HARMONIZED EVALUATION CRITERIA FOR ICR AND OED EVALUATIONS[1]

1. OUTCOME

Definition: *the extent to which the operation's major relevant objectives were achieved, or are expected to be achieved, efficiently.*

Guidance

- The outcome rating completes the series of DO ratings used in ISRs.
- For evaluation purposes, an operation's objectives encompass both the PDOs stated in board documents and key associated outcome targets.[2] This means that whenever the PDOs stated in the board documents are so broad and/or vaguely worded as to preclude any meaningful evaluation, intended objectives are inferred by the evaluator from key associated outcome targets (and/or the operation's design features as relevant).
- The rating of O*utcome* should encompass the extent to which the operation's institutional objectives 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) and/or the *extent* to which one or more objectives are not achieved (or are not expected to be achieved).
- Shortcomings in efficiency may have to do with the extent to which the operation 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 DPL operations).
- Shortcomings in relevance may have to do with the extent to which an operation's objectives, design, or implementation are inconsistent with the country's current development priorities and with current Bank country and sectoral assistance strategies and corporate goals (expressed in PRSPs, CASs, SSPs, OPs).[3]
- The evaluator must use judgment in weighing possible shortcomings in the achievement of the operation's objectives, in its efficiency, or in its relevance, and arrive at an assessment of how they affect the overall rating.

---

[1] / These guidelines apply to both investment and development policy lending. They were developed by a working group comprising staff from OPCS, OED and the Regions.

[2] / As the Bank is an objectives-based institution achievements against the PDO are paramount, while restructuring provides opportunity to internalize positive unintended results, or they can be taken into account as additional achievements if convincingly documented.

[3] / The ICR Guidelines and OED practice have been identical on this point for some years. 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.

Rating Scale[4]

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

In the rare instances where lack of sufficient information, or other circumstances, make it impossible to assign one of the above ratings, "*Not-rated*" should be recorded.

2. RISK TO DEVELOPMENT OUTCOME

Definition: *the risk, at the time of evaluation, that development outcomes (or expected outcomes) will not be maintained (or realized)*

Guidance

The *Risk to Development Outcome* rating has two dimensions:

(i) the likelihood that some changes may occur that are detrimental to the ultimate achievement of the operation's development outcome; and

(ii) the impact on the operation's development outcomes of some or all of these changes materializing.

[4] / In the guidance for these evaluation criteria links will be provided to ICRs and OED evaluations to demonstrate examples of the application of the ratings' definitions to real cases.

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

Other risks arise from factors outside the operation, be they country level such as price changes or global such as technological advances. The impact on outcomes of a change in the operating environment depends on the severity and nature of the change as well as the adaptability (or lack there of) of the operation's design to withstand that change.

The rating helps to identify operations that require close attention in managing risks that may affect the long-term flow of net benefits. Rating the *Risk to Development Outcome* requires an assessment of the uncertainties faced by an operation over its expected remaining useful life and whether adequate arrangements are in place to help avoid or mitigate the impact of those uncertainties. The impact will increase if the design or implementation of the operation is not well aligned with the operating environment, or mitigation measures are inappropriate to deal with foreseeable risks.

Whereas the *Outcome* rating reflects the evaluator's best estimate of the expected overall development outcome, the *Risk to Development Outcome* rating reflects the evaluator's judgment of the uncertainties faced by the operation's development outcomes over its expected remaining useful life, taking account of any risk mitigation measures already in place at the time of evaluation.[5] *Risk to Development Outcome* says nothing about the absolute level of the expected net benefits. So, for example, an operation can have a high expected rate of return and a satisfactory outcome but the R*isk to Development Outcome* may be high in its particular operating environment.

Criteria

The overall *Risk to Development Outcome* is rated by assessing both the probability and likely impact of various threats to outcomes, taking into account how these have been mitigated in the operation's design or by actions taken during its initial implementation. The evaluator should take account of the operational, sector, and country context in weighing (in each case) the relative importance of these individual criteria of risk as it may affect planned outcomes.

- Technical (e.g. where innovative technology and systems are involved);
- Financial (incl. the robustness of financial flows and financial viability);
- Economic (both at country and global level);
- Social (e.g. in terms of the strength of stakeholder support and/or mitigation of any negative social impacts);
- Political (e.g. volatility of political situation);

[5] In statistical terms, *Outcome* would be analogous to the "expected value/mean" whereas *Risk to Development Outcome* would be analogous to the "dispersion/standard deviation". Of course, in considering risk we are interested only in the downside risk – that is the possibility that the positive development outcomes, even for unsatisfactory projects, will be reduced. This is akin to a one-tailed test, which is the statistical basis for much hypothesis testing in the research literature. In that sense, the analogy is to a test that the benefits will be less than expected.

- Environmental (incl. both positive and negative impacts) ;
- Government ownership/commitment (e.g. continuation of supportive policies and any budgetary provisions);
- Other stakeholder ownership (e.g. from private sector/civil society);
- Institutional support (e.g. from project entities; and/or related to legal/legislative framework);
- Governance; and
- Natural disasters exposure

Rating Scale

Taking into account the above, the overall *Risk to Development Outcome ─ the risk at the time of evaluation that development outcomes (or expected outcomes) will not be maintained (or realized)* ─ should be rated using the following four-point scale:

(i) *Negligible to Low*
(ii) Moderate
(iii) Significant
(iv) High

Whenever the lack of sufficient information, or other circumstances, makes it impossible to assign one of the above ratings, "*Non-evaluable*" should be recorded.

3. BANK PERFORMANCE

Definition: *the extent to which services provided by the Bank ensured quality at entry of the operation and supported effective implementation through appropriate supervision (including ensuring adequate transition arrangements for regular operation of supported activities after loan/credit closing), towards the achievement of development outcomes.*

Dimensions

Bank Performance is rated by assessing two dimensions, viz., (i) Bank performance in ensuring quality at entry; and (ii) quality of Bank supervision. Quality at Entry refers to the extent to which the Bank identified, facilitated preparation of, and appraised the operation such that it was most likely to achieve planned development outcomes and was consistent with the Bank's fiduciary role. Quality of supervision refers to the extent to which the Bank proactively identified and resolved threats to the achievement of relevant development outcomes and the Bank's fiduciary role.

The lists of assessment criteria below are taken from QAG's criteria for QEA7 and QSA6, which may be seen in full here.**[Insert links]**

(1) Ensuring Quality at Entry

Criteria

Bank performance is rated against the following criteria, as applicable to a particular operation. The evaluator 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 IL operations)
- Structural, Financial and Macro-economic Aspects (for DPL operations)
- Poverty, Gender and Social Development Aspects
- Environmental Aspects
- Fiduciary Aspects
- Policy and Institutional Aspects
- Implementation Arrangements
- Monitoring and Evaluation Arrangements
- Risk Assessment
- Bank Inputs and Processes

Rating Scale

With respect to the relevant criteria that would enhance development outcomes and the 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

(2) Quality of Supervision

Criteria

Bank performance is rated against the following criteria, as applicable to a particular operation. The evaluator 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/Credit closing)

Rating Scale

With respect to relevant criteria that would enhance development outcomes and the 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. Quality at entry and quality of supervision should each be rated using their respective six-point rating scales, and for transparency, the individual ratings for quality at entry and quality of supervision should be presented separately. The quality at entry and quality of supervision ratings should be combined into a rating of overall *Bank Performance*. Ratings for the more common combinations of ratings of quality at entry and quality of supervision are provided below, followed by additional guidance on other combinations.

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.

Guidance

When the rating for one dimension is in the satisfactory range while 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, except when Bank performance did not significantly affect the particular outcome.

4. BORROWER PERFORMANCE

Definition: *the extent to which the borrower (including the government and implementing agency or agencies) ensured quality of preparation and implementation, and complied with covenants and agreements, towards the achievement of development outcomes.*

Dimensions

Borrower Performance is rated by assessing two dimensions, viz., (i) government performance (central and/or local government as relevant); and (ii) implementing agency or agencies performance. Where the government and implementing agency are indistinguishable, particularly for DPL operations, only an overall rating is necessary taking into account relevant criteria from both lists below.

(1) Government Performance

Government performance is rated against the following criteria, as applicable to a particular operation. The evaluator should take account of the operational, sector, and country context in weighing the relative importance of each criterion of Government Performance as it affected outcomes.

Criteria

- Government ownership and commitment to achieving development objectives
- Enabling environment including supportive macro, sectoral, and institutional policies (legislation, regulatory and pricing reforms etc.)
- Adequacy of beneficiary/stakeholder consultations and involvement
- Readiness for implementation, implementation arrangements and capacity, and appointment of key staff
- Timely resolution of implementation issues
- Fiduciary (financial management, governance, provision of counterpart funding, procurement, reimbursements, compliance with covenants)
- Adequacy of monitoring and evaluation arrangements, including the utilization of M&E data in decision-making and resource allocation
- Relationships and coordination with donors/ partners/stakeholders
- Adequacy of transition arrangements for regular operation of supported activities after Loan/Credit closing

Rating Scale

With respect to the relevant criteria either in identifying opportunities for, or resolving threats to, development outcomes or sustainability, rate government performance using the following scale:

Highly Satisfactory	There were **no** shortcomings in government performance
Satisfactory	There were **minor** shortcomings in government performance
Moderately Satisfactory	There were **moderate** shortcomings in government performance
Moderately Unsatisfactory	There were **significant** shortcomings in government performance
Unsatisfactory	There were **major** shortcomings in government performance
Highly Unsatisfactory	There were **severe** shortcomings in government performance

(2) Implementing Agency or Agencies' Performance

Implementing agency or agencies' performance is rated against the following criteria, as applicable to a particular operation. The evaluator should take account of the operational, sector, and country context in weighing the relative importance of each criterion of agency performance as it affected outcomes.

Criteria

- Agency commitment to achieving development objectives
- Adequacy of beneficiary/stakeholder consultations and involvement
- Readiness for implementation, implementation arrangements and appointment of key staff
- Timely resolution of implementation issues
- Fiduciary (financial management, governance, procurement, reimbursements, compliance with covenants)
- Adequacy of monitoring and evaluation arrangements, including the utilization of M&E data in decision-making and resource allocation
- Relationships and coordination with partners/stakeholders
- Adequacy of transition arrangements for regular operation of project supported activities after Loan/Credit closing

Rating Scale

With respect to the relevant criteria either in identifying opportunities for, or resolving threats to, development outcomes or sustainability, rate implementing agency performance using the following scale:

Highly Satisfactory	There were **no** shortcomings in implementing agency or agencies' performance
Satisfactory	There were **minor** shortcomings in implementing agency or agencies' performance
Moderately Satisfactory	There were **moderate** shortcomings in implementing agency or agencies' performance
Moderately Unsatisfactory	There were **significant** shortcomings in implementing agency or agencies' performance
Unsatisfactory	There were **major** shortcomings in implementing agency or agencies' performance
Highly Unsatisfactory	There were **severe** shortcomings in implementing agency or agencies' performance

Rating of Overall Borrower Performance

The rating of *Borrower Performance* is based on the ratings for each of the two dimensions (i) government performance; and (ii) implementing agency or agencies' performance. Government performance and implementing agency or agencies' performance should each be rated using their respective six-point rating scales, and for transparency, the individual ratings for them should be presented separately. The ratings for government performance and implementing agency or agencies' performance should be combined into a rating of *Borrower Performance*. Ratings for the more common combinations of the ratings for government and implementing agency performance are provided below, followed by additional guidance on other combinations.

Highly Satisfactory	Performance was rated *Highly Satisfactory* on both dimensions.
S*atisfactory*	Performance was rated *Satisfactory* on both dimensions, OR was rated *Satisfactory* on one dimension and *Highly Satisfactory* on the other dimension.
Moderately Satisfactory	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	Performance was rated *Moderately Unsatisfactory* on both dimensions. (Also see guidance below.)
Unsatisfactory	Performance was rated *Unsatisfactory* on both dimensions, OR was rated *Unsatisfactory* on one dimension and *Moderately Unsatisfactory* on the other dimension.
Highly Unsatisfactory	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.

Guidance

When the rating for one dimension is in the satisfactory range while the rating for the other dimension is in the unsatisfactory range, the rating of overall *Borrower Performance normally* depends on the *Outcome* rating. Thus, overall *Borrower Performance* is rated *Moderately Satisfactory* IF *Outcome* is rated in the satisfactory range, or *Moderately Unsatisfactory* IF *Outcome* is rated in the unsatisfactory range, except when Borrower performance did not significantly affect the particular outcome.