects
- Provide accountability and transparency at the level of individual projects with respect to the activities of the World Bank, the borrower, and involved stakeholders
- Provide a vehicle for realistic self-evaluation of performance by the World Bank and borrowers
- Contribute to databases for aggregation, analysis, and reporting, especially by the Independent Evaluation Group (IEG), on the effectiveness of lending projects in contributing to development strategies at the sector, country, and global levels

ICRs are intended to contribute to accountability and learning both internally (members of the Board of Executive Directors and World Bank managers and staff) and externally (governments and their agencies, stakeholders, beneficiaries in partner countries, and the general public). The final ICR is publicly disclosed when it is submitted to the Board.



## **What Is an Implementation Completion and Results Report Review?**

The Implementation Completion and Results Report Review (ICRR), conducted by IEG, is an independent, desk-based, critical review of the evidence, results, and ratings of the ICR in relation to the project's design documents. Based on the evidence provided in the ICR and an interview with the last task team leader (TTL), IEG arrives at its own ratings for the project, based on the same evaluation criteria used by the World Bank.<sup>1</sup> IEG reviews each ICR that is submitted to IEG.

The ICRR is an independent validation of the World Bank's self-evaluation and ratings; it is not an independent evaluation of the project based on evidence collected outside the World Bank's self-evaluation process.<sup>2</sup> The ICRR is intended to critically assess the evidence provided in the ICR, its quality, and the attribution of results to the activities or actions supported by the project under review. It is not simply a summary of what is in the ICR.

ICRRs serve as an independent validation of the results reported in the ICR and contribute to both learning and accountability. They also provide a systematic way for IEG to critically review the evolving portfolio as projects close and to summarize the projects' objectives and key results, in addition to the ratings. The write-ups are stored in a searchable database within IEG and, for all projects that closed from fiscal 2011 onward, are posted on IEG's external website. They are often useful as a starting point for IEG's ICR reviewers as a quick way to identify projects of different types—with specific objectives or activities—in preparing to undertake larger country, sector, or thematic evaluations.

## **On What Basis Does Independent Evaluation Group Assess Projects?**

The World Bank and IEG share a common, objectives-based project evaluation methodology for World Bank projects to assess achievements against each project's stated objectives, and the relevance of the objectives and the efficiency of resource use in achieving the objectives. An advantage of this methodology is that it can accommodate country context in terms of setting objectives that are reasonable. The World Bank and the governments are accountable for delivering results based on those objectives.

## **What Are the Main Ratings?**

There are three main project ratings that IEG validates through the ICRR, and one rating that is assigned by IEG only.<sup>3</sup> The three main project ratings are the following:

- **Outcome.** The extent to which the project’s major relevant objectives were achieved, or are expected to be achieved, efficiently. Both the World Bank and IEG rate outcome.
- **Bank performance.** The extent to which the services provided by the World Bank ensured quality at entry of the project and supported effective implementation through appropriate supervision (including ensuring adequate transition arrangements for regular operation of supported activities after loan or credit closing), toward the achievement of development outcomes. Both the World Bank and IEG rate Bank performance. IEG also rates the two constituent elements—quality at entry and quality of supervision.
- **Monitoring and evaluation quality.** The quality of the design and implementation of the monitoring and evaluation arrangements of the project and the extent to which the results are used to improve performance. Both the World Bank and IEG rate monitoring and evaluation quality.<sup>4</sup>

IEG assigns one additional rating based on the material presented in the ICR:

- **ICR quality.** The quality of the evidence and analysis in the ICR; the extent to which the lessons are based on evidence; the results orientation of the ICR; and its conciseness, internal consistency, and the consistency with World Bank guidelines. Only IEG rates ICR quality.

## Structure of this Manual

The manual is organized as follows, with appendixes.

- Section 1 provides an overview and explains the responsibilities of the IEG ICR reviewer, the materials to consult, and the ICRR process.
- Sections 2–15 cover identification of the objectives and the criteria for the main ratings and definitions and criteria for other issues covered in the ICRR that are not rated, such as the risk to development outcome; safeguards, fiduciary issues, and unintended outcomes; and lessons.
- Section 16 provides a basic checklist.
- Section 17 provides guidance on reviewing ICRRs in fragility, conflict, and violence settings.
- Section 18 discusses assessing canceled projects, for which the World Bank will issue a Note of Canceled Operation in lieu of an ICR. These Notes are also reviewed using the ICRR form.



# 1. Procedures for the Implementation Completion and Results Report Review

## Responsibilities of the Independent Evaluation Group Implementation Completion and Results Report Reviewer

The Independent Evaluation Group (IEG) Implementation Completion and Results Report (ICR) reviewer is responsible for the following:

- Correctly completing the Implementation Completion and Results Report Review (ICRR) form in the online IEG ICRR Portal system, following the guidelines and procedures in this manual and in the Onboarding Guidelines for IEG ICR reviewers
- Assigning ratings based on the evidence in the ICR and that are gleaned by consulting the other key documents listed in the Assembling the Key Documents subsection of this chapter.
- Meeting with the project's last TTL, recording a summary of the meeting, and updating the draft ICRR and ratings to reflect any new and relevant information
- Revising the ICRR and ratings based on comments from any or all of
  - A member of the review panel (composed of senior staff and consultants in IEG),
  - The ICRR coordinator, or
  - The IEG manager.
- Reviewing written comments from the Global Practice or Region that managed the project, incorporating any new and relevant information, correcting any inaccuracies, updating ratings if warranted, and drafting a response to the Regional director to explain any updates

The review process and the specific steps involved and expectations of the IEG ICR reviewer at each stage are discussed in the next subsection.

## Independent Evaluation Group's Implementation Completion and Results Report Review Process

The ICR for a project arrives in IEG electronically after it has been approved in the Operations Portal and sent to the Board through the relevant system. The ICRR

coordinator then assigns it to an IEG ICR reviewer for review. A blank ICRR form for the project is automatically created in the IEG ICRR Portal, based on its project ID. Certain fields in the basic data portion of the form will be automatically populated (the project name, project ID, and TTL).

## Assembling the Key Documents

In preparing the first draft of the ICRR, the IEG ICR reviewer assembles the key documents listed here. The IEG ICR reviewer is not expected to go beyond these documents to look for additional evidence:

- **The Financing Agreement** (loan, credit, or grant agreement)—primarily for use in verifying the project’s original objectives and components. If the legal agreement was amended, the amended agreement(s) will also be provided.
- The **Project Appraisal Document** (PAD; for investment projects)—primarily for use in identifying the project’s original objectives, components, planned amounts, cofinanciers, results framework, planned monitoring and evaluation (M&E) and the presence of baseline information, safeguard category (for investment projects), and other aspects of design. If the project has been restructured, there will also be a project paper.
- The **Country Partnership Framework** (CPF; previously known as Country Assistance Strategy) in effect at project closing (and the CPF in effect at approval if different)—primarily for use in assessing the project’s relevance.
- **Implementation Completion and Results Report** (ICR)—the World Bank’s self-assessment of the project and the main document to review. The ICR includes information on *revisions to the design* (such as restructuring and changes in objectives or components, allocations, cofinanciers, or expected counterpart contributions); the *implementation of project activities*; the implementing unit’s assessment of the project’s *outcomes*, the *relevance of the project*, the *achievement of its objectives*, the project’s *efficiency*, and *safeguard and fiduciary compliance*; operational staff’s *self-ratings* (on outcome, Bank performance, and M&E quality); and the lessons learned from the experience. These reports often include an assessment by the borrower as an appendix and, occasionally, an assessment by cofinanciers of the results, in addition to financial or economic analysis and the results of beneficiary surveys.<sup>1</sup>

## **Preparing the Initial Draft**

The IEG ICR reviewer is expected to read all the above documents and use that information as indicated in this manual and its appendixes, and as in the Operations Policy and Country Services (OPCS) ICR Guidelines, to draft the ICRR. Before finalizing the draft, the IEG ICR reviewer must contact the last TTL of the project to set up an interview. The detailed checklist for completing the ICRR is in appendix A, and illustrative examples of assessment are provided in appendix C of the “Reference Annex: Illustrative Examples for Independent Evaluation Group Validators.”

## **Interviewing the Last Task Team Leader**

This interview, which is conducted before the draft ICRR is finalized, provides an opportunity for the last TTL of the project to offer additional views or information to the IEG ICR reviewer (beyond what is in the ICR) about the project experience. The interview also provides an opportunity for the IEG ICR reviewer to pose any follow-up questions that arose in the course of reading the ICR, to improve the accuracy and quality of the ICRR. The IEG ICR reviewer should contact the TTL early in the process of drafting the ICRR, to check availability and travel schedule, and to propose a meeting either in person or by audio. If the TTL is not responsive after 10 business days to the request for a meeting, the ICR reviewer may proceed to finalize the draft. Note that the finalized draft ICRR is later shared with the Global Practice for comments, and at that point the TTL also will have an opportunity to provide any additional information. The IEG ICR reviewer should not share the draft Review with the TTL, however, or share the proposed ratings.

After the TTL interview, the IEG ICR reviewer writes a summary of the meeting, uploads the summary to the activity history for this ICRR in the IEG ICRR Portal system, and shares it with the panel reviewer when they are identified. The detailed protocol for the TTL interview is in appendix B of the “Reference Annex: Illustrative Examples for Independent Evaluation Group Validators.”

## **Submitting the Draft for Panel Review**

After the interview with the TTL, the IEG ICR reviewer updates the ICRR with any new and relevant information from the TTL. After a final “Save” and “Exit,” the IEG ICR reviewer uses the activity history pane to send the ICRR back to the ICRR coordinator, who then identifies an appropriate panel reviewer and assigns it to a panel reviewer.

## **Panel Review**

After the IEG ICR reviewer sends the completed draft ICRR to the ICRR coordinator, the coordinator selects an appropriate panel reviewer from the IEG evaluation panel, comprising senior evaluators. The tasks of the panel reviewer are to

- Review the same documents as the IEG ICR reviewer;
- Read the ICRR;
- Ensure that the objectives have been properly identified;
- Ensure that the guidelines have been properly applied;
- Ensure that the ICRR is complete, internally consistent, and sufficiently critical of the quality of the data and analysis; and
- Comment on the ratings.

The panel reviewer provides comments to the IEG ICR reviewer by sending a message through the IEG ICRR Portal system. Many panel reviewers find it convenient to download the draft ICRR as a Word document, add comments to that document using track changes or comment boxes, and attach the Word document along with the message to the IEG ICR reviewer.

Comments from a panel reviewer generally indicate areas of agreement but also areas for improvement, areas of disagreement, and queries about the evidence.

The IEG ICR reviewer then revises the ICRR as needed and responds to the panel reviewer as needed. The discussion can go back and forth several times. If the panel reviewer and IEG ICR reviewer cannot reach agreement, the ICRR coordinator can step in. When the panel reviewer is satisfied that the ICRR is ready and there is agreement, they clear the ICRR in the system to send it to the ICRR coordinator.

Recording these interactions in the portal is required to ensure consistency across ICRRs.

## **Quality Check by Implementation Completion and Results Report Review Coordinator, Independent Evaluation Group Unit Manager, or Both**

After clearance by the panel reviewer, the ICRR coordinator or IEG manager (or both) reviews the draft and sends any questions or comments to the IEG ICR reviewer and panel reviewer, who address the questions and comments as needed.

## **Regional Director Review**

After clearance by the relevant IEG manager, the ICRR coordinator or IEG manager (or both) sends the draft ICRR to the Regional director for comment, through the IEG ICRR Portal system.<sup>2</sup> The Regional director is responsible for forwarding it to the people on the country team most familiar with the project for comment and for coordinating the response to IEG. The Regional director has the option of inviting the borrower to comment. The standard review period for World Bank teams and the borrower is two weeks (10 business days).<sup>3</sup>

In the response, the Regional director or borrower may point to factual corrections needed, suggest changes in the text, or indicate disagreement on the ratings. Often, certain information already presented in the ICR is repeated, but the Regional director may also provide additional relevant and credible information concerning achievement of the objectives (or other aspects) not already in the ICR.

## **Finalizing Implementation Completion and Results Report Review and Responding to Regional Director Comments**

The IEG ICR reviewer should incorporate any additional relevant information from Regional director comments that is credible and will improve the accuracy of the assessment but indicate in the text that this additional information was “provided by the Region” (to distinguish it from the information in the ICR), and cite the source of the information, if known.

Since the IEG ICR reviewer has already fully assessed the project with respect to all the information in the ICR, a response from the Regional director that simply reiterates that same information would not be expected to result in changes in the ICRR. However, new information, corrections, or new (and compelling) lines of argument could result in changes in the ICRR text or ratings.

The comments from the Regional director should be discussed with the panel reviewer, and any proposed changes should be cleared by them. The ICRR form is then modified and saved, and the IEG ICR reviewer drafts a response to the Regional director, also cleared by the panel reviewer, to be sent from the ICRR coordinator or IEG manager (or both) back to the Regional director.

The ICRR coordinator or IEG manager (or both) sends the response and the final ICRR and instructs a designated staff member (not the IEG ICR reviewer) to post the ICRR.



## **Posting the Implementation Completion and Results Report Review**

After IEG's disclosure policy, each finalized IEG ICRR is publicly disclosed. IEG ICRRs are searchable on the IEG website.

## **Requests for Meetings during and after the Review**

In some instances, during the review period and occasionally after receiving the final ICRR, the Regional director will request a meeting with IEG. Normally, such a request would come with the written comments. The IEG ICR reviewer, the panel reviewer, and either the ICRR coordinator or the IEG manager generally attend these meetings. The purpose of the meeting is primarily to listen to the concerns of the operational team about the draft ICRR. The meeting also provides an opportunity for the IEG team to request clarification of specific points and seek additional information. Before agreeing to a meeting, IEG is expected to already have received the written comments on the ICRR.

## 2. Guidance Manual

### Section 1: Project Data

The IEG ICRR Portal system automatically pulls in project data from the Operations Portal for the project. The IEG ICR reviewer should note any discrepancies between the data appearing automatically and the information presented in the ICR document.

**Appraisal amounts.** The source of the project costs, loan or credit amounts, and cofinancing amounts at appraisal is the PAD (for investment lending). The project costs include the contribution of the World Bank, the government (counterpart funding), and any official cofinanciers.<sup>1</sup> Only the total project costs, the World Bank's contribution, and the official cofinancing (as elaborated in Board documents) are mentioned here. The government's contribution to the project is not considered cofinancing and is not recorded in section 1. The costs should be recorded in millions of US dollars.

*Cofinancing* refers to any arrangement under which World Bank funds or guarantees are associated with funds provided by third parties for a particular project or program. The third parties may be official or private. There are two ways of channeling cofinancing:

- **Joint cofinancing.** A joint project in which expenditures from a common list of goods and services are jointly financed in agreed proportions by the World Bank and the cofinancier.
- **Parallel cofinancing.** A project in which the World Bank and the cofinancier finance different services, goods, or parts of the project.

**Actual amounts.** Actual total project costs, loan or credit amounts, and cofinancing should be copied from the ICR. If there was additional or supplemental financing, the amount actually disbursed should be included in the total actual project cost; it should not be added to the block on appraisal amounts.

**Cofinanciers** should include donors other than the World Bank that provide official cofinancing, as mentioned in the PAD or program document, but should not include donors or partners acting as project executors or implementors.

### Section 2: Project Objectives and Components

#### Identifying the Objectives

The World Bank's evaluation architecture—both self-evaluation (reflected in the ICR and in supervision reports) and independent evaluation (IEG's assessments)—is objectives based. All the elements of the project outcome rating are linked to the

objectives: the relevance of the objectives, whether the objectives were achieved (efficacy), and whether they were achieved efficiently. Accurately identifying the objectives is thus essential to the entire evaluation exercise and is critical for ensuring accountability.

This section explains where the objectives can be found and the guidelines for interpreting them for the purposes of the ICR review.

The project's objectives include the statement of objectives, as articulated in the loan, credit, or grant agreement, and key associated outcome targets, if any. The statement of objectives should be lifted directly from the lending agreement (for investment lending)—not from the ICR.

See also the discussion in the next three sections on identifying the objective by parsing the project development objective (PDO) statement.

### **What Constitutes a Project's Objectives?**

A project's objective is a statement of what it intends to achieve, expressed in terms of an intermediate or final development outcome, as opposed to a financed deliverable (output). In its guidelines for the content of the PAD, OPCS recommends that the project's development objective(s) should

- (a) be stated as concisely as possible; (b) indicate the expected outcomes for the targeted project beneficiaries (specific group of people or institutions); and (c) focus on outcomes that the operation can realistically achieve at closing, given its duration, resources, and approach. It should neither encompass higher-level objectives beyond the purview of the project, nor be a restatement of the project's components or outputs. World Bank. 2024. *Preparing the Project Appraisal Document (PAD) for Investment Project Financing*, Page 5)

### **Where to Find the Objectives**

**The IEG ICR reviewer should always assess the PDO stated in the original and revised legal documents, and not take as given the PDO statement in the ICR, which is being assessed.**

For **investment projects**, take the PDO as stated in the lending agreement, development credit agreement, or grant agreement (the legally binding document negotiated between the World Bank and the government), in schedule 2, at the end of the agreement, entitled "Project Description."

If needed, additional information on the PDO can usually be found in the PAD. Such information may be found in the front matter or summary; in the section on “Project Development Objectives;” and in the technical appendix, “Detailed Project Description.” However, if the wording of the PDO diverges from that in the legal agreement, it is important to take the wording as in the legal agreement.

## Assessing Global Environmental Objectives

Projects wholly or partly financed by the Global Environment Facility will likely include global environmental objectives in the PAD, in addition to PDOs. Both the PDO and the global environmental objective from the PAD should be listed on the ICRR form, in addition to the objectives noted in the grant or legal agreement. However, the project is assessed based on the wording in the grant or legal agreement.

## Project Components, Cost, and Dates

### Components

In this section, provide a **summary** description of components (matching the PAD or lending agreement, not the ICR), with sufficient detail to make clear what activities were supported by project funds. No evaluation is needed here—only description—but any discrepancies in the description of components across the PAD, credit, or loan agreement and ICR should be noted.

List each component separately, followed (in parentheses) by both the appraisal and actual costs for that component in millions of US dollars. The estimated and actual component costs should add up to the estimated and actual total project costs, respectively.<sup>2</sup> If they do not, the reason for the difference should be noted—for example, if the estimates at appraisal exclude contingencies, but the actual costs do not.

### Comments on Project Cost, Financing, Borrower Contribution, and Dates

In this section, record the following information:

- **Project cost.** Estimated at appraisal and actual amount
- **Project financing.** Estimated at appraisal and actual amount
- **Borrower contribution.** Government contribution, both as estimated at appraisal and actual amounts
- **Dates.** Provide the approval date, effectiveness date, the Mid-Term Review date, and closing date (original and actual). Include the number of extensions to the project closing date and reasons for extensions.

- **Restructurings.** Discuss all project restructuring. In the case of formally restructured projects, use the proportion that was disbursed before and after the revision to establish what weight to give to revised objectives or targets in the overall outcome rating.
- **Split rating.** If the evaluation warrants a split rating, explain the rationale for carrying out split rating.

## Split Rating

If the project’s development objectives (or key associated outcome indicator targets, or both) have been revised (through a formally approved restructuring or additional financing), the ICR and the ICRR should take into consideration both the original and formally revised objectives in deriving the project’s overall outcome rating.

**Principle.** The principle of reporting on both the original and revised objectives is to ensure that the project demonstrates accountability for reaching the promised level of outcome, without penalizing the project for reasonable adaptations during implementation. To apply the split rating method, the ICR and ICRR assess achievement across the entire project time to assign separate efficacy ratings against both the original and revised project objectives or outcome targets, and relevance and efficiency ratings are given for the entire project at closing. Based on these ratings, separate outcome ratings are then derived (original PDO and targets and revised PDO, target, or both after project restructuring). To derive the overall outcome rating, separate outcome ratings are weighted in proportion to the share of actual loan or credit disbursements made in the periods before and after approval of the revision.

This methodology aims to reward project teams that recognize issues early and restructure accordingly. It also aims to provide an incentive against the late “lowering of the bar,” that is, lowering the ambition of a project to match achievement in the field rather than to account for changes in context.

## When to Apply a Split Rating

Revision of the PDO statement or outcome targets may or may not call for application of a split rating, depending on whether the scope of the project narrowed, expanded, or remained the same:

1. **Scope of the project narrowed.** If the project became overall less ambitious, generally a **split rating is applied** regardless of whether project funding increased (for example, through Additional Financing), decreased (for example, through cancellation), or remained the same—unless good reasons can be presented as to why the split rating does not make sense in a specific case.

2. **Scope of the project expanded.** If the project became overall more ambitious, generally a **split rating is not applied** regardless of whether project funding increased (for example, through Additional Financing) or remained the same— unless good reasons can be presented as to why the split rating makes sense in a specific case. Generally, the operation can be assessed on the basis of the more ambitious revised objective or outcome targets.
3. **Scope of the project remained the same.** If a project was restructured, but its scope remained the same, for example, geographic coverage was moved from one region to another similar region, or objectives or outcomes that were too broad or vaguely worded were just clarified—there may be no need to apply a split rating, and the ICR and ICRR can simply assess the project against the revised objectives, outcome targets, or both. However, the onus is on the ICR to make the case that the level of ambition, difficulty, and scope remained unchanged.

**Changes such as dropping or reformulating indicators and adding new ones do not necessarily trigger a split rating** because such changes may reflect different (and presumably better) measures of a project’s achievement rather than raising or lowering of a project ambition. Project teams are encouraged to adopt better indicators whenever necessary and appropriate. Similarly, changes in a project’s components, output targets, or both do not trigger a split rating, since these kinds of changes reflect a different (and presumably better) path, in the absence of a revision of PDO or outcome targets, to achieve the same expected outcomes. The IEG ICR reviewer should exercise judgment, based on the project’s scope and theory of change (TOC), to determine whether changes of PDOs or outcome indicators represent a better measurement of the expected outcomes. This judgment should be described and defended in the ICRR text.

In instances where a project’s objectives, outcome targets, or both are revised more than once, the above judgments and procedures should be repeated as necessary, determining the overall outcome rating according to the percentage of loan, credit, or grant disbursements under each restructuring.

**Section 2(b)** responds to the question, “Were the project objectives or key associated outcome targets revised during implementation?” If yes, then enter the approval date.

**Section 2(c)** responds to the question, “Will a split evaluation be undertaken?”

**Section 2(e)** should provide the reason why or why not the split rating methodology would be used for evaluation outcome.

## Section 3: Relevance of Objectives

### Relevance of Objectives

#### Definition

**Relevance of objectives** is the extent to which an operation's objectives are consistent with current World Bank and country strategies (expressed in the CPF) and are addressing significant constraints to identified development challenges. By assessing the relevance of objectives, the evaluation considers whether the World Bank's implementation support was responsive to changing needs and whether the operation remained important to achieving CPF or Country Partnership Strategy development objectives (which may change over time). If country circumstances changed significantly during implementation, discuss how these changes were accommodated (by changing the objectives through formal restructuring or other means) to retain the relevance of the objectives. In case there is no CPF, review consistency with the Country Engagement Note. For the purposes of the Review, "current" refers to the time of project closing.

#### Criteria

Relevance of objectives is assessed with respect to the country strategy agreed between the World Bank and the country. Assessment of relevance of objectives is based on the following criteria:

- **Alignment with strategy.** Look for an explanation to answer the question, "How (and to what extent) does the objective align with the country strategy at project closing?" This explanation should provide details on the strategic nature of the alignment, that is,
  - What development problem the project sought to address,
  - Where this development problem fits within an area where the World Bank contributes, and
  - How this problem was or is being specifically addressed through the project.

**Optional.** Looking for alignment between the PDO and World Bank CPF or Country Partnership Strategy at appraisal can provide good context regarding the project design.

- **Country context.** Country context helps in understanding the range of constraints affecting implementation of the operation. Look for an explanation to answer the question, "Is the objective outcome oriented and appropriately pitched for development status and capacity in the country as described in the CPF?" This judgment would incorporate government capacity, fragility issues,

and what is a reasonable expectation for the objective given the possible wide range of constraints in the operational context.

- **Sector context.** What were the main challenges in the sector the project planned to address?
- **Previous sector experience.** The World Bank's previous sector experience in the country would help in assessing how challenging the project's objectives should be in terms of the World Bank's engagement. Look for an explanation to answer the question, "What is the historical experience of the World Bank in the relevant country and sector?"

### Guidelines

The information can be found in the context and background section of the PAD. Explanations of the areas mentioned in the previous subsection are likely to appear in the ICR section on background and context, "Key Factors That Affected Implementation and Outcome," and in the sections on relevance of PDOs and efficacy, but any other section may also include relevant explanations.

The PDO should include a clear formulation that captures the intended project outcomes. This is important because it captures the transformational effect of the project and is linked to the ambition of the project.

The reviewer needs to assess the extent to which the PDO represents a meaningful contribution toward solving the development problem identified by the project. Also, assess the extent to which the PDO was pitched at a level that was commensurate with the development status and capacity of the country as described in the CPF.

The ICR review should clarify

- The nature of the development problem that the project intended to address;
- How this development problem fits in with the activities to which the World Bank contributed; and
- How the project addressed the development problem.

**Responsiveness to changing country needs.** If country circumstances changed significantly during implementation, did the project's objectives remain responsive to changing needs and did the project objectives remained fully relevant? An assessment of this must be provided in the ICRR. If the stated objectives are vague or not sufficiently monitorable, a relatively low rating can be appropriate for the relevance of objectives and for quality at entry.



## Rating of Relevance of Objectives

Relevance of objectives is rated on a four-point scale: **high**, **substantial**, **modest**, or **negligible**:

High	High relevance to country strategy. Full alignment between project objectives and country strategy. The project fully addresses the development problem and sought to address or contribute to the solution for the development problem.
Substantial	Substantial relevance to country strategy. Almost full alignment between project objectives and country strategy, or minor misalignments in limited areas.
Modest	Modest relevance to country strategy. Partial alignment between project objectives and country strategy. Or, if circumstances changed, the PDOs were not changed accordingly to keep objectives fully relevant.
Negligible	Negligible relevance to country strategy. Very little alignment between project objectives and country strategy.

## Section 4: Achievement of the Objectives (Efficacy)

### Definition

Efficacy is defined as the extent to which the project's objectives were achieved or are expected to be achieved, taking into account their relative importance, and are attributable to the activities or actions supported by the project.

For the purposes of this section, the objectives refer to each of the key *outcomes* indicated in the statement of PDOs from the legal agreement (credit, lending, or grant agreement).

### Criteria

Each objective is assessed based on the level of achievement and the concept of plausible causality. To establish this, for each objective, the IEG ICR reviewer should do the following:

1. Assemble and succinctly present the evidence from the ICR for each part of the results chain or causal chain supported by the project—the inputs and outputs—and the observed intermediate outcomes or impacts for each objective. Explain the interpretation of the extent to which evidence presented in the ICR supports the conclusion that the causal relationships asserted in the results chain are true, or likely, or, at least, plausible. Also, assess the extent to which the ICR presents evidence that it was at least plausible that the outcomes achieved arose from the activities and outputs of the project, as distinct from other nonproject factors—such as other interventions, policy changes unrelated to the project, natural events, or market factors. This type of analysis helps in making a reasonable assessment of plausible attribution of observed outcomes to the project’s specific interventions.
2. Discuss and (to the extent feasible) present evidence from the ICR of the contribution of other, nonproject factors leading to these outcomes (**the counterfactual**), with the intent of examining whether the achieved outcomes can plausibly be attributed to the government program or project supported by the World Bank.

### General Principles for Assessment of Efficacy

Efficacy is to be assessed at the time of evaluation (that is, at the time of the ICR, the ICRR, or the Project Performance Assessment Report, as are the other two constituent elements of outcome, that is, relevance of objectives and efficiency). For example, if a flood (or other natural disaster) had wiped out all project achievements *after* project closing such that at the time of evaluation there was nothing to be seen in the field, efficacy would not be rated favorably. (Bank performance could, of course, be rated in the satisfactory range depending on whether the World Bank had done everything possible to avoid the unfavorable outcome.)

Each stated objective is to be rated, even if the objective is stated in output terms. If an objective is stated in output terms, however, this can be an example of “setting the bar” too low. If the IEG ICR reviewer can make the case that a higher bar, or promising intermediate or final outcome (rather than outputs) would have been possible and desirable in the particular country circumstances, then this would provide support for a relatively low rating of relevance of objectives.

### Organizing the Assessment of Efficacy

In section 4 of the online ICRR form, the IEG ICR reviewer adds a section for each of the outcomes (objectives) that make up the statement of the objectives being assessed.

In the objective header field, the IEG ICR reviewer enters the text of the objective, as stated within the PDO.

In the objective rationale field, the IEG ICR reviewer should do the following:

1. Provide an overview of the TOC for the given objective. This should be based on the theory as articulated in the completion report but should also critique where there might be gaps or unjustified issues.
2. Assemble and succinctly present the evidence from the ICR (or fieldwork in the case of a Project Performance Assessment Report) that documents the realization of the complete results chain, from outputs to intermediate outcomes to final outcomes.
3. Comment on the extent to which the outcomes can be attributed to the project or program in question. These two elements are discussed below, in turn. A rating will be assigned to each objective.

### **Clarify the Specific Objectives by Parsing the Project Development Objective Statement**

Organize the assessment of PDOs in relation to each objective or outcome captured in the statement of objectives. Compound PDOs with multiple outcomes should be treated separately and unpacked. An example of a compound objective is several outcomes being linked together in a single sentence.

Some PDO statements articulate objectives (that is, expected outcomes) but also include components or activities or outputs contributing to those objectives, usually after words or phrases such as *by means of*, *through*, or *by*. In this case, the element whose achievement should be assessed is only the expected outcomes. Components, activities, and outputs should be factored into the results chain analysis—in other words, they may help demonstrate the causal (or, at least, plausibly causal) relationship between the project's interventions and achieved outcomes. If the phrase *in order to* is used, however, then what comes *after* that phrase usually constitutes the main objective(s) whose achievement is to be assessed.

If the PDO statement expresses a goal to support a government's program, the objectives of that program (normally found in the PAD) should be used. If the project supports a subset of the government's program objectives, the assessment of efficacy should include only the program objectives specifically supported by the project.

### Parsing the Project Development Objective Statement, Example 1

Here is the PDO statement of a rural poverty reduction project:

To improve access to small-scale socio-economic infrastructure and services, raising incomes through investment in productive activities, and strengthening the capacity of Municipal Councils and Community Associations to raise funding and harmonize policies and institutional arrangements for delivery of public investments intended to benefit the rural poor.

In this example, assess efficacy separately for each of the project's three objectives (or outcomes):

Objective 1: Improving access to small-scale socioeconomic infrastructure and services

Objective 2: Raising incomes through investments in productive activities

Objective 3: Strengthening the capacity of municipal councils and community associations to raise funding and harmonize policies and institutional arrangements for delivery of public investments intended to benefit rural poor people.

### Parsing the Project Development Objective Statement, Example 2

Here is the PDO statement of a health services project:

To support the borrower's efforts to further strengthen its health delivery services and the current health policy framework for NCDs [noncommunicable diseases] through

- The expansion of access and the quality of primary health care services related to NCD early detection; and
- The provision of specialized medical care to avoid or reduce exposure to NCD risk factors and their health effects

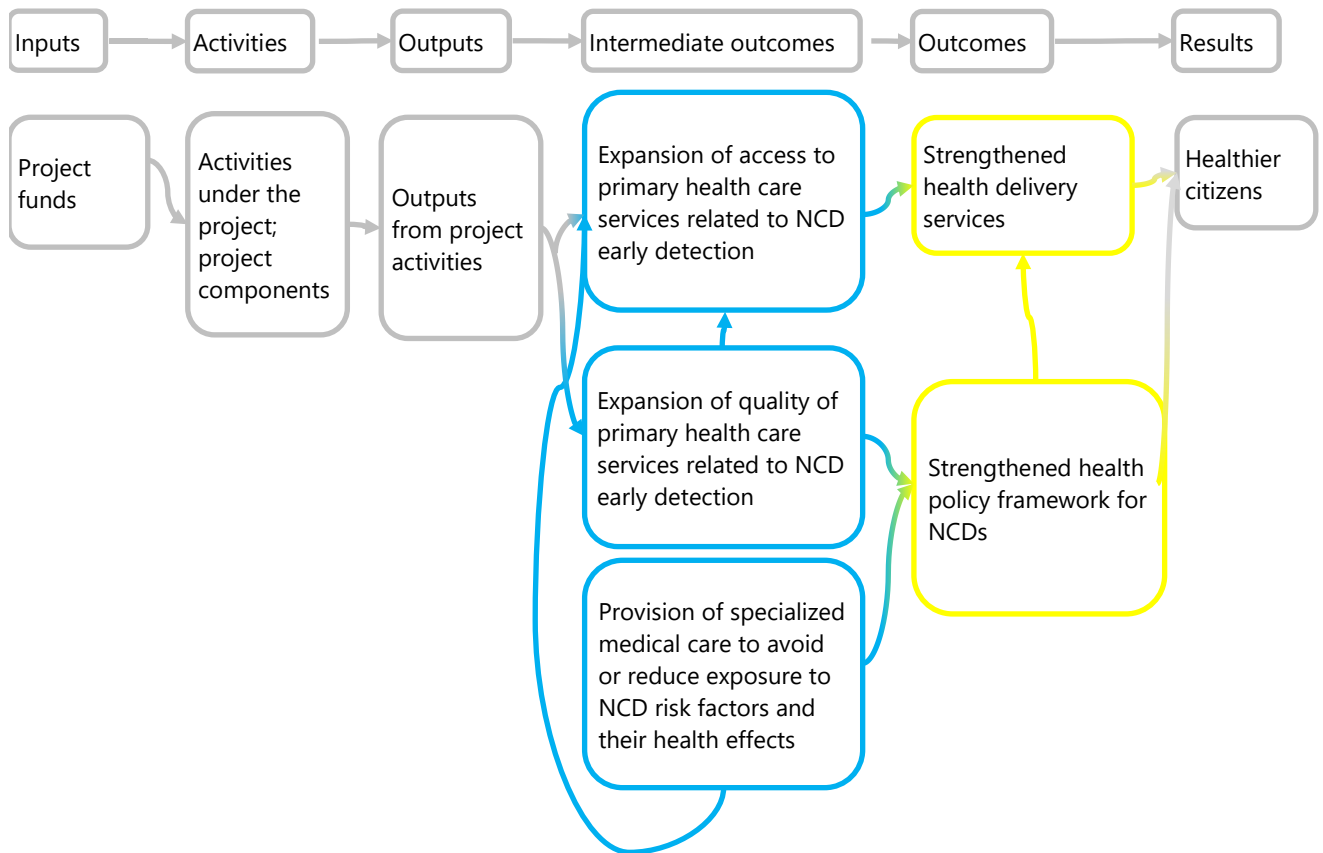
In this example, assess efficacy separately for two elements. These constitute the main objectives to rate in the efficacy section:

Objective 1: Strengthened health delivery services

Objective 2: Strengthened health policy framework for NCDs

The elements highlighted in blue above are considered intermediate outcomes. Figure 2.1 shows a sketch of the results chain suggested by the PDO statement.

**Figure 2.1. Results Chain**



Source: Independent Evaluation Group.  
Note: NCD = noncommunicable disease.

**Components, activities, and project outputs should not be included as objectives, even if they appear in the same sentence as the outcome-level objectives.** The PDO statement of the Kyrgyz Agricultural Support Services Project incorporates both multiple outcomes and a listing of components in a single objective statement. The statement of objectives was as follows:

To improve the incentive framework for, and productivity, profitability, and sustainability of Kyrgyz agriculture by means of: assisting the government in implementing land and agrarian reforms; providing emerging private farms with advisory and development services; developing the seed industry; establishing a legal framework, organizations, and procedures for crop protection and plant quarantine; establishing an agricultural market information system; and

enhancing institutional capacity of the Ministry of Agriculture and Water Resources.

Up to the phrase *by means of*, there are four outcomes that the project seeks to effect, listed below. These four are, in effect, the main outcomes that the project sought to achieve and would be the main headings in section 4 of the ICRR form, on achievement of objectives (see below). All of the activities *after* the words *by means of* are the project's components, which are activities and outputs. Thus, they should not be considered as objectives but instead as outputs within the project's results chain leading to the three main outcomes.

- Objective 1: Improve incentive framework for Kyrgyz agriculture
- Objective 2: Improve productivity of Kyrgyz agriculture
- Objective 3: Improve profitability of Kyrgyz agriculture
- Objective 4: Improve sustainability of Kyrgyz agriculture

## Theory of Change

**For each objective or outcome, succinctly describe the TOC** or the results chain or logic behind the objective. This could be derived from the TOC presented in the ICR under the section "Context at Appraisal." The TOC illustrates the results chain, explaining the links between the operation's interventions, outputs, intermediate results, and desired outcomes, along with underlying assumptions. The TOC discussion is structured in a way that the objective statement is explicit at the level of the desired outcomes. The teams are also advised to include longer-term outcomes expected to occur beyond the project's closing to demonstrate the project's contribution to the borrower's higher-level objectives for the sector or the relevant CPF objective(s). The TOC discussion also identifies any critical assumptions and external factors that might affect or contribute to the outcomes.

**Provide a brief assessment of the TOC.** Evaluating a project TOC involves assessing the plausibility and coherence of the logic and assumptions, and the potential effectiveness of the project design. Some guiding questions to reflect on when assessing the project's TOC are the following:

- **Assess the coherence of the logic.** Are the links between project activities and outcomes plausible and logical? Are there any missing links that need to be made more explicit? Is there alignment between the outcomes specified in the PDO and the high-level outcomes or CPF objectives to which the project contributes?

- **Assess the plausibility of assumptions.** Were the critical assumptions and external factors that may have affected or contributed to the outcomes clear? Are the assumptions realistic given the project context?
- **Consider alternative explanations.** Are there other possible approaches or factors that might contribute to achieving the outcomes? Discuss the other factors that could have affected the outcome or what has been left out of the TOC.
- **Link with the results framework.** Was the TOC used as a basis for developing the operation's results framework?
- **Balance between the complexity and clarity of causal links.** Did the TOC strike a good balance between the complexity and clarity of the causal links when presenting the operation design?

Note that the TOC can be analyzed for each objective separately under their respective sections in the efficacy category. Alternatively, the discussion can be combined for the PDO and presented up-front in the efficacy section.

### Discussion of Validity of Indicators

As part of the discussion of efficacy, the ICR should contain an assessment of the validity of the PDO-level indicators in the results framework, and the ICRR should comment on the validity of the PDO-level indicators used to show the extent of achievement of objectives. If the indicators defined in the results framework had shortcomings (for example, in validity, measurability, timeliness, and so on), check whether the ICR included other sources of information that would speak to the achievement of project objectives or outcomes.

Even in cases where indicators defined in the results framework were excellent for assessing the outcomes, the ICR may include additional data and evidence that speaks to achievements. Use of multiple sources of information helps with “triangulation” – when different sources point to the same types of achievement, this convergence provides a richer, more accurate assessment of achievement of objectives, and, if it is well done, the ICRR may highlight this as a commendable practice. An important element of complementary data and evidence is the perspective of beneficiaries.

### Discussion of Attribution and the Counterfactual

The **counterfactual** is defined as what would have happened in the absence of the government intervention, project, or program supported by the World Bank. Establishing the evidence for the elements of the results chain for each outcome is a necessary but not sufficient condition for attributing the outcomes to the project. In most

cases, other factors beyond the scope of the project also affect these outcomes, contributing to or detracting from them. These factors might include the influence of weather or rainfall, economic crises, natural disasters, favorable or unfavorable international prices for farmers' production, other government policies outside the project, or the activities of other donors.

Programs supported by the World Bank are rarely subjected to an evaluation design, such as a randomized experiment, capable of contrasting results with and without an intervention or program. In some cases, rigorous impact evaluations can be conducted on specific parts or interventions within a program; the results from these studies can be useful in understanding the counterfactual, at least in determining what parts of a program worked.

However, most projects or programs supported by the World Bank involve large-scale and multifaceted interventions or country- or sectorwide policies for which it would be difficult or impossible to establish an airtight counterfactual as the basis for attributing outcomes to the project.

To assess efficacy, for each objective, the IEG ICR reviewer should nevertheless identify and discuss the key factors outside the project that might have contributed to or detracted from the outcomes, and any evidence (from the ICR) for the actual influence of these factors.<sup>3</sup>

The following types of information—in addition to evidence from the results chain—have been found useful in assessing evidence on the extent to which the outcomes achieved can plausibly be attributed to the project or program:

- A timeline of key events, showing the relationship among project activities, events beyond the project, and changes in outcomes flagged by the objectives
- Evidence of trends in the outcomes before, during, and after the project or program
- Evidence of trends in outcomes in project and nonproject areas, taking into account the ways the two areas may differ in baseline characteristics and other factors that may be affecting the two areas
- Trends in other factors that plausibly could have influenced the outcomes independently of the project, such as weather, natural disasters, economic trends, other government policies, or the activities of other donors

For all ratings, the IEG ICR reviewer needs to assess whether the outcomes achieved are attributable to the project. When the expected outcome is achieved but the evidence is



not presented to show how the activities and outputs of the project have led to the outcomes, then the rating of efficacy should be lower than would be the case with the same level of expected outcome accompanied by evidence of how the project activities and outputs led to the outcomes. Similarly, when evidence indicates that expected outcomes achieved occurred because of factors other than project activities and outputs, the rating of efficacy should be lower than would be the case with similar outcomes accompanied by evidence of how the project activities and outputs led to the outcomes.

It is important to note that the efficacy rating reflects the incremental contribution of the project or program to observed outcomes, regardless of whether the observed outcomes moved in the “right” or “wrong” direction. For example:

- If the expected outcome target was met or exceeded, but there is evidence that the change was due mainly (or solely) to external factors, an efficacy rating of modest (or negligible) may be warranted.
- If the outcome deteriorated, falling short of the target, but there is evidence that the decline would have been even worse in the absence of the project, a rating of substantial (or high) could be warranted.

To justify these judgments, a high standard of evidence is expected. For example, it is insufficient for the ICR to claim that the project fell short of achieving its objective because of macroeconomic conditions without strong evidence that these conditions were responsible for the trend in the outcome indicator. The burden of proof is on the ICR itself to show that improved outcomes were the result of the project, and that declining outcomes were not.

### **How to Treat Overarching Objectives and Objectives across Projects**

When there are both overarching and specific objectives, the ICR should report both. This is very likely to be the case for a series of projects, which constitute a series of investment projects (phases) offered over a medium- to long-term period, with objectives for each phase and for the overall program. The ICR and the ICRR should assess achievement of the overarching objectives of the overall program and the achievement of the specific objectives of each completed phase and their contribution to the overall program outcome. Regarding the overarching objective, the ICRR should also comment on the likelihood that the overarching objective will be achieved in the future and should cite the reasoning and information on which this assessment is based. The overarching development objective should not be rated, however. (In the electronic ICRR form, open an objective section for the overarching objective, and use the rationale field to comment on it. Then, in the rating field, select “Not Rated/Not Applicable.”)

## Other Considerations in the Assessment of Efficacy

If objectives are not yet achieved at the time of project closing, the ICR or the ICR review, or both can still make a case that they are “likely to be achieved,” but the IEG ICR reviewer would need to look for, and present, convincing evidence of the likelihood of such achievement through a strong results chain. If sufficient evidence does not exist, no leaps of faith should be made.

A downgrade in the efficacy rating is warranted in both of the following cases: (i) when there is insufficient evidence of impact; or (ii) there is evidence of insufficient impact.

## Rating of Efficacy

In the ICRR, the efficacy of each objective (intended outcome) is rated on a four-point scale: **high, substantial, modest, negligible**. The ICR should include a discussion of efficacy of each objective but is not required to contain a rating of efficacy for each objective. Both in the ICRR and the ICR, an overall efficacy rating is given. The ratings are defined as follows:

- **High.** The project fully achieved or exceeded its objectives (or intended outcomes) or is likely to do so.
- **Substantial.** The project almost fully achieved its objective (or intended outcomes) or is likely to do so.
- **Modest.** The project partly achieved its objectives (or intended outcomes) or is likely to do so.
- **Negligible.** The project barely achieved its objectives (or intended outcomes) or is likely to do so.

## Arriving at the Overall Efficacy Rating in Projects with Multiple Objectives

In projects with multiple objectives, IEG’s ICRR rates the achievement or likely achievement of each individual objective, and also provides a single overall efficacy rating covering all objectives. The World Bank’s ICR should include a discussion of achievement of each individual objective but will provide only a single overall efficacy rating covering all the objectives. Both the World Bank’s ICR and IEG’s ICRR use the same approach for arriving at the overall efficacy rating, and use a harmonized rating scale: high, substantial, modest, and negligible.

If neither the legal agreement nor the PAD explicitly indicates the relative importance or weight of objectives, then the World Bank’s ICR and IEG’s ICRR assume equal importance (or equal weight) for each of the objectives.

In any case, the amount of resources allocated to an objective should not be used to impute the relative importance (or weight) of that objective.

As a rule of thumb, a high or substantial efficacy rating is warranted when all three of the following are true:

- There is sufficient evidence of outcomes or impact;
- There is evidence of sufficient outcomes or impact; and
- The observed outcomes or impact can be attributed to the project interventions or activities.

If most of the project objectives are rated high or substantial, the overall efficacy rating will tend to be high or substantial. If a majority of the project objectives are rated modest or negligible, the overall efficacy rating will tend to be modest or negligible.

If half the objectives are rated high or substantial and half modest or negligible, then there is the possibility to either round up or round down based on the strength of the evidence. Rounding up would be justified in the case of “strong” high or substantial ratings (that is, they are at the top of their range, so that, for example, the substantial rating is nearly high). Rounding down would be justified in the case of “weak” high or substantial ratings (that is, they are at the bottom of their range, so that, for example, the substantial rating is barely so).

The overall efficacy rating derived based on the guidance above should then be cross-checked against best judgment, in other words, stepping back and asking the questions: “To what extent did the project achieve the objectives promised?” and “Are the shortcomings in achievement of the objective absent, minor, moderate, significant, severe, or major?”

Both of the above approaches should be used to guide and arrive at the final overall efficacy rating.

## **Section 5: Efficiency**

### **Definition**

Efficiency is a measure of how economically resources and inputs are converted to results. For a development project, the central question is whether the costs involved in achieving project objectives were reasonable in comparison with both the benefits and with recognized norms (“value for money”; World Bank, 2021, Appendix G). To what

extent did the project achieve the maximum possible benefits (outputs, outcomes, and impacts) with the minimum possible inputs or costs?

## **Guidelines**

This section should report on all available measures of efficiency both ex ante and ex post and highlight any data gaps and methodological weaknesses in the World Bank's assessment of efficiency. Criteria for the assessment of efficiency can vary across the range of sectors that an investment operation might support. The analysis should discuss both the traditional measures of efficiency (as applicable and practical)—for example, net present value, economic rate of return (ERR), cost-effectiveness, unit rate norms, service standards, least-cost analysis and comparisons, and financial rate of return—and aspects of design and implementation that either contributed to or reduced efficiency. The ICR should also analyze the project's efficiency using any appropriate cost-effectiveness criteria to determine whether the project represented the expected least-cost solution to attain identified and measurable benefits by an analysis of either cost per unit of input or cost per unit of output.

## **Economic Analysis**

### **At Appraisal**

The methodology used at appraisal should clearly be stated and discussed, including (i) benefits; (ii) costs; and (iii) assumptions.

If an ERR or internal rate of return (IRR) has been calculated, the assumptions should be fully explained and transparent in the ICR. Any data or other gaps and methodological strengths or weaknesses in the World Bank's assessment of efficiency should be noted.

The ICR normally would indicate what the ERR or IRR or net present value was at project appraisal. Underlying assumptions about costs and benefits, and any other information supporting the analysis (for example, output volumes, major cost items, or prices) including a sensitivity analysis should be presented.

### **At Closing**

Normally, if an ERR or IRR or net present value (or similar) was estimated at project appraisal in the PAD, the ICR should repeat the calculation based on information available at the time of closing, presenting what it is when reestimated at completion, and on what percentage of total project costs the original and revised estimates were based.

The ICR should also indicate the components and the percentage of total project costs covered by any such analyses (noting any differences from the analyses at appraisal).

If an ERR is calculated at project appraisal (ex ante), state the methodology and explicitly discuss the benefits, costs, and assumptions. Subsequently report ex post ERR. Clearly indicate if the ex ante and ex post methodologies are the same or comparable. The review should discuss the reasons for differences in the ex post and ex ante ERR.

In the efficiency fields in the electronic ICRR form, “Coverage/scope (%)” refers to the percentage of total project cost for which the ERR, IRR, or financial rate of return was calculated. A comment should be included, where possible, on the reliability of the ERR, IRR, or financial rate of return calculation(s) presented in the ICR.

### **Implementation and Administrative Efficiency**

The ICR reviewer should discuss various aspects of operational and administrative efficiency that contributed to the efficient use of resources under the project, and include all available indicators of efficiency, including efficient use of project funds, in the assessment. Shortcomings in efficiency may have to do with the extent to which the project fails to achieve (or is not expected to achieve) a return higher than the opportunity cost of capital and is not the least-cost alternative.

Implementation efficiency may be affected by

- The complexity of the project or program and its organizational arrangements;
- The commitment demonstrated by government, its agencies, and other participants to the objectives of the project;
- Whether risks were identified and their mitigation adequate;
- The adequacy of participatory processes; and
- Unforeseen security and natural events.

### **When a Project Objective Is to Improve the Efficiency of a Sector, Avoid Confusing Efficiency of the Project with Efficiency of a Sector**

In some projects, one of the objectives is to improve the efficiency of a sector (for example, the health sector or the agricultural sector) within the country, or to improve the efficiency of a government program being supported. In such cases, achievement of improved efficiency of the sector or the government program represents achievement of a project objective, and therefore should be assessed under efficacy, not under efficiency.

The rating of efficiency in the ICR is intended to cover the extent to which project resources were used wisely and achieved good value for the money.

Especially in such cases, it is important to avoid confusing the efficiency of the project with achievement of improved efficiency of the sector or program being supported. The latter is an outcome and would be included in the assessment of efficacy. For example, in an education system, repetition and dropout rates might decline as the result of an education investment, which would suggest improved internal efficiency of the education system. It would not necessarily indicate that project resources were used efficiently (that is, that the project was implemented cost effectively or at least cost). Likewise, efficiency is about the cost-effectiveness of project resources, not the use of World Bank budgetary resources.

### **Rating of Efficiency**

Efficiency should be assigned an overall rating, based on a four-point scale: negligible, modest, substantial, or high.

High	Efficiency exceeds expectations.
Substantial	Efficiency is what would be expected in the operation's sector.
Modest	Efficiency is below expectations in the operation's sector.
Negligible	Efficiency is very low in comparison with both the benefits (if any) and with recognized norms in the operation's sector.

## **Section 6: Project Outcome**

### **Definition**

The project outcome is defined as “the extent to which the project’s major relevant objectives were achieved, or are expected to be achieved, efficiently.”

Thus, the outcome rating is derived from the prior assessment of the relevance of objectives, efficacy in achieving each objective, and efficiency.

To ensure consistency across IEG ICR reviewers, IEG has developed guidelines for deriving the project outcome rating from the subratings on relevance, efficacy, and efficiency in the previous sections.

## Guidance

The IEG-OPCS Harmonized Evaluation Criteria (World Bank, 2021, page 10) provide the following guidance in assigning an outcome rating, based on relevance, efficacy, and efficiency:

The World Bank's project evaluation architecture – both self-evaluation (reflected in the ICR and in ISRs) and independent evaluation (IEG's assessments) – is objectives-based. In keeping with this objectives-based methodology, the ICR assesses and rates the outcomes of operations at closing against the objectives and outcome targets defined at the outset and any new or revised objectives or outcome targets introduced through restructuring.

### Rating of Outcome

Highly satisfactory	There were no shortcomings in the project's achievement of its objectives, in its efficiency, or in its relevance.
Satisfactory	There were minor shortcomings in the project's achievement of its objectives, in its efficiency, or in its relevance.
Moderately satisfactory	There were moderate shortcomings in the project's achievement of its objectives, in its efficiency, or in its relevance.
Moderately unsatisfactory	There were significant shortcomings in the project's achievement of its objectives, in its efficiency, or in its relevance.
Unsatisfactory	There were major shortcomings in the project's achievement of its objectives, in its efficiency, or in its relevance.
Highly unsatisfactory	There were severe shortcomings in the project's achievement of its objectives, in its efficiency, or in its relevance.

### Deriving the Outcome Rating from Independent Evaluation Group's Subratings of Relevance, Efficacy, and Efficiency

The outcome rating is derived from the assessment of the relevance of objectives, efficacy in achieving each objective, and efficiency. Because the World Bank and IEG

have adopted an objectives-based evaluation methodology, achievements are assessed against the PDOs.

For consistency, the following rules of thumb will derive the project outcome rating from the subratings for relevance, efficacy, and efficiency (table 2.1):

- To receive an outcome rating of moderately satisfactory or higher, a project must be rated high or substantial on efficacy.
- A rating of modest on efficacy produces an outcome rating that is, at best, moderately unsatisfactory.
- A rating of negligible on any one of relevance, efficacy, and efficiency produces an outcome rating that is at best unsatisfactory.
- In order for the outcome to be rated highly satisfactory, a project must be rated high on efficacy, high on any one of the other subratings, and substantial on the third.
- Tables 2.1 and 2.2 provide guidance on other possible scenarios.

**Table 2.1. Deriving the Overall Outcome Rating for a Project, Tree View**

Subratings			Outcome Rating
Relevance	Efficacy	Efficiency	
High	High	High or substantial	Highly satisfactory
		Modest	Moderately satisfactory
		Negligible	Unsatisfactory
	Substantial	High or substantial	Satisfactory or moderately satisfactory <sup>a</sup>
		Modest	Moderately satisfactory
		Negligible	Unsatisfactory
	Modest	High, substantial, or modest	Moderately unsatisfactory
		Negligible	Unsatisfactory
		High, substantial, or modest	Unsatisfactory
Substantial	High	Negligible	Highly unsatisfactory
		High	Highly satisfactory
		Substantial	Satisfactory
	Substantial	Modest	Moderately satisfactory
		Negligible	Unsatisfactory
		High or substantial	Satisfactory or moderately satisfactory <sup>a</sup>
	Modest	Modest	Moderately satisfactory
		Negligible	Unsatisfactory
		High, substantial, or modest	Moderately unsatisfactory
		Negligible	Unsatisfactory



Chapter 2  
Guidance Manual

Relevance	Subratings		Outcome Rating	
	Efficacy	Efficiency		
Modest	Negligible	High, substantial, or modest	Unsatisfactory	
		Negligible	Highly unsatisfactory	
	High	High or substantial	Moderately satisfactory	
		Modest	Moderately unsatisfactory	
		Negligible	Unsatisfactory	
		Substantial	Moderately satisfactory	
	Substantial	High or substantial	Moderately satisfactory	
		Modest	Moderately unsatisfactory	
		Negligible	Unsatisfactory	
		Modest	Moderately unsatisfactory	
	Negligible	Negligible	High or substantial	Moderately unsatisfactory
			Modest	Unsatisfactory
High		High, substantial, or modest	Unsatisfactory	
		Negligible	Highly unsatisfactory	
		Substantial	Moderately satisfactory	
		Modest	Moderately unsatisfactory	
Substantial		High, substantial, or modest	Unsatisfactory	
		Negligible	Highly unsatisfactory	
		Modest	Unsatisfactory	
		Negligible	Highly unsatisfactory	
Negligible		High, substantial, modest, or negligible	Highly unsatisfactory	

Source: Independent Evaluation Group and Operations Policy and Country Services.

a. For a project in which there is a modest achievement of one or more of the objectives or outcomes used in the assessment of the overall efficacy.

**Table 2.2. Deriving the Overall Outcome Rating for a Project, Table View**

Rating	Definition	Subratings: Relevance, Efficacy, and Efficiency	Comment
Highly satisfactory	There were no shortcomings in the operation's achievement of its objectives, in its efficiency, or in its relevance.	High on any two criteria—one of which must be efficacy—and at least substantial on the third	<i>No shortcomings</i> requires efficacy to be one of the high ratings.
Satisfactory	There were minor shortcomings in the operation's achievement of its objectives, in its efficiency, or in its relevance.	<ul style="list-style-type: none"> <li>Substantial on all three criteria</li> <li>Substantial on two criteria and high on the third, or</li> <li>Substantial efficacy but high relevance and efficiency</li> </ul>	<i>Minor shortcomings</i> is implicitly defined as substantially achieving the objectives and substantial or better on the other two criteria.
Moderately satisfactory	There were moderate shortcomings in the operation's achievement of its objectives, in its efficiency, or in its relevance.	Substantial (or high) on two criteria—one of which must be efficacy—and modest on the third	<i>Moderate shortcomings</i> is implicitly defined as modest on one criterion.
Moderately unsatisfactory	There were significant shortcomings in the operation's achievement of its objectives, in its efficiency, or in its relevance.	<ul style="list-style-type: none"> <li>Modest on any two criteria and substantial (or high) on the third, or</li> <li>Modest efficacy with substantial (or high) on the other two criteria</li> </ul>	<i>Significant shortcomings</i> is implicitly defined as modest on two criteria or modest efficacy; would also apply if one were high and two were modest.
Unsatisfactory	There were major shortcomings in the operation's achievement of its objectives, in its efficiency, or in its relevance.	<ul style="list-style-type: none"> <li>Modest on all three criteria, or</li> <li>Negligible on one criterion and modest, substantial, or high on the other two</li> </ul>	<i>Major shortcomings</i> is implicitly defined as modest on three criteria or at least one is negligible.
Highly unsatisfactory	There were severe shortcomings in the operation's achievement of its objectives, in its efficiency, or in its relevance.	<ul style="list-style-type: none"> <li>Negligible on all three criteria, or</li> <li>Negligible on two criteria and modest, substantial, or high on the third</li> </ul>	<i>Severe shortcomings</i> is implicitly defined as negligible on at least two criteria.

Source: Independent Evaluation Group and Operations Policy and Country Services.

In addition, the Harmonized Evaluation Criteria specifically mention that the variation in achievement of different objectives is to be taken into account:

Shortcomings in the achievement of objectives may have to do with either the number of objectives that are not achieved (or are not expected to be achieved) and/or the extent to which one or more objectives are not achieved (or are not expected to be achieved).

When insufficient information is provided by the World Bank for IEG to arrive at a clear rating, this can be a reason for an unfavorable rating.

The rating of outcome should encompass the extent to which the project's objectives were relevant and were achieved, or are expected to be achieved, efficiently.

- Shortcomings in the achievement of objectives may have to do with either the number of objectives that are not achieved (or are not expected to be achieved), the extent to which one or more objectives are not achieved (or are not expected to be achieved), or both.
- Shortcomings in efficiency may have to do with the extent to which the project fails to achieve (or is not expected to achieve) a return higher than the opportunity cost of capital and is not the least-cost alternative (this criterion may not apply for development policy loan operations).
- Shortcomings in relevance may have to do with the extent to which a project's objectives are inconsistent with current World Bank country strategies (as expressed in Poverty Reduction Strategy Papers, and country assistance strategies, Systematic Country Diagnostics, or CPFs). It is important to ensure that achievement of objectives reflects continuing priorities at the PDO level, not out-of-date priorities that should have triggered restructuring.

The IEG ICR reviewer must use judgment in weighing possible shortcomings in the achievement of the project's objectives, in its efficiency, or in its relevance, and arrive at an assessment of how they affect the overall rating.

### **Rating the Outcome of Projects with Formally Revised Objectives**

See guidance on split rating in appendix section C1 of the "Reference Annex: Illustrative Examples for Independent Evaluation Group Validators."

#### **Procedure for Applying Split Rating of Outcome**

- Determine the actual total World Bank disbursements before and after the date when the revised project objectives were formally approved.

- Provide two outcome ratings—one for achievements against the original objectives or outcome targets, and one for achievements against the revised objectives or outcome targets.
  - Outcome ratings are determined on the basis of individual ratings for relevance, efficacy, and efficiency.
  - Separate efficacy ratings are derived based on the objectives or targets before the restructuring and for the objectives or targets after the restructuring. It is important to note that, when objectives are revised, the project is rated against both sets of objectives separately, for the *entire duration of the project* (not just the period for which each of the objectives was in effect). Achievement of each individual objective (efficacy), both original and revised, is assessed across the project's entire lifetime.
  - Relevance and efficiency are given only one rating each (relevance at closing and efficiency at closing).
  - The first outcome rating is derived using the first efficacy rating and the ratings for relevance and efficiency at closing. The second outcome rating is derived using the second efficacy rating and the ratings for relevance and efficiency at closing.
- Assign a numeric value for each of the outcome ratings: highly satisfactory = 6, satisfactory = 5, moderately satisfactory = 4, moderately unsatisfactory = 3, unsatisfactory = 2, highly unsatisfactory = 1.
- Arrive at an overall rating by weighting the two ratings by the proportion of actual total disbursement under each set of objectives and rounding to the nearest whole number (1 to 6).
- If a project's objectives, outcome targets, or both are revised more than once, additional outcome ratings are determined for each restructuring as necessary, and the overall outcome rating is determined according to the percentage of loan, credit, or grant disbursements under each restructuring.

## Example of the Calculation of Split Ratings for a Skills Development Project

**Table 2.3. Overall Outcome Ratings**

Rating Dimension	Original Objectives	Objectives after First Revision	Objectives after Second Revision
Relevance of Objectives		Substantial	
Efficacy			
Objective 1: Improve access to vocational training	Modest	Substantial	High
Objective 2: Improve quality of vocational training	Modest	Substantial	Substantial
Objective 3: Improve the demand-responsiveness of vocational training	Modest	Modest	High
Overall efficacy	Modest	Substantial	High
Efficiency		Modest	
Outcome rating	Moderately unsatisfactory	Moderately satisfactory	Moderately satisfactory
Outcome rating value	3	4	4
Amount disbursed (US\$, millions)	2.9	20.5	5.8
Disbursement (%)	9.9	70.2	19.9
Weight value	0.30	2.81	0.79
Total weights		3.9 (rounds up to 4.0)	
Overall outcome rating		Moderately satisfactory (4.0)	

*Source:* adapted from ICRR of P118101, Rwanda Skills Development Project.

See appendix section C1 of the “Reference Annex: Illustrative Examples for Independent Evaluation Group Validators” for more examples.

## Section 7: Risk to Development Outcome

### Definition

The risk to development outcome is the risk, at the time of evaluation, that development outcomes (or expected outcomes) will not be maintained (or realized). This refers to outcomes that have actually been achieved (or are expected to be achieved).

### Guidance

The risk to development outcome has two dimensions:

- The likelihood that changes may take place that are detrimental to the ultimate achievement of the project’s development outcome

- The impact of some or all of these changes on the project's development outcomes

Some risks are internal or specific to a project. They are primarily related to the suitability of the project's design to its operating environment.

Other risks arise from factors outside the project. These may appear at the country level, such as price changes, or on a global scale, such as technological advances. The impact on outcomes of a change in the operating environment depends on the severity and nature of the change and the adaptability (or lack thereof) of the project's design to withstand that change.

Assessment of risk to development outcome requires an assessment of the uncertainties faced by a project over its expected useful life and of whether adequate arrangements are in place to help avoid or mitigate the impact of those uncertainties. This risk is understood to be higher if the design or implementation of the project is not well aligned with the operating environment or mitigation measures are inappropriate to deal with foreseeable risks.

## Criteria

The overall risk to development outcome is based on both the probability and the 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 its initial implementation. The IEG ICR reviewer should consider the operational, sector, and country context in weighing (in each case) the relative importance of these individual criteria of risk as they may affect planned outcomes:

- Technical (where innovative technology and systems are involved)
- Financial (including the robustness of financial flows and financial viability)
- Economic (at both the country and global levels)
- Social (in terms of the strength of stakeholder support, mitigation of any negative social impacts, or both)
- Political (for example, volatility of the political situation)
- Environmental (including both positive and negative impacts)
- Government ownership or commitment (for example, continuation of supportive policies and any budgetary provisions)

- Other stakeholder ownership (for example, from private sector or civil society)
- Institutional support (from project entities; related to legal or legislative framework)
- Governance
- Exposure to natural disasters

## Section 8: Bank Performance

### Definition

The World Bank’s performance is defined as the extent to which services provided by the World Bank ensured quality at entry of the project and supported effective implementation through appropriate supervision (including ensuring adequate transition arrangements for regular operation of supported activities after loan or credit closing), toward the achievement of development outcomes.

Bank performance is rated by assessing two dimensions: (i) Bank performance in ensuring quality at entry and (ii) quality of World Bank supervision. Based on the criteria discussed below,<sup>1</sup> the IEG ICR reviewer rates the World Bank’s quality at entry and quality of supervision separately and uses the IEG-OPCS Harmonized Evaluation Criteria Guidelines to arrive at an overall rating for Bank performance.

### Quality at Entry

#### Definition

Quality at entry, which is shorthand for “Bank performance in ensuring quality at entry” refers to the extent to which the World Bank identified, facilitated preparation of, and appraised the project so that it was most likely to achieve planned development outcomes and was consistent with the World Bank’s fiduciary role.

#### Criteria

Bank performance in ensuring quality at entry is rated against the following criteria, as applicable to a particular project. The IEG ICR reviewer should take account of the operational, sector, and country context in weighing the relative importance of each criterion of quality at entry as it affected outcomes:

- Strategic relevance and approach
- Technical, financial, and economic aspects (for investment lending projects)

- Poverty, gender, and social development aspects
- Environmental aspects<sup>2</sup>
- Fiduciary aspects
- Policy and institutional aspects
- Implementation arrangements
- M&E arrangements
- Risk assessment
- World Bank inputs and processes

Information on which to base the assessment of quality at entry may be found throughout the ICR, but sections of particular importance are the description of components and the section on background and context, “Key Factors That Affected Implementation and Outcome.”

### Rating of Bank Performance in Ensuring Quality at Entry

With respect to the relevant criteria that would enhance development outcomes and the World Bank’s fiduciary role, rate Bank performance in ensuring quality at entry using the following scale:

Highly satisfactory	There were no shortcomings in identification, preparation, or appraisal.
Satisfactory	There were minor shortcomings in identification, preparation, or appraisal.
Moderately satisfactory	There were moderate shortcomings in identification, preparation, or appraisal.
Moderately unsatisfactory	There were significant shortcomings in identification, preparation, or appraisal.
Unsatisfactory	There were major shortcomings in identification, preparation, or appraisal.



Highly unsatisfactory      There were severe shortcomings in identification, preparation, or appraisal.

## Quality of Supervision

### Definition

Quality of supervision refers to the extent to which the World Bank proactively identified and resolved threats to the achievement of relevant development outcomes and the World Bank's fiduciary role.

### Criteria

Bank performance in quality of supervision is rated against the following criteria, as applicable to a particular project. The IEG ICR reviewer should take account of the operational, sector, and country context in weighing the relative importance of each criterion of quality of supervision as it affected outcomes:

- Focus on development impact
- Supervision of fiduciary and safeguard aspects (when applicable)
- Adequacy of supervision inputs and processes
- Candor and quality of performance reporting
- Role in ensuring adequate transition arrangements (for regular operation of supported activities after loan or credit closing)

In assessing Bank performance, it can be helpful to keep in mind three groups of implementation factors that remain outside the control of the World Bank:

- Factors **outside the control of the World Bank or government or implementing agencies**, such as the following:
  - Changes in world markets and prices
  - Unexpected and unforeseeable technical difficulties
  - Natural disasters (including extraordinary weather and sudden disease epidemics)
  - War and civil disturbances, including the effects overflowing from neighboring territories (such as refugees)

- Factors generally **subject to government control**, such as the following:
  - Macroeconomic and sector policies
  - Government commitment
  - Governance and corruption (see Handbook on Governance and Corruption)
  - Appointment of key staff
  - Provision of counterpart funds
  - Efficient administrative procedures
- Factors general **subject to implementing agency control**, such as the following:
  - Management effectiveness
  - Staffing adequacy and quality

### Rating of Quality of Supervision

In consideration of relevant criteria that would enhance development outcomes and the World Bank's fiduciary role, rate quality of supervision using the following scale:

Highly satisfactory	There were no shortcomings in the proactive identification of opportunities and resolution of threats.
Satisfactory	There were minor shortcomings in the proactive identification of opportunities and resolution of threats.
Moderately satisfactory	There were moderate shortcomings in the proactive identification of opportunities and resolution of threats.
Moderately unsatisfactory	There were significant shortcomings in the proactive identification of opportunities and resolution of threats.
Unsatisfactory	There were major shortcomings in the proactive identification of opportunities and resolution of threats.
Highly unsatisfactory	There were severe shortcomings in the proactive identification of opportunities and resolution of threats.

## Rating of Overall Bank Performance

The rating of overall Bank performance is based on the ratings for each of the two dimensions (i) Bank performance in ensuring quality at entry and (ii) the quality of supervision. The quality at entry and quality of supervision ratings should be combined into a rating of overall Bank performance.

In general, the lower of the two ratings (of Bank performance in ensuring quality at entry and quality of supervision) determines the rating of overall Bank performance. The reason for this is that Bank performance is considered to be entirely within the control of the World Bank. A mistake or shortcoming in either one would have been avoidable.

Ratings for the most common combinations of ratings of quality at entry and quality of supervision are provided below, followed by additional guidance on other combinations (table 2.4).

Highly satisfactory	Bank performance was rated highly satisfactory on both dimensions.
Satisfactory	Bank performance was rated satisfactory on both dimensions, or was rated satisfactory on one dimension and highly satisfactory on the other dimension.
Moderately satisfactory	Bank performance was rated moderately satisfactory on both dimensions, or was rated moderately satisfactory on one dimension and satisfactory or highly satisfactory on the other dimension. (Also see guidance below.)
Moderately unsatisfactory	Bank performance was rated moderately unsatisfactory on both dimensions. (Also see guidance below.)
Unsatisfactory	Bank performance was rated unsatisfactory on both dimensions, or was rated unsatisfactory on one dimension and moderately unsatisfactory on the other dimension.
Highly unsatisfactory	Bank performance was rated highly unsatisfactory on both dimensions, or was rated moderately unsatisfactory or unsatisfactory on one dimension and highly unsatisfactory on the other dimension.

**When the rating for one dimension is in the satisfactory range (moderately satisfactory or better), and the rating for the other dimension is in the unsatisfactory range, the rating for overall Bank performance normally depends on the outcome rating.** Thus, overall Bank performance is rated moderately satisfactory *if* outcome is rated in the satisfactory range, or moderately unsatisfactory *if* outcome is rated in the unsatisfactory range.

**Table 2.4. Relation among Quality at Entry Quality of Supervision, and Overall Bank Performance Ratings**

Quality at Entry	Quality of Supervision	Overall Bank Performance
HS	HS	HS
S	S or HS	S
S or HS	S	S
MS	MS, S, or HS	MS
MS, S, or HS	MS	MS
MU	MU	MU
U	U or MU	U
U or MU	U	U
HU	HU, MU, or U	HU
HU, MU, or U	HU	HU
HS, S, or MS	MU, U, or HU	MS if outcome rating is MS or higher; MU if outcome is MU or lower
MU, U, or HU	HS, S, or MS	MS if outcome rating is MS or higher; MU if outcome is MU or lower

Source: Independent Evaluation Group.

Note: HS = highly satisfactory; HU = highly unsatisfactory; MS = moderately satisfactory; MU = moderately unsatisfactory; S = satisfactory; U = unsatisfactory.

## Section 9: Quality of Monitoring and Evaluation

### Definition

The M&E quality rating is based on an assessment of three main elements:

- M&E design
- M&E implementation
- M&E utilization

Monitoring and evaluation are distinct, and the rating is informed by both the quality of monitoring and the quality of evaluation.

## Guidance

In assessing the M&E quality rating, the IEG ICR reviewer should note that there may be good M&E mechanisms located outside the project as well as inside—for example, national surveys related to child educational achievements. Such alternative arrangements, provided they exist and serve the purpose, are fully acceptable as the basis for assessing the quality of M&E rating, and they are often more sustainable than project-specific M&E systems. Moreover, although monitoring is an essential part of any project management system, impact studies relevant to a sector and a project, such as impacts on child health, may be more efficiently done through broader national assessments.

Rating M&E quality is not intended to call for a focus only on quantitative evidence. In addition, good M&E will always rely on sound qualitative evidence, on the triangulation of that evidence with quantitative findings, and on the link of the array of evidence with the postulated causality chain.

In rating M&E quality, the IEG ICR reviewer is asked to look at three sequential elements: (i) *M&E design*, as reflected in the project design and proposed methodologies mapped out in the documents up to the point of Board approval; (ii) *M&E implementation*, as reflected in the actual project M&E inputs and the methodologies applied over the period of project effectiveness; and, (iii) *M&E utilization*, as reflected in the changes made in the ongoing project or changes in subsequent interventions attributable to this work. These three elements are common to both investment and policy lending.

The IEG ICR reviewer is asked to discuss each of the three elements of M&E quality separately and to arrive at an overall *quality of M&E* rating on a four-point scale.<sup>3</sup>

## Criteria

### Monitoring and Evaluation Design

The IEG ICR reviewer should assess to what extent the M&E design was sound and was designed to collect, analyze, and provide decision makers with methodologically sound assessments, given the stated objectives. The IEG ICR reviewer also needs to assess the extent to which the methodology proposed in the PAD would enable the assessment of attribution. The specific questions in assessing M&E design are the following:

- To what extent was the TOC (documenting how the key activities and outputs led to the outcomes) sound and reflected in the results framework?
- To what extent were the objectives clearly specified?
- To what extent were there indicators encompassing all outcomes of the PDO statement?
- To what extent were the intermediate results indicators adequate to capture the contribution of the operation's components (activities) and outputs toward achieving PDO-level outcomes?
- To what extent were the indicators specific, measurable, achievable, relevant, and time bound? To what extent were there baselines and targets available for all indicators?
- To what extent were the proposed sampling methods, data collection methods, and analysis appropriate for all indicators? How were comparators selected and handled?
- To what extent did the design ensure that a baseline, if relevant, would be done in time?
- To what extent were the M&E design and arrangements well embedded institutionally?

### Monitoring and Evaluation Implementation

In *M&E implementation*, the IEG ICR reviewer should assess to what extent the input, output, outcome, and impact evidence anticipated in the design was collected and analyzed in a methodologically sound manner. Specifically, these questions should be answered:

- To what extent was planned baseline data collection carried out?
- To what extent were the indicators included in the results framework measured and reported in the Implementation Status and Results Reports? (As a reminder, the ICR itself should report on this. The IEG ICR reviewer does review the Implementation Status and Results Reports.)
- To what extent were any weaknesses in M&E design, including specification of indicators, corrected during implementation?

- To what extent did the agency responsible for M&E (and any other relevant stakeholders) ensure attention to effective M&E implementation?
- To what extent are the data found to be reliable and of good quality? Important elements of this include sound methodology, independence of analysts, and quality control.
- If relevant, to what extent were beneficiaries involved in defining target indicators and assessing their achievement?
- To what extent are M&E functions and processes likely to be sustained after project closing?

### Use of Monitoring and Evaluation Data

The IEG ICR reviewer should assess, first, to what extent the M&E findings were communicated to the various stakeholders and, second, to what extent this informed strategic redirection and resource reallocation or is expected to lead to these in follow-on interventions. The specific questions to be answered in assessing M&E utilization are the following:

- To what extent were M&E findings communicated to the various stakeholders?
- To what extent can positive (or negative) shifts in the implementation direction of the project or program be attributed to the M&E activities?
- To what extent were the M&E data used to provide evidence of achievement of outcomes, as opposed to only providing evidence of application of inputs or achievement of outputs?
- To what extent did the M&E data or findings inform subsequent interventions, or to what extent are they expected to influence subsequent interventions in the near term?

### Determining the Overall Monitoring and Evaluation Quality Rating

The quality of M&E is rated on a four-point scale—negligible, modest, substantial, or high.

- **High.** There were minor shortcomings, or only minor shortcomings, in the M&E system's design, implementation, or utilization. The M&E system as designed and implemented was more than sufficient to assess the achievement of the objectives and to test the links in the results chain. M&E findings were

disseminated and used to inform the direction of the project, strategy development, or future projects.

- **Substantial.** There were moderate shortcomings in the M&E system’s design, implementation, or utilization. The M&E system as designed and implemented was generally sufficient to assess the achievement of the objectives and test the links in the results chain, but there were moderate weaknesses in a few areas.
- **Modest.** There were significant shortcomings in the M&E system’s design, implementation, or utilization. There were significant weaknesses in the design or implementation (or both) of the M&E system, making it somewhat difficult to assess the achievement of the stated objectives and test the links in the results chain, or there were significant weaknesses in the use and impact of the M&E system.
- **Negligible.** There were severe shortcomings in the M&E system’s design, implementation, or utilization. The M&E system as designed and implemented was insufficient to assess the achievement of the stated objectives and test the links in the results chain, and the use and impact of the M&E system were limited.

### Relation of Monitoring and Evaluation Quality to Other Ratings

Strengths and weaknesses in M&E design should also be reflected in the World Bank **quality at entry rating**, and those in M&E implementation should be reflected in the **World Bank supervision rating** and the **implementing agency performance rating**. The **relevance of project design** includes an assessment of the results framework and its relevance to the objectives. Rating overlap is acceptable and may be expected.

Beyond these direct links to other ratings, weak M&E may often impede effective project management, and thus it can indirectly affect the project efficacy and efficiency rating. If M&E is insufficient, it can affect the IEG ICR reviewer’s ability to assess the achievement of the project’s objectives.

### Section 10: Other Issues—Safeguards, Fiduciary Compliance, and Unanticipated Impacts

OPCS requires that the ICR summarize key safeguard and fiduciary issues in the project, compliance with the World Bank policy and procedural requirements, and any problems that arose and their resolution, as applicable. It also asks that the ICR record any significant deviations or waivers from the World Bank safeguards or fiduciary policies and procedures.



## Safeguards

### What Is the Purpose of Environmental and Social Safeguards?

The Environmental and Social Standards Framework is designed to help borrowers to manage the risks and impacts of a project, and improve their environmental and social performance, through a risk- and outcomes-based approach. The World Bank's environmental and social standards are designed to ensure that the potentially adverse impacts of World Bank-supported programs on the environment and on people are avoided or minimized and that unavoidable adverse impacts are mitigated.

Projects supported by the World Bank through investment policy financing are required to meet the following Environmental and Social Standards:

- **Environmental and Social Standard 1:** Assessment and Management of Environmental and Social Risks and Impacts
- **Environmental and Social Standard 2:** Labor and Working Conditions
- **Environmental and Social Standard 3:** Resource Efficiency and Pollution Prevention and Management
- **Environmental and Social Standard 4:** Community Health and Safety
- **Environmental and Social Standard 5:** Land Acquisition, Restrictions on Land Use and Involuntary Resettlement
- **Environmental and Social Standard 6:** Biodiversity Conservation and Sustainable Management of Living Natural Resources
- **Environmental and Social Standard 7:** Indigenous Peoples/Sub-Saharan African Historically Underserved Traditional Local Communities
- **Environmental and Social Standard 8:** Cultural Heritage
- **Environmental and Social Standard 9:** Financial Intermediaries
- **Environmental and Social Standard 10:** Stakeholder Engagement and Information Disclosure

This framework replaces the following Operational Policy (OP) and Bank Procedures (BP): OP/BP 4.00, Piloting the Use of Borrower Systems to Address Environmental and Social Safeguard Issues; OP 4.01, Environmental Assessment; OP 4.04, Natural Habitats; OP 4.09, Pest Management; OP 4.10, Indigenous Peoples; OP 4.11, Physical Cultural

Resources; OP 4.12, Involuntary Resettlement; OP 4.36, Forests; and OP 4.37, Safety of Dams.

This framework does not replace OP/BP4.03, Performance Standards for Private Sector Activities; OP 7.50, Projects on International Waterways, and OP 7.60, Projects in Disputed Areas.

During project identification, the World Bank screens a project with these safeguard policies in mind and classifies the project (including projects involving financial intermediaries) into one of four classifications—high, substantial, moderate, or low risk—based on the significance of environmental and social risk. The World Bank will review the risk classification assigned to the project on a regular basis, including during implementation, and will change the classification where necessary to ensure it continues to be appropriate.

### **How Are Safeguard Requirements Addressed in Projects?**

The World Bank will agree on an Environmental and Social Commitment Plan (ESCP) with the borrower. The ESCP will set out the material measures and actions required for the project to meet the Environmental and Social Standards over a specified time frame. The ESCP is part of the legal agreement. The draft ESCP is disclosed as early as possible, and before project appraisal.

### **Who Is Responsible for Complying with Safeguard Policies?**

The World Bank will require the borrower to implement the measures and actions identified in the ESCP, in accordance with the time frames specified in the ESCP, and to review the status of implementation of the ESCP as part of its monitoring and reporting.

### **What Information Are the PAD and ICRR Expected to Provide on Safeguard Compliance?**

The PAD should include a section on Environmental and Social Standards, including a table on the environmental and social (E&S) standards relevance given its context at the time of appraisal. The PAD will highlight the overall E&S risks—high, substantial, moderate, or low. It also briefly identifies the main E&S risks and impacts of the proposed project. For a high-risk or substantial-risk project, the World Bank will indicate in the PAD the project-related documents that will be prepared and disclosed after Board approval. The PAD includes a summary of the key environmental and social management measures, as set out in relevant E&S documents that have been prepared by appraisal and any E&S documents that will be prepared after Board approval and as committed in the ESCP. These E&S documents could include Environmental and Social Management Framework, Environmental and Social Impact Assessment, Resettlement

Policy Framework, Resettlement Action Plan, Indigenous People's Framework, Labor Management Procedures, and Stakeholder Engagement Plan.

### **What Should the Reviewer Record in the Safeguard Section?**

The reviewer should note the following information:

- The applicable E&S standards, if any, the overall E&S risk of the project at appraisal (high, substantial, moderate, or low risk), and (for high- and substantial-risk projects) the assessment instrument and mitigation plan;<sup>4</sup> these can be found in the PAD
- For category high- and substantial-risk projects, evidence that the project completed the planned mitigation activities, from the ICR
- The findings of any independent review of safeguards implementation (for high-risk projects) or monitoring reports (for others)
- If the physical components of the project that generated environmental or social effects were modified—through additional financing or project restructuring, for example—the reviewer should note whether the environmental or social assessment was updated or a new assessment prepared.

The absence of any of the above required information in the PAD or ICR should be noted.

### **Are Any of the Ratings Affected by Performance on Safeguard Compliance?**

There is currently no formal rating for safeguard compliance. However, the results in the safeguards section will affect other ratings.

- If the ICR fails to document any of these issues, that should also be mentioned in the ICR quality section and contribute to the assessment of that rating. Likewise, an exemplary explanation of safeguard issues should also feed into the ICR quality rating.
- Good or weak performance in preparation (identifying the applicable policies, preparing the assessment and mitigation plans) should be a factor in the Bank performance quality at entry rating.
- The World Bank's performance in supervising safeguard compliance should be reflected in the Bank performance quality of supervision assessment.

- Good or weak performance in adequately mitigating the impacts of safeguard issues should enter into the World Bank supervision rating and the borrower performance rating.

## **Fiduciary Issues**

### **What Constitutes a Fiduciary Issue?**

Fiduciary issues refer to compliance with operational policies on financial management (OP/Bank Procedure 10.02), procurement (OP/BP 11.00), and disbursement (OP/BP 12.00). This material is to be culled from throughout the ICR.

### **Financial Management**

Financial management issues involve the adequacy of the project's institutional financial management arrangements, reporting and accounting provisions, internal control procedures, planning and budgeting, counterpart funding, flow-of-funds arrangements, external audit reporting, and project financial management and accounting staff issues. Particular attention should be paid to the timeliness of project external audits and whether the external auditors' opinions were qualified. If so, the nature of the qualifications (that is, whether they were serious or merely administrative) and the measures taken to address them should be included. If the ICR does not offer comments on the latter, the review should note the absence of information.

Other important aspects of financial management include the following:

- The extent of compliance with financial covenants (this should be reported in the ICR)
- Whether all World Bank, International Development Association, and (where relevant) trust fund resources were fully accounted for by the time of project evaluation
- Issues of corruption or misuse of funds associated with the project, and how they have been addressed
- Whether all audit recommendations had been addressed by the time of project evaluation

### **Procurement**

Procurement issues include the following: the extent to which World Bank procurement guidelines were followed; significant implementation delays because of procurement-related issues and their causes; and evidence of timely World Bank intervention in

resolving procurement difficulties, providing procurement advice, or in giving nonobjections. Common causes of procurement-related delays or issues include misprocurement, low procurement capacity in the implementing agency, and lack of consistency between World Bank and national procurement laws and regulations. Any issues of this nature should be discussed in the section and mentioned in the Bank performance quality of supervision section.

### **Unanticipated Positive and Negative Effects**

Even when a project's objectives are not achieved, implementation often yields many benefits. However, those benefits are not taken into account in the assessment of the objectives. An unanticipated benefit is a positive or negative benefit or externality that occurred outside the framework of the stated objectives of the project. To be included in this section, they must be truly unanticipated (in the PAD or program document), attributable to the project, quantifiable, of significant magnitude, and at least as well evidenced as the project's other outcomes. Where there are unintended benefits, an assessment should be made of why these were not internalized through project restructuring by modifying either project objectives or key associated outcome targets.

### **Section 11: Ratings Summary**

This table displays main ratings from the ICR and compares them with IEG's ratings in the ICRR. The ICR ratings are automatically pulled from the Operations Portal, and the IEG ratings are pulled from the earlier sections in the ICRR form.

In cases where the ratings diverge, the IEG ICR reviewer should explain the reasons for divergence. The explanations can be short and link back to summary statements in earlier sections of the ICRR. The explanation for divergences in outcome rating can repeat elements in section 6, which summarizes the outcome rating.

### **Section 12: Deriving Lessons**

What are lessons?

- They are knowledge or experience gained by participating or completing a process, activity, project, or program.
- They are derived from decisions that had positive or negative consequences or tradeoffs during project or program preparation, design, implementation, and supervision.

- They are about situations or decisions that were within project or program control. They may relate to how the project or program shifted or adapted to situations outside its control.
- They flow logically from the available evidence.
- They require analysis and sense-making by implementing teams and evaluators.
- They are not facts, findings, or recommendations. Lessons are not suggestions on what to do differently.
- They are a prescription and guide others into action. Lessons are about behavior change.
- Most importantly, they should not overgeneralize (that is, extend beyond the available evidence). If the operation is in a fragility, conflict, and violence (FCV) context, the lesson should not be targeted to all FCV, but at most, to FCV in situations similar to that in which the project is being implemented. Even better, except under certain circumstances, lessons should be situated within the specific country context.

Why do we write lessons?

- They support accountability and learning.
- They help lead to behavior change.
- They help others understand what worked and what did not, and what factors influenced success or failure.
- They are aggregated and used to help others change behavior within the World Bank.

The purpose of the ICR is twofold: accountability and learning. Getting the ratings right is important, but learning what works, what does not work, and why is the key to greater effectiveness in the future. An ICR without good lessons is a missed opportunity to learn and do better. ICRs for projects that do not achieve their objectives often produce some of the most valuable lessons.

The ICRR typically presents three to five key lessons that emerge from the information in the report. They may come from the ICR, or they may be reflections on this project from the IEG ICR reviewer based on the ICR, compared with other projects the IEG ICR reviewer has reviewed (or, for example, confirming that the findings in this ICR underscore evaluative findings or lessons from other IEG evaluations). Whatever the

case (whether the lesson is from the ICR or from IEG), it is incumbent on the IEG ICR reviewer to identify the source.

Even if lessons in the ICR are not well formulated, the ICRR should formulate them well. The two biggest issues in formulating lessons—in the ICR and by IEG ICR reviewers—are the following: (i) they are formulated as facts, findings, or recommendations, rather than lessons; and (ii) they are not underpinned by the evidence in the ICR. Table 13.1 distinguishes among facts, findings, lessons, and recommendations.

**Table 2.5. The Difference among Facts, Findings, Lessons, and Recommendations**

	What Is It?	Example
Fact	What happened—an event and data (results); not in dispute	“The project manager was dismissed in year 5.”
Finding	What the analyst interpreted or concluded from the facts specific to the project; can be disputed	“Mainly because replacement of the project manager was delayed, the project did not meet its targets.”
Lesson	The broader significance of a finding; draws a conclusion from experience that may be applicable beyond the project under review	“Inability to effectively address poor project management performance and hire a replacement efficiently caused financial and procurement delays that negatively affected activity implementation, contributing to the project not meeting its timelines, outcomes, and targets.”
Recommendation	Suggests how to proceed in the future in the light of this experience; proposes actions	“The borrower should ensure that key project management positions are filled with competent staff. The World Bank should help ensure this through appropriate covenants and prompt supervision.”

Source: Independent Evaluation Group.

Facts and findings are found throughout the ICR; they are the material from which lessons are built. If something is repeated verbatim from the ICR, then it probably is not a lesson, but a fact or finding. If the draft lesson has the words *should* or *needs to or ideally*, then it is very likely a recommendation and not a lesson.

Lessons should be clearly and concisely stated.

In the ICRR form, the “lessons” stated (whether from IEG or the ICR) should be properly formulated and evidenced, and the source cited (whether from the ICR or IEG). Comments on the quality of the lessons (including the extent to which they are evidence based) belong in the section on the quality of the ICR. Remember that even projects rated unsatisfactory with weak M&E generate important lessons.

## How to Write a Lesson

- Structure: One topic sentence supported by three to five sentences explaining the evidence. Direct the information toward the target audience and identify the decision that led to the lesson.
- If the lesson has the words *should*, *likely*, *need to*, *consider*—it is likely a recommendation and not a lesson.
- Validate and engage with the TTL and ask, “If there is another project like this being developed, do you have some up-front advice on the process or implementation? What would you do differently or the same and why?”

To write lessons, consider the following guidance:

- **To be actionable.** The lesson needs to describe the decision or lack of decision that was taken, when the decision was made, and the type of project or context that took place.
- **To be relevant.** The lesson needs to be broad enough to be applicable to as many similar situations or projects as possible. Although there is much experimentation taking place at the World Bank, the lesson needs to be applicable to similar contexts and be evidence based.
- **To be nongeneric.** The lesson needs to have detailed information on the specific decision or cause of the effect to be applicable. For example, if the lesson is about indicators, it needs to discuss which type of indicators aimed to measure which outcome.
- **To be specific.** The lesson needs to include information about the project. Sometimes it is difficult to include all the information in one statement. That is why it is important to follow up the lesson statement with sufficient three to five sentences that provide context, data, and specific project or decision-related information.
- **To be valid.** The lesson needs to be evidence based and avoid sector-based speculations. This means that there needs to be enough data to support the lesson. Writers should avoid picking and choosing which evidence they liked most. The lessons and data need to be consequential.



## Section 13: Assessment Recommended

An assessment is recommended for any one (or more) of the following reasons:

1. Requested by Region, executive directors, or cofinanciers
2. Input into IEG's cluster or country sector review
3. Input into IEG's sector or thematic study
4. Input into Country Assistance Evaluation
5. Potential impact evaluation
6. Innovative project/new instrument
7. Major disagreement with ICR (rating or evidence), low ICR quality, or both
8. Underevaluated country/sector/theme
9. Major safeguard compliance issue

The first six reasons correspond to IEG's learning function, and the remaining three to its accountability function. If the project is recommended for a field assessment, the reviewer should offer a rationale. This will flag the project as a possible candidate, but it will not necessarily lead to a field visit.

## Section 14: Quality of the ICR

Because the ICRR is almost entirely based on the information found in the ICR, the reliability of IEG's ratings based on the desk review depends critically on the accuracy and quality of the evidence it provides. For this reason, IEG rates the quality of the ICR.

### Criteria

The assessment of the quality of the ICR is based on the following criteria:

- **Quality of evidence.** Is the evidence from a credible source, appropriately referenced, and presented in a parsimonious fashion? Does the ICR, including annexes or appendixes, present a complete and robust evidence base to support the achievements reported? Is there an attempt to triangulate evidence and to fill in gaps that might exist in the results framework? Is the evidence aligned to the associated TOC?
- **Quality of analysis.** Has there been sufficient interrogation of the evidence, concise summarizing of salient points, and clear linking of evidence to findings?

Does the analysis seek to engage with and respond to the associated TOC in the report?

- **Lessons.** Regarding the extent to which lessons are based on evidence and analysis, are the lessons appropriately responding to the specific experiences and findings for the project? Are they sufficiently linked to the narrative and ratings included in the report? Are they illustrative of the operational realities encountered through the project?
- **Results orientation.** Does the report emphasize and highlight how activities inform outcomes, which in turn is linked to the impact of the project's intervention? Is the report focused on what occurred as a consequence of the project? Is there sufficient clarity on the expected transformative impact of the project?
- **Internal consistency.** Is there a logical linking and integration of the various parts of the report and are the results mutually reinforcing? Does the report consistently revert to and articulate progress against the objectives and the TOC?
- **Consistency with guidelines.** Has the report followed and responded to the guidelines, with regard to both ratings and the performance narrative?
- **Conciseness.** Is there sufficient clarity in the report's messaging? Is the performance story direct, well informed, and tightly presented? Does it meet OPCS guidelines on length? Is there redundant or unnecessary content?

## Guidelines

In commenting on the quality of the ICR, it is generally a good strategy to begin by highlighting the strengths of the ICR before touching on the weaknesses. Candor, for instance, is highly valued. A successful overview of quality of the ICR may not include comments answering all seven criteria outlined above, but it should provide a response addressing at least four relevant criteria. See the ratings profiles in the next subsection for more complete information on what to look for in a high-quality ICR.

Problems in ICR quality that should be flagged include the following: inadequate evidence; incomplete ICRs (such as missing data in tables, no discussion of efficiency); failure to assess the objectives; and relying too much on monitoring indicators instead of using all available data. IEG does not downgrade the ICR quality simply because of a difference in opinion about the ratings. There does not need to be a correlation between the project's outcome rating and the quality of the ICR. Some of the best ICRs have been written for projects that were unsatisfactory.

## Rating of the Quality of the ICR

Based on these criteria, the rating profile is as follows:

- **High.** The ICR is tightly written and provides a complete critique of the project. There is a clear link between the TOC, the narrative, the ratings, and the evidence. It provides a candid, accurate, and substantiated set of observations that are aligned to the PDO. The report is concise, follows the guidelines, seeks to triangulate data to reach conclusions, and is focused on results. The quality of evidence and analysis is substantial and informs all aspects of the ICR, and there are few lapses in the quality of data and information. There is a well-articulated TOC informing the reader as to how the ratings have been reached and the lessons are specific, useful, and based on evidence of what actually occurred in the project.
- **Substantial.** The ICR provides a detailed overview of the project. The narrative supports the TOC, the ratings, and available evidence. It is candid, accurate, and generally aligned to the PDO. The report is concise, follows the majority of the guidelines, makes an attempt to triangulate data to reach conclusions, and is focused on results. The quality of evidence and analysis is aligned to the messages outlined in the ICR, though there may be some minor shortcomings in the completeness of data and information. There is a reference to the project's TOC that helps the reader to understand how the ratings have been reached. The ICR's lessons are clear, useful, and based on evidence outlined in the ICR.
- **Modest.** The ICR provides a comprehensive overview of the project. The narrative loosely supports the TOC and the ratings, and there are some gaps in evidence. It is relatively candid, predominantly accurate, and generally aligned to the PDO. The report covers a wide range of issues, follows most of the guidelines, and is focused on results. There is an attempt to link the quality of evidence and analysis to the messages outlined in the ICR. There is an effort to articulate how the ratings have been reached, and there may be some gaps in data and in various sections of the ICR. The ICR's lessons are generally useful and based on evidence outlined in the ICR but may be overly general and not targeted to specific events or actions.
- **Negligible.** The ICR provides a basic overview of the project. The narrative loosely supports the TOC and the ratings, and there are obvious and consistent gaps in evidence. It is not particularly candid, may have evident inaccuracies, and is not always aligned to the PDO. The report covers a range of issues, follows some of the guidelines, and is irregularly focused on results. The link between the quality of evidence and analysis in the ICR is not always reflected in

the report's messages. There is some attempt to articulate how the ratings have been reached, but there are evident gaps in data and in various sections of the ICR. The ICR's lessons are likely to be overly general and not targeted to specific events or actions.

### **3. Other Considerations for the Implementation Completion and Results Report**

**Does the ICRR confine itself only to evidence on the project’s key performance indicators that were identified in the PAD or Implementation Status and Results Reports?**

No—all evidence, regardless of the source, is to be brought to bear in preparing the ICRR, so long as the evidence is of quality. The ICR should provide enough information for IEG to be able to assess the quality of the evidence (for example, the methodology of the beneficiary assessment, how the control group or comparison group was selected, and so on).

**How should interlinks between ratings be addressed?**







For the most part, each of the key ratings—outcome, Bank performance—measure distinct dimensions of development effectiveness and are independent of each other. So, for example, outcome may be rated highly unsatisfactory (for example, in a fragile state where a political coup erodes government commitment and the project objectives remain unachieved) while Bank performance may be rated highly satisfactory (if the political coup was wholly unpredictable and the World Bank had done the best it could under the circumstances), or vice versa.

In practice, however, there can be a number of interlinks among the ratings that must be borne in mind to ensure internal consistency among the ICRR ratings. Some of these interlinks are deliberate and obvious, but others are not so obvious (Refer Table 3.1):

- The IEG-OPCS Harmonized Evaluation Criteria introduce a deliberate interlink between the outcome and the Bank performance ratings—when the two elements of Bank performance are in opposite directions (one above the line and the other below), the outcome rating becomes the tiebreaker.
- There is another deliberate link between the Bank performance rating and the following subratings or dimensions: M&E quality rating, safeguard compliance, fiduciary compliance, and unintended positive and negative effects. These subratings or dimensions were introduced as separate sections in the ICRR form to specifically zoom in on and give prominence to particular aspects of Bank performance. So, if, for example, M&E quality is rated unfavorably, Bank performance cannot be rated too favorably. Similarly, if there is weak fiduciary compliance, Bank performance may be affected.

- Not-so-obvious interlinks can manifest themselves between the Bank performance ratings and the outcome rating depending on the extent and nature of weaknesses in the above mentioned subratings or dimensions. So, for example, if M&E quality is extremely weak or M&E is nonexistent, that would raise questions about how effectively project implementation could have occurred and, therefore, how favorable efficacy (and hence outcome) could be. Similarly, if there were significant unintended negative effects attributable to the project (for example, a project in which a road was built in line with the stated project objectives and was also efficiently built, but the surrounding areas were deforested in the process), outcome could not be rated favorably. Also, if fiduciary compliance was weak and there was evidence of substantiated corruption, then that would signal an inefficient use of project resources and outcome could not be rated favorably.
- There is a final set of not-so-obvious interlinks that the IEG ICR reviewer needs to also be aware of: although the outcome rating and the Bank performance ratings can certainly go in different directions, the IEG ICR reviewer should be able to explain the reasons for any divergence. Generally, if outcome is rated unfavorably, Bank performance would typically also be rated unfavorably. There are some exceptions, however: First, an unfavorable outcome rating could be associated not with inadequate implementation performance on the part of the World Bank but rather on the part of other donors or cofinanciers, if there were such other donors or cofinanciers in the project. Second, an unfavorable outcome rating could be associated with an exogenous shock (for example, an earthquake that wipes out the project roads). Finally, it could be that the World Bank supports a high-risk project, the risk materializes, and the World Bank makes an informed decision not to cancel or restructure the project because the rewards could be extraordinarily high if the project succeeds—in such a case, if the project fails to achieve its objectives, outcome could be rated unfavorably while Bank performance could be rated favorably. In any event, when the ratings diverge, the IEG ICR reviewer should be able to explain the reasons for the divergence. Otherwise, the divergence may be unjustified.

**Table 3.1. Red Flags” or Examples of Ratings Patterns to Check**

Scenario	Red Flag or Likely OK?	Explanation
Outcome MS+ M&E quality modest	 Red flag	Although there can be a good M&E system showing a lack of outcomes, it would be strange to have evidence of good outcomes with a weak M&E system, although it could be that an independent impact evaluation was done that showed the achievement of outcomes despite weak M&E.
Outcome MU– M&E quality substantial	 likely OK	
Outcome: MS+ Bank performance: MU–	 Red flag	Unless there was a negative unexpected shock that caused the outcome to disappear, it would not make sense to have an outcome rating rated MS or higher together with a low rating for Bank performance. (If there was a positive shock that created the outcome then it cannot be attributed to the World Bank.)
Outcome: MU– Bank performance: MS+	 likely OK	However, if there was a negative shock (for example, an earthquake) that eliminated the outcomes, we could still have Bank performance being rated MS or higher (they did a good job but could not have foreseen the earthquake).
Efficacy: substantial+ Efficiency: modest–	 likely OK	It is very possible for a project to achieve outcomes but achieve them inefficiently; however, if the project did not achieve outcomes, then efficiency would be quite unlikely to be high, because efficiency, in essence, compares project achievements to project costs.
Efficacy: modest– Efficiency: substantial+	 Red flag	

Source: Independent Evaluation Group.

Note: M&E = monitoring and evaluation; MS = moderately satisfactory; MU = moderately unsatisfactory.

## 4. Guidance Specific to Fragility, Conflict, and Violence Context

Assessing results and performance in FCV situations is more challenging compared with regular environments<sup>1</sup>. Although some of the issues can be similar to non-FCV environments, the scale of the challenges is often larger because of a fast-changing and difficult operating environment, involving either generally weak state institutional capacity, high levels of exclusion, limited provision of basic services to the population, or a combination of the aforementioned and necessitating longer-term engagements to achieve any observable changes. Insecurity and low capacity in the field may significantly affect the possibility of collecting data or assessing results. Each FCV situation is different and some of the guidance may apply more to one situation than another.

### Project Context and Development Objectives

#### Context at Appraisal

**Context.** Provide a brief story line of the operation's context, including the FCV drivers of fragility, sources of resilience, and risk factors (DRRs).

**TOC.** Effective interventions in FCV situations are rooted in a good TOC. Teams need to document the operation's TOC including how it is affected by or will affect the FCV context. Links between the operation design and the FCV DRRs in the country need to be clear. Some specific considerations to bear in mind when establishing a TOC in an FCV environment are as follows:

- The DRRs should be addressed to facilitate achievement of the PDOs in the FCV context. The TOC explicitly illustrates any links between the operation's interventions and the FCV dynamic in the country. It shows how the interventions will affect the FCV identified driver of fragility in the country. Together with consideration of the resilience factors, this will feed into the underlying assumptions and help improve the understanding of how the operation intends to contribute to high-level outcome(s) related to the country's FCV situation (for example, improved social cohesion, increased legitimacy of government institutions, economic development in lagging regions).
- The TOC should reflect the capacity limitations in the FCV context and risks underlying the operation's design.



- The TOC should reflect changes to the intervention logic during implementation and any adjustments made in response to volatility, characteristic of many FCV environments.

**PDO.** Operations in an FCV environment may seek to achieve outcomes at different levels. If a fragility-related element is included in the PDO, make sure that it is measurable and within the reach of the operation (based on the TOC). That is, the PDO should be set at a level that is achievable at closing given the project or program design. Consideration here should be made for lower-level objectives.

### **Significant Changes during Implementation (if Applicable)**

**Changes.** The political, social, or conflict environment in an FCV situation may change very quickly and this may have a significant impact on the achievability of the original PDO. Changes to PDOs during implementation because of changes in country context need to be described (for example, change in the beneficiaries targeted or change of the geographic focus of operations). State the reason(s) for changes and describe how the changes affected the TOC, the original PDO, and the operation design, including any impact on the FCV DRRs that the operation was designed to address. Where possible, it would be advisable to include all versions of the TOC to clearly illustrate the rationale behind any PDO adjustments.

### **Outcome**

As with other investment operations, to derive the overall outcome rating, this section assesses the operation's relevance, efficacy, and efficiency.

### **Relevance**

The application of relevance to FCV environments should focus on the suitability and realism of the PDO as it pertains to the given context in which the operation is to be delivered. The PDO should reflect conditions in the field and be pitched to the capacity of relevant delivery partners. The assessment of relevance of objectives in FCV environments is based on the following specific criteria, in addition to the requirements outlined in existing OPCS policy guidelines:

- There should be a clear articulation of the specific development challenge the operation seeks to address and the role the World Bank intends to play in responding to this challenge. The explanation should

- Provide details on whether the operation is responding to fragility drivers, sources of resilience and risk factors (DRR) identified in the Risk and Resilience Assessment, or both;
- Explain how the operation is aligned with the development partners' commitments and interventions; and
- Articulate the wide range of constraints and risks that may impede progress toward addressing the problem.
- There should be an explanation on the suitability of the objective's outcome orientation. Is the PDO appropriately pitched for the country context? This judgment includes government capacity, relevant fragility issues, and risks in the given context. Are the specific objectives the operation seeks to achieve based on the World Bank's recent experience in the given sector in the country and past attempts to tackle the same problem by the World Bank or others? For example, is the operation occurring in an immediate postconflict period and therefore likely to confront considerably greater challenges than operations responding to more endemic FCV constraints in more settled fragile environments? Second-, third-, and fourth-phase operations within a given sector in an FCV environment should include outcomes consistent with reasonable progress over time or explain continued barriers to progress toward achieving expected outcomes.

**Relevance of disbursement-linked indicators (DLIs; for Program-for-Results).** For Program-for-Results, the assessment of relevance comprises two parts: (i) the relevance of the program's objectives (as described above); and ii) the relevance of the DLIs in providing incentives for the achievement of results. In an FCV context, the following should also inform the assessment of the relevance of DLIs:

- Are the DLIs consistent with the underlying fragility and conflict dynamics and the need to strengthen public institutions?
- Is the number of DLIs appropriate given the capacity and implementation constraints in the FCV context?

### **Efficacy**

In some FCV contexts, it is difficult to accurately plan the results to be achieved over a given time horizon (usually more than five years). Operations usually include flexibility that allows for rapid adaptation to changing circumstances. However, this rapid adaptation and course correction needs to be timely to have any development impact by

operation closing. The team should still take into consideration the requirements outlined in existing OPCS policy guidelines regarding the application of a split rating.

If the PDOs include a broad or higher-level outcome related to FCV environment, specific objectives can be defined from the expected measurable lower-level objectives. The contribution of the selected lower-level outcomes to the broad or higher-level outcome should be described in the operation's TOC; in which case, consideration here can be made for assessing lower-level objectives and those that might directly respond to the DRRs. For example, the objective of "bringing social cohesion to the conflict-affected population" could be made more specific by linking it to "improved access to local conflict resolution mechanisms." It is expected that a clear and concise narrative be provided on the line of sight for contribution of the operation to high-level outcome(s) related to FCV environments.

M&E data to measure project results often prove difficult to collect in an FCV environment. Teams need to be realistic about what type of data can be collected and used to assess achievement of the PDOs. The choice of indicators in the results framework needs to reflect this situation or challenge. In this regard, it is helpful to mobilize both quantitative and qualitative (for example, perception surveys) data available, even beyond the results framework indicators. It is important to leverage and document the sources of data and add secondary sources, qualitative assessment, data triangulation, or stakeholder interviews and beneficiary feedbacks when reliable data are not available.

Starting early with preparations for completion may help undertake data retrofitting exercises needed to document operation achievement. Smaller, rapid, and targeted evaluations on specific topics can be commissioned to underpin the operation's achievements related to FCV issues, which may not have been well captured in the results framework or during implementation.

Since one of the objectives of World Bank Group engagement in FCV is to strengthen core state functions and support national systems, the ICR may wish to consider the "improved institutional resilience and capacity" as part of the contributing results to the PDO (if it is not an explicit objective in the PDO statement). For example, if client M&E institutional capacity was built as part of the implementation of the M&E system, this needs to be mentioned in the completion report, since it is a key contribution to strengthen (fragile) institutions and to enhance sustainability of development outcomes (in FCV countries). It could also be a contribution to improving transparency and accountability of institutions if data were shared with the beneficiaries during the operation's implementation. Capturing the (positive) impact the operation may have

had on drivers of fragility (if any) and how this may have helped reduce the fragility and improved institutional capacity and resilience in a country, or at least in the operation target area, could be an important objective to be assessed as part of efficacy in the FCV context. Therefore, task teams have the option to capture this positive impact related to “capacity strengthening” as an additional objective to count toward the overall efficacy rating.

**Assessment of institutional capacity objective.** This objective is the opportunity for the task teams to highlight the importance of the institutional capacity strengthening in low-capacity environment characteristic of FCV settings. The use of the operation’s actions and outputs related to institutional capacity, systems, and procedures may provide the basis for assessing the improvement in client institutional capacity. It is also important to provide evidence that improved institutional capacity, systems, and procedures contributed to an improved response to the drivers of fragility and sources of resilience (for example, improved transparency, accountability of institutions, and state local presence).

- **Example 1.** PDO statement: “To improve access to critical socioeconomic services and increase employability of youth in FCV-impacted communities.” In this example, there are three objectives (or outcomes) to be assessed. Assess separately each of the two outcomes included in the PDO statement: (i) to improve access to critical socioeconomic services, and (ii) to increase employability of youth. In addition, assess a third implicit objective related to institutional capacity strengthening.
- **Example 2.** PDO statement: “To improve access to social services for vulnerable communities and strengthen institutional capacity of the local government institutions in selected regions.” In this example, since institutional capacity strengthening is already included as an explicit objective in the PDO statement, there are only two objectives (or outcomes) to use for assessing efficacy: (i) to improve access to social services for vulnerable communities, and (ii) to strengthen institutional capacity of the local government institutions in selected regions.

Overall efficacy (extent of achievement of the PDOs) is rated on a four-point scale (high, substantial, modest, and negligible). In FCV operations with multiple objectives, provide a single overall efficacy rating covering all objectives or outcomes (including the institutional strengthening if it was not included as an explicit outcome in the PDO statement). Even though each objective does not need to be separately rated, each needs to be discussed and information for demonstrating its achievement provided.

## **Bank Performance, Compliance Issues, and Risk to Development Outcome**

### **Quality of Monitoring and Evaluation**

The results measurement framework in FCV should be adapted to be sensitive to FCV dynamics. Managing for results and developing an appropriate M&E system in FCV contexts is more challenging compared with non-FCV environments. Although some of the M&E issues can be similar, the magnitude and scale of the challenges in FCV is often larger because of a fast-changing and difficult operating environment, with generally low and fragile institutional capacities. Insecurity and violence in the field may also significantly affect the ability to access project sites and beneficiaries to collect data or measure project progress.

The assessment of the quality of M&E involves the following two main elements: whether the M&E framework and system was appropriately adapted to the FCV context, and whether it considered effects on fragility and conflict itself, where appropriate. Some specific considerations to bear in mind when assessing the quality of M&E in an FCV context are as follows.

### **Adaptation of the Monitoring and Evaluation Framework and System to Fragility, Conflict, and Violence Capacity and Constraints**

- Does the TOC include clear links between project interventions and the drivers of fragility and conflict in the country, and has it been used as a basis for developing the operation's results framework?
- Are results indicators closely aligned to the intended PDO and TOC? Was the project M&E framework tailored to the specific conditions, constraints and capacity limitations related to FCV (for example, data paucity, insecurity and inability to access project sites and affected people, political instability, low capacity of government, changing conditions and fluid situations in the field, and so on)?
- Was the M&E framework designed to capture indicators related to fragility and conflict drivers, FCV risks, and factors of inclusion and resilience, as relevant? This involves
  - Disaggregating indicators where feasible to track differential effects on social groups, regions, and so on with specific attention on vulnerable, excluded, or aggrieved populations; and

- Identifying and collecting indicators to monitor the FCV context if the project aimed to address drivers of fragility and conflict directly or aimed to do no harm through conflict sensitive design.

### Use of Monitoring and Evaluation Framework and System to Assess Progress and Respond to Changing Conditions in the Field

- Were M&E data collected and analyzed using practical and context-appropriate information and communication technology tools, and were any issues with data availability addressed on time?
- Was the client's M&E capacity developed during the project implementation?
- Was the framework simple enough for the country context and did it allow flexibility to respond to constraints and changing conditions in the field?
- Did the M&E system and data collection take into account coordination and coherence among different donors and partners to minimize reporting requirements for government counterparts, clients, and beneficiaries?
- Did the M&E system allow for frequent check-ins to monitor progress and risks in real time? Were the inputs from the system used to inform course corrections and promote learning during project or program implementation?
- Were client and project beneficiaries involved in the M&E implementation to improve transparency and accountability?
- Did the data collection and management efforts include ethical considerations related to handling sensitive information and risks to respondents?

### Bank Performance

Bank performance in FCV environments should provide detail and explanation on the extent to which the World Bank responded appropriately to the challenging environment in which the operation is to be delivered. This requires a response outlining how the World Bank adapted to existing and emerging fragility drivers in the country, both in design and during implementation. In addition to existing guidance on standard Bank operations, teams should provide detail on the following:

- **Design.** Did the operation respond to previous experience in the country? Did it identify relevant risks to success including fragility drivers outlined in the Risk and Resilience Assessment? To what extent did the operation mitigate these risks? This necessitates a broader explanation of the extent to which World Bank

staff identified and managed ex ante risks to achieving the PDO. It should also include an assessment of the up-front collaboration and coordination with development partners.

- **Implementation.** How did the World Bank team and its partners respond to emerging challenges and unanticipated developments? To what degree was the operation (if at all) changed? Importantly, the explanation should include the way partners were engaged and the centrality (or not) of dialogue in responding to relevant constraints.

### Risk to Development Outcome

The risk to development outcome in FCV environments needs to reflect the increased emphasis on managing risk to achieve the desired outcome. In addition to the details outlined in existing OPCS guidelines, teams should also provide details on the following:

- **Implementation risk.** Are the results sustainable and capable of supporting subsequent interventions? Teams need to explain what (if any) progress was made in building the capacity of relevant development partners and whether appropriate measures were in place to respond to given implementation challenges.
- **Risk mitigation.** Did the operation appropriately respond to the FCV challenges in the country? Were risk mitigation approaches required? Were they effective in diluting negative impacts? Teams should describe the level of achievement in managing risk and responding to implementation delays and problems. This description should include details on how the mitigation approach is likely to affect longer-term outcomes in the country.

### Lessons and Recommendations

Generate actionable knowledge and lessons learned that reflect Bank Group operational experience to inform future engagements in the sector, the country, and FCV context. Some specific considerations to bear in mind when capturing lessons learned in an FCV environment are as follows:

- Are they context specific?
- Are they drawn from specific implementation experience and operation adaptation and evolution experience?

- Are they oriented toward given fragility constraints and drivers?

Can they help build the evidence base for future World Bank Group and other partner and government interventions?



## **5. Note on Canceled Operations**

### **What Is a Note on Canceled Operation?**

A Note on Canceled Operation (NCO)<sup>1</sup> is prepared for a project that fails to become effective or is canceled before significant implementation is initiated.<sup>2</sup> The cut-off point for significant implementation is defined as final actual disbursement of less than 5 percent of the initial commitment or \$1 million (whichever is smaller), excluding any Project Preparation Facility and front-end fees. The NCO, which describes the project and explains why it was not implemented, is sent to the Board. The ICR guidelines also cover NCO requirements.

### **Which Sections of the ICRR Review Should Be Completed, and What Ratings Assigned?**

The OPCS guidelines for writing ICRs do not indicate which ratings are to be completed for the NCO: “the text should generally follow the relevant sections of the ICR Guidelines and briefly cover the project’s rationale and objectives, main events and factors leading to cancellation, and any lessons learned” (World Bank, 2021, appendix A).

For the purposes of the ICRR, sections 1, 2, 3, 8, 9, 10, 14, and 15 should be completed. The following ratings or subratings should be assessed:

- Relevance of objectives
- Bank quality at entry
- ICR quality (in this case, the NCO quality)

### **Rating the Quality of the Note on Canceled Operation**

To enable an assessment of its quality, the NCO is expected to discuss the main events leading to cancellation, steps taken to resolve problems, exogenous factors, identification of causes, and parties responsible if the project failed, and the implications of failure. Above all, the purpose of the NCO is to clearly explain why the project was canceled; if the NCO does not convincingly explain or document the reasons for cancellation, the quality would be rated unsatisfactory.

## Notes

### Introduction

<sup>1</sup> When insufficient information is provided by the World Bank for the Independent Evaluation Group (IEG) to arrive at a clear rating, IEG downgrades the relevant ratings as warranted. This practice began on July 1, 2006.

<sup>2</sup> For a subset of operations—on the order of 20–25 percent—IEG conducts Project Performance Assessments in the field.

<sup>3</sup> For Implementation Completion and Results Reports (ICRs) written before July 1, 2017, ICRs and IEG also rated risk to development outcome and borrower performance. Risk to development outcome is the risk, at the time of evaluation, that development outcomes (or expected outcomes) will not be maintained (or realized). Borrower performance is the extent to which the borrower (including the government and the implementing agency or agencies) ensured quality of preparation and implementation and complied with covenants and agreements, toward the achievement of development outcomes. Also, before July 1, 2017, IEG discussed and rated monitoring and evaluation quality in the ICR Review, although the World Bank did not do so in the ICR.

<sup>4</sup> For the World Bank, rating monitoring and evaluation quality began on July 1, 2017.

### Chapter 1

<sup>1</sup> For greater detail on ICR guidelines, see World Bank, OPCS 2021.

<sup>2</sup> The draft Review is also usually copied to the country director, the Regional sector managers, the last task team leader of the project, and key individuals in the relevant Global Practice. Identification of the individuals to be copied is the responsibility of the ICR Review coordinator.

<sup>3</sup> The Regional director may ask for an extension if necessary (for example, if key staff are on annual leave or a mission and unavailable to reply).

### Chapter 2, Sections 1–7

<sup>1</sup> The definition of cofinancing and the types are from OP14.20 (1995, revised in April 2013) and annex A to OP14.20. In joint cofinancing, procurement is carried out in accordance with the World Bank's procurement and consultant guidelines. In parallel cofinancing, the World Bank and cofinanciers finance their different components according to their own rules.

<sup>2</sup> The sum of the estimated costs of the components may differ from the total project cost because the component-costing often excludes contingencies. If there are two versions of component costs in the Project Appraisal Document, with and without contingencies, then the one with contingencies should be used. Otherwise, the IEG ICR reviewer should note that the appraisal estimates for the components exclude contingencies.

<sup>3</sup> In conducting Project Performance Assessments, the evaluator should plan to search for evidence of the influence of nonproject factors, such as economic trends, other government policies, other donor support, and exogenous factors that may also be affecting the anticipated outcomes.

#### **Chapter 2, Sections 8–14**

<sup>1</sup> The lists of assessment criteria below are taken from the Quality Assurance Group’s criteria for its seventh Quality of Entry Assessment (World Bank 2007) and its sixth Quality of Supervision Assessment (Hari Prasad 2005).

<sup>2</sup> This would include provisions for safeguard policy compliance.

<sup>3</sup> IEG does not provide ratings on the three separate elements of monitoring and evaluation quality.

<sup>4</sup> Any instruments or plans mentioned for other categories should also be mentioned, but they are required for category high and substantial risks.

#### **Chapter 4**

<sup>1</sup> This information is from appendix L of World Bank, OPCS 2021

#### **Chapter 5**

<sup>1</sup> This information is from appendix A of World Bank, OPCS 2021

## References

World Bank. 2021. *Implementation Completion and Results Report (ICR) for Investment Project Financing (IPF) Operations*. Operations Policy OPS5.03-GUID.161. World Bank, Washington, DC.

World Bank. 2024. *Preparing the Project Appraisal Document (PAD) for Investment Project Financing (IPF) Operations* Policy OPS5.03-GUID.165 World Bank, Washington, DC.

World Bank. 2016. *"P118101, Rwanda Skills Development Project."* Independent Evaluation Group, Implementation Completion and Results Report Review. ICR0020564, World Bank, Washington, DC.