EVALUATION BRIEF

Evaluability - Is it relevant for the EBRD?

June 2012

EBRD EVALUATION DEPARTMENT

European Bank
for Reconstruction and Development

EvD produces occasional Evaluation Briefs. These deal with evaluation or other topical issues in a concise manner. They draw on material from a range of sources. Their purpose is to stimulate discussion.

Contents

Abbreviations

ADB	Asian Development Bank
DOTS	Development Outcome Tracking System [of the IFC]
EBRD	European Bank for Reconstruction and Development
EvD	Evaluation department
IDB	Inter-American Development Bank
IFC	International Finance Corporation [of the World Bank Group]
IFI	International financial institution
MIGA	Multilateral Investment Guarantee Agency [of the World Bank Group]

Defined terms

Evaluability	The extent to which the value generated or the expected results of a project are verifiable in a reliable and credible fashion
Impact	The positive or negative long-term effects produced by an intervention, directly or indirectly, intended or unintended; an impact generally results from a series of causal factors of which the project is but one
Indicator	Quantitative or qualitative factor or variable that provides a simple and reliable means to measure achievement, to reflect the changes connected to an intervention, or to help assess the performance of a development actor
Outcome	The short-term and medium-term effects that are a consequence of delivering the intervention's outputs
Output	The products, capital goods and services which result from an intervention—its deliverables
Quality-at-entry	A comprehensive check on all aspects of design integrity and alignment with polices and strategies—incorporates evaluability
Result	The output, outcome or impact (intended or unintended, positive or negative) of a development intervention
Results chain	The causal sequence for a development intervention that stipulates the necessary sequence to achieve desired results—beginning with inputs, moving through activities and outputs, and culminating in outcomes and impacts

Conclusions and recommendations

Conclusions

i) The concept of evaluability is highly relevant for the EBRD.

ii) While the Bank's guidance for the preparation of new operations does require that elements of evaluability be incorporated, the requirements are not sufficiently specific, comprehensive or detailed to ensure that projects would consistently be rated as evaluable or better.

iii) Practice generally lags guidance such that many of the EBRD's operations are unlikely to be rated as evaluable or better (further work would be needed to quantify this).

iv) Most other international financial institutions (IFIs) use evaluability, either on its own or as part of quality-at-entry assessments, to improve project performance and to position themselves to better "tell their story" about the results they are achieving.

v) The EBRD, unlike most other international financial institutions, does not require results frameworks for projects, strategies or policies.

vi) A greater focus on evaluability and results more generally does have structural and process implications. Most other IFIs have created units at high levels within management to be responsible for ensuring effective results management from design to delivery and beyond, and for reporting on results actually achieved. Corporate scorecards commonly include indicators derived from evaluability or quality-at-entry assessments as well as aggregate data on results actually delivered.

vii) Evaluability assessments are relatively cheap and easy to administer. Payoffs are large.

Recommendations

i) Evaluability assessments should become a routine part of the approval process for new EBRD operations with a minimum acceptable level of evaluability established. It is suggested that this start with grants (technical cooperation and so on) with a progressive roll-out to other operations.

ii) Management and EvD should work together to come up with simple checklists of evaluability for the varying types of products and services to which the tool will be applied.

iii) Consideration should be given to requiring implementation of a results framework or another means of ensuring that the EBRD's work becomes more evaluable.

1. Purpose and structure

The paper introduces the idea of evaluability, a concept widely used by IFIs to improve the performance of their operations, and their ability to report on the results achieved. Since the European Bank for Reconstruction and Development (EBRD) does not explicitly use evaluability, this brief explores the relevance of doing so. The paper aims to stimulate internal discussion on whether the concept offers value to the EBRD and whether it should be adopted.

Chapter 2 explains what evaluability is, how it is carried out and by whom, and outlines why it matters. The chapter briefly covers quality-at-entry assessment as a more comprehensive process, one that incorporates evaluability. Chapter 3 looks at how three other IFIs use the evaluability concept—the Asian Development Bank (ADB), Inter-American Development Bank (IDB) and International Finance Corporation (IFC). Chapter 4 considers current practice in the EBRD as revealed by some internal reviews and recent EvD evaluations.

2. Evaluability - why it matters

2.1 What is evaluability?

The Evaluation Cooperation Group[1] defines evaluability as "the extent to which the value generated or the expected results of a project are verifiable in a reliable and credible fashion ."[2] This is a "narrow" but useful definition of evaluability.

Others define evaluability more broadly and/or in different ways. For example, the OECD Development Assistance Committee says that through an evaluability assessment:

> "the feasibility of an evaluation is assessed…it should be determined whether or not the development intervention is adequately defined and its results verifiable, and if evaluation

[1] The Evaluation Cooperation Group brings together the heads of evaluation of the major IFIs to develop good practice standards and share experience on evaluation matters relevant to IFIs. The Evaluation department (EvD) was a founder member of the Evaluation Cooperation Group.

[2] The definition does appear to limit the use of evaluability to projects but it is equally applicable to country or sector strategies and policies.

> is the best way to answer questions posed by policymakers or stakeholders."[3]

For others, evaluability is more about the feasibility of conducting an evaluation or determining whether a programme is ready for evaluation. Morra and Rist define an evaluability assessment as:

> "a brief preliminary study undertaken to determine whether an evaluation would be useful and feasible…it may also redefine the purpose of the evaluation and methods for conducting it." [4]

Other organisations outside the IFI and development world often define evaluability in terms of whether a programme is ready for evaluation.[5]

For the purposes of this paper, the Evaluation Cooperation Group definition is used, partly because it is the one adopted by the heads of evaluation of the IFIs so is most likely to be relevant to the EBRD, but also for another important reason. Other definitions state or imply that an evaluability assessment is something carried out before the conduct of an *ex post* evaluation. While this type of assessment would be useful to help the Evaluation department (EvD) avoid wasting time and effort trying to evaluate something not capable of being evaluated in a reliable and credible way, it is too late to do anything to change the reality. The Evaluation Cooperation Group definition allows for early assessment of evaluability, even before approval so that changes can be made to improve evaluability, if this is required. This is judged to be much more useful for the EBRD since, as will be shown below, this provides direct benefits to the Bank as well as ensuring that *ex post* evaluation is more reliable, credible and useful.

Evaluability is not an absolute condition—rather, it occurs along a continuum from more to less evaluable. Some mix of the following elements is generally used to assess whether an operation is more or less evaluable:

– clarity, specificity and plausibility of the stated expected results

– existence and quality of indicators and data for assessing the achievement of results

– existence of information about the current situation in each of the expected results areas (the so-called baseline or "before project" situation)

– adequacy of risk identification and mitigation strategy

---

3 Development Assistance Committee of the Organisation Cooperation and Development. (2011). *Quality Standards for Development Evaluation*.

4 Morra Imas, L. and Rist, R. (2009), *The Road to Results: Designing and Conducting Effective Development Evaluations*, The World Bank, Washington, DC.

5 See for example Juvenile Justice Evaluation Center. (2003). *Evaluability Assessment: Examining the Readiness of a Program for Evaluation*. Available at http://www.jrsa.org/pubs/juv-justice/evaluability-assessment.pdf

– adequacy of provisions for monitoring.

Each of these elements is further described below.

2.1.1 Expression of expected results

Results is a “catch-all” term for what the EBRD states as objectives, projected financial performance, transition impact, environmental impact or social impact, additionality and investment performance.

For an operation to have a high degree of evaluability the stated expected results should meet the following conditions*:*

i) They should be clearly and unambiguously expressed so that everyone has a similar understanding of what is intended.

ii) They should be stated in terms specific enough so their achievement can be recognised and measured either quantitatively or qualitatively.

iii) There should be a clear hierarchy of results since the production of some results depends on the successful delivery of others—the most common and useful hierarchy used is that of outputs, outcomes and impacts. These results flow from the provision of inputs.[6]

iv) The expected results should be realistic and plausible—the claims made for expected results at the outcome and impact levels should be a solution to a clearly understood problem, and they should realistically and logically flow from the inputs provided and the outputs delivered. The claimed expected results should be in whole or large part attributable to the project (or at the impact level, the project should plausibly contribute to impact attainment).

v) If results are to be targeted to a specific group or market segment this should be specified in a way that allows for targeting to be measured.

2.1.2 Indicators

Indicators are yardsticks by which the achievement of expected results is measured, either quantitatively or qualitatively. For a high degree of evaluability, indicators should have the following characteristics:

i) There should be at least one indicator for each expected result.

6 For a definition of these terms see the defined terms below the list of abbreviations at the start of this report. These draw on OECD-DAC, (2002), *Glossary of Key Terms in Evaluation and Results Based Management*. OECD-DAC, Paris.

ii) Indicators should be valid—that is, they should actually measure achievement of the expected result, either directly or via a proxy indicator.

iii) They should be measurable—data should exist or be capable of generation for the indicator of success chosen.

iv) Indicators should have target levels of achievement associated with them and these targets should have achievement dates stated.

While not required for determining the evaluability of individual operations, it is highly desirable that indicators be capable of aggregation to the extent possible. Increasingly, IFIs and their shareholders want to know about overall performance in terms of aggregate or overall results. For this reason, a number of IFIs have put considerable effort into developing standard output and outcome indicators, so that aggregate results can be reported in corporate scorecards or results frameworks.

2.1.3 Baseline

Baseline levels describe the situation at approval for each of the expected results areas to provide the basis for a "before and after" assessment.[7] Each indicator should ideally have a baseline figure be it quantitative or qualitative. Credible and reliable evaluation of the performance of the EBRD's operations is difficult if the starting point is not known or is only vaguely specified.

2.1.4 Risks

Risks are those events (most importantly those outside the control of the operation, or only partly under its control), which if they occur would negatively affect the achievement of results. Risks are important for an assessment of evaluability because they can have such a dramatic effect on the achievement of results. Risk occurrence and management need to be accounted for when evaluating the performance of operations.[8] Risk identification should have the following characteristics:

[7] As well as a "before and after" assessment, a robust evaluation also requires a "with and without" [EBRD financing for example] assessment, the so-called counterfactual. However, the existence of a counterfactual has to date not generally been included in evaluability assessments although it would be desirable to do so.

[8] Without going into detail here, an evaluation of performance must take into account risk occurrence and risk management. For example, when an identified risk occurs that requires significant changes to the operation, and these are handled in a timely, efficient and effective manner, this is seen as a positive in terms of assessed performance. Conversely, the occurrence of an unforeseen risk, which should realistically have been foreseen, or a failure to address emerging risks in a timely manner, is seen as a negative in terms of project performance.

i) All the main risks should have been identified while of course accepting that not all risks can be foreseen, and that not all foreseen risks are worth identifying (a focus on those risks that are more likely and of greater potential impact is called for).

ii) The potential severity of the risks and likelihood of their occurrence should be identified.

iii) If "killer risks" exist, the project should be re-thought—killer risks are those that have a high likelihood of occurring and if they occur, would have fatal consequences for the success of the operation.

iv) Identified risks should be accompanied by a plausible mitigation strategy.

v) Responsibility and mechanisms for monitoring the occurrence and emergence of risks should be clear.

2.1.5 Monitoring

If an operation is highly evaluable in all other aspects but there is no realistic provision for monitoring and data collection during implementation it is most unlikely that monitoring will occur. Therefore, monitoring data will not be available for subsequent evaluation or to support claims of success. Characteristics required of monitoring systems for a highly evaluable operation include:

i) a clear identification of who is responsible for monitoring

ii) demonstrated availability of resources to carry out monitoring—people with the right skills and the needed financial resources

iii) identified sources of required information.

2.2 How and by whom is evaluability assessed?

Evaluability is generally assessed by means of a checklist of varying degrees of complexity using some mix of the elements of evaluability listed above. At its most basic, the checklist may only require a set of yes or no answers with an overall percentage score. The use of different weights for the various elements of evaluability can also be used. Some take the score and give a ranking to one of an even number of categories—for example, *poor, marginal, good*, or *excellent*. A cut-off point of acceptability is established. An assessment of evaluability is a fairly simple exercise that need not take too much time. However, some time and effort will be required to ensure a project is evaluable.

The assessment is generally carried out by the team responsible for the operation and is submitted as part of the management approval process for new operations. The independent evaluation department of the IFI may carry out periodic independent checks on the robustness of evaluability assessments.

2.3 Why evaluability matters

Using the evaluability concept and approving operations that are more evaluable can have the following advantages:

- Projects that are more evaluable are more likely to be successful.
- Evaluability assessment can provide an early indicator of likely success or potential problems.
- Conducting an evaluability assessment before approval provides the opportunity to make changes to increase evaluability rather than waiting to find out during implementation or even after project completion when little if anything can be done about the problem.
- Subsequent evaluation findings are more reliable and credible so providing a sounder basis for accountability.
- Lessons derived from more evaluable projects are more likely to be useful because they can focus on those aspects that contribute to, or work against, results achievement—what worked, what did not and why.
- It enables the institution to better tell its story about the results it is achieving thereby meeting its obligations to shareholders and external stakeholders and, potentially, attracting incremental resources.
- It provides a useful indicator for tracking institutional performance.
- It helps demonstrate that the IFI is fulfilling its obligations under international accords such as the Paris, Accra and Bussan agendas for action.

Of course, there are also some countervailing factors, including:

- The conduct of evaluability assessments has a cost although, as noted in section 2.2, the cost is not great if simple checklists are used.

- There can be an implicit incentive to not make operations evaluable as it may be perceived that this confers a commitment to deliver on those results for which individuals or teams will be held to account—a reality they may wish to avoid if at all possible.
- Related to the previous point, institutional incentives may not reward a results focus, or the results incentivised may be more narrowly cast in volume, value and profitability terms.
- There can be a perceived need to submit projects for approval that contain unrealistic expectations in order to get them approved.

Some commentary follows on the preceding points.

2.3.1 Relationship between evaluability and project success

There is empirical evidence to support the contention that more evaluable projects tend to be more successful. For example, World Bank reports that since 1997 its Independent Evaluation Group has rated the quality of design and implementation for around 3,500 cases. Results showed that about 64 per cent of operations with successful outcome ratings also had high ratings for design quality while only 10 per cent succeeded when they had design flaws. Interestingly, the report notes: "A similar but stronger finding is true for the quality of implementation support. More than 70 per cent of operations with satisfactory ratings had high-quality implementation support, and only 4 per cent succeeded when evaluation found weaknesses in implementation."[9] Implementation support no doubt includes effective monitoring to identify emerging problems and effective risk management and mitigation, both provided for in evaluability assessments.

Figure 1: Relationship between development outcomes and work quality at IFC

> "In IFC we have measured the statistical relationship between Development Outcomes and overall work quality (covering 905 operations), as well as the three individual components of work quality (screening, appraisal and structuring – upfront work quality; supervision and administration; and, role and contribution – IFC additionality).
>
> The latest results show that all three components of IFC work quality are statistically significant at the 5 per cent level. Also, based on the coefficient values, we can conclude that: role and contribution (additionality); screening, appraisal and structuring (quality-at-entry); and supervision and administration, in that order, are important factors in determining success.

[9] World Bank, 2011, *World Bank for Results*, World Bank, Washington, DC.

> In terms of the channels of effect, broadly speaking, we believe that IFC work quality-at-entry is a critical instrument for mitigating external risk factors such as sponsor risk, market risk, business climate, and project type (greenfield versus expansion). For instance, in projects with high sponsor risks, we expect a thorough screening, appraisal and structuring work quality to compensate for these high risks. Such projects also may require high supervision and administration work quality."
>
> Marvin Taylor-Dormond
> Director, IEG Private Sector Evaluation, World Bank Group
> Comment provided for this Evaluability Brief

There are sound reasons why superior evaluability (or quality-at-entry, of which evaluability is a part—see section 2.4) is a factor that contributes to better project performance including:

i) Many of the elements contributing to superior evaluability come from sound diagnostic work and understanding of the problem(s) to be addressed, which result in better-designed projects that are more likely to produce expected results.

ii) The aphorism "what gets measured gets managed" also helps explain why projects that are more evaluable may be more successful—tracking progress towards results achievement through monitoring and staying in touch as to whether potential risks are eventuating can give early warning of emerging problems to allow timely corrective action to be taken.

iii) Another reason for the relationship is that projects that achieve high evaluability scores may have more realistic expected results. The EBRD and other IFIs primarily evaluate performance against results expected at approval. The question being answered by evaluation is *did we achieve what we set out to achieve?* Of course, evaluation also considers unexpected results, whether positive or negative, but the main focus is on the achievement of expected results. Since performance is mostly assessed against expected results, more realistic results should lead ultimately to more projects being rated as successful.

iv) Also, projects which are more evaluable are more capable of being rated successful. If it is not clear what an operation is setting out to achieve, it will be difficult to assign a performance rating so the tendency might be for a lower rating to be given than would have been the case with a better specification of expected results. It is hard to rate an operation successful or highly successful if it is not clear what it was setting out to achieve or if basic performance data are lacking.

2.3.2 More reliable and credible evaluations

If it is not clear what results projects are setting out to achieve, evaluation will not be able to fully determine what worked, what did not, and why. Lessons will be hard to derive and learning diminished.

Also, if projects are not evaluable, performance assessments by evaluation will probably be less reliable and credible thereby diminishing their utility for accountability purposes.

2.3.3 Telling the story of results

Importantly, it will be hard for an organisation to "tell its story" of what it has achieved if actual results cannot be verified in a reliable and credible fashion. It is much more persuasive to have credible evidence of how IFI involvement produced tangible benefits than it is to talk about the volume of lending and extent of repayment. Most IFIs now report on achieved results in corporate scorecards and via annual development effectiveness reports (see Chapter 3). They also produce a range of publications highlighting results achieved at the country and sector levels. The response to these initiatives has been very positive.

2.4 What is quality-at-entry and how does it differ from evaluability?

A number of IFIs conduct quality-at-entry rather than evaluability assessments. Evaluability forms part of quality-at-entry but the latter is a much more comprehensive process that looks at all aspects of operation design, due diligence, alignment with policies and strategies, and so on. Evaluability asks the question "can this operation be evaluated?" Quality-at-entry, on the other hand, asks the questions "are we doing the right things and are we doing them right?" The Quality Assurance Group of the World Bank was responsible for assessing the quality-at-entry of recently approved operations. It did so along eight major dimensions:

– strategic relevance and approach
– technical, financial and economic aspects
– poverty and social aspects
– environmental aspects
– fiduciary aspects
– policy and institutional aspects

– implementation arrangements

– risk assessment and management.

To this list one should add the extent to which past experience has been adequately identified, and the extent to which this has influenced choices made in the design of the new operation.

The point is that quality-at-entry is an altogether more complex, costly and time-consuming exercise. Nonetheless, a number of IFIs conduct regular quality-at-entry assessments as noted in the following chapter.

3. How other IFIs use evaluability

3.1 Asian Development Bank

ADB uses the concepts of evaluability as part of its comprehensive managing-for-development-results framework, which has evolved since the introduction of the logical framework in 1995. The managing-for-development-results framework is seen as a way to help ADB "apply its limited resources in a way that delivers the best possible development results [by providing]...the process and tools needed to deliver optimal results and ensure transparency and accountability."[10]

Figure 2: Quality-at-entry at ADB

> "At ADB, we undertake an assessment of quality-at-entry every two years. Since we instituted this in 2006, it has gained institutional support and is seen as an integral tool in helping us set quality standards, monitor trends and improve our operations. The quality-at-entry exercise, forces us to take a step back and ask ourselves, did we do the right things to begin with? Were the objectives worthwhile? Were the rewards commensurate with risks? Did we articulate the underlying development rationale and monitoring framework clearly?
>
> Benefiting from the outcome of the assessment, we have introduced amended processes in the preparation of our Country Partnership Strategies; identified specific improvements on impact and outcome in our design monitoring framework for sovereign projects; and identified the need to do further work on feasibility analysis, risk assessment and management for non-sovereign projects. Equally important, quality-at-entry is an indicator of operational efficiency together with portfolio performance for our corporate results assessment which is presented to our Governors at our Annual Meeting. Overall, quality-at-entry has been a worthwhile exercise for measuring and improving ourselves."

[10] ADB, 2011. *Driving Results at ADB*. ADB, Manila. Available at http://www.adb.org/sites/default/files/driving-results-at-adb.pdf.

Rajat Nag, Managing Director-General, Asian Development Bank
Comment provided for this Evaluability Brief

In 2004, ADB launched its first corporate results framework with a performance scorecard introduced in 2009. Refined in 2011, the results framework now covers four levels of results:

- Asia and Pacific Development Outcomes
    - core Outputs and Outcomes—a set of standard output and outcome indicators covering education, energy, finance, transport and water
- operational effectiveness
- organisational effectiveness.

Reporting against the results framework is the basis for an annual Development Effectiveness Review prepared by a Results Management Unit located within the Strategy and Policy department, which reports to the President. The Development Effectiveness Review is always considered by the Development Effectiveness Committee of the Board of Directors (to whom ADB's Independent Evaluation department reports in the first instance). Now in its fifth year, the Development Effectiveness Review is stated as being a "*corporate performance assessment. The review evaluated ADB's progress against the Strategy 2020 framework, and highlighted performance trends and needed Management actions*." This quote highlights the fact that ADB is using results measurement as a management tool.[11]

In the results framework under operational effectiveness, there are three indicators that draw on quality-at-entry assessments; the percentage of country strategies, sovereign-guaranteed and non-sovereign/private sector operations rated satisfactory or better. For example, the 2011 Development Effectiveness Review reported that for the indicator of quality-at-entry of non-sovereign operations ADB was "off track" with only 57 per cent rated satisfactory versus a target of 85 per cent but with an improving trend.

Quality-at-entry assessments are carried out every two years by consultants overseen by an inter-departmental panel. The assessment covers 100 per cent of approved operations (sovereign and non-sovereign, and technical assistance) and country strategies. This is a management exercise although each time it has been conducted management requested that the Independent Evaluation department provide the chair of the inter-departmental panel with a greater degree of neutrality.

[11] ADB, 2012. *Development Effectiveness Review: 2011 Report*. ADB, Manila. Available at http://www.adb.org/sites/default/files/defr-2011.pdf.

At the level of individual operations, all operations (including non-sovereign/private sector ones) must have a design and monitoring framework, which is ADB's version of the logical framework. This contains almost all the elements required for assuring evaluability (the framework itself does not contain details of risk mitigation or details of monitoring arrangements [aside from the identification of data sources] although these other aspects are expected to be included in other parts of the project documentation). Managers and department heads attest to the quality of these results frameworks as part of the due diligence process.

Over many years, ADB has put a lot of effort into improving the quality of design and monitoring frameworks by providing training, preparing guidelines, requiring high-level "sign-off" on quality, and regularly independently checking quality. Initially significantly led by the Independent Evaluation department, this is now entirely a management responsibility with only periodic checks carried out by the Evaluation department. A small unit within the Central Operations Services Office is responsible for ensuring the integrity of the design and monitoring framework process, supported by staff embedded in each operational department. A checklist for design and monitoring framework quality is used.

Since 2006, all ADB policy and strategy papers must have results frameworks although the evaluability of these is not systematically assessed.

3.2 Inter-American-Development Bank

In 2008, IDB approved a comprehensive Development Effectiveness Framework. This mandates that a project design has to be evaluable *ex ante*—specifically it states "...if an operation is scheduled for discussion at the Operations Policy Committee, the Office of Strategic Planning and Development will provide its technical opinion on the evaluability of the operation at that time." New country strategies and operations must have satisfactory scores in evaluability dimensions.

Like ADB, IDB has a corporate results framework. The results framework has four levels:

- lending programme estimates
- regional development goals
- bank output contribution to regional development goals
- operational effectiveness and efficiency.

Evaluability plays a significant role in the last level, namely:

- percentage of country strategies with satisfactory scores in evaluability dimensions
- percentage of country strategies that have satisfactory results that can be validated at completion for sector outcomes, financial outcomes and progress on building and using country systems
- percentage of both sovereign and non-sovereign operations with satisfactory scores on evaluability dimensions
- percentage of completed knowledge and capacity building products with results that can be validated.

The Development Effectiveness Overview is "IDB's annual corporate report that accounts for the effectiveness of its work, stating the results achieved with the implementation of the Bank's Development Effectiveness Agenda." It reports against the four levels of results in the results framework. With regards to evaluability, the 2011 Development Effectiveness Overview recorded that from a baseline of only 27 per cent of country strategies with satisfactory scores in evaluability dimensions in 2006–09, 100 per cent were so rated in 2011. The report also indicated that 100 per cent of non-sovereign operations also achieved a satisfactory rating on evaluability dimensions.[12]

All sovereign and non-sovereign projects, country strategies and knowledge and capacity building products must have a development effectiveness or country results matrix. It is assessed using a checklist of questions that elicit yes/no answers. Lending projects must achieve at least five out of a possible 10 points on an equally weighted evaluability score to be approved (there are different development effectiveness matrices for sovereign and non-sovereign operations). The checklist is included for every project that goes to the Board for approval. The evaluability score is validated by the IDB's independent evaluation department, the Office of Evaluation and Oversight.

There are four elements that make up the evaluability score.[13] These are:

- Does the project have a clear logic supported by analytical underpinnings?
- Is there a cost-benefit analysis or cost-effectiveness analysis?
- Does the project have specific methodologies and metrics for monitoring and evaluating change?
- Have the risks of the operation been identified? If so, does the

[12] IDB, 2012. *Development Effectiveness Overview 2011*. IDB, Washington, DC.
[13] Evaluability in IDB goes beyond the Evaluation Cooperation Group's more narrow definition to include elements of quality-at-entry.

project have metrics for monitoring implementation of mitigation measures?

For sovereign projects, the development effectiveness matrix also considers the following:

- Is the project aligned to IDB's strategic objectives (aligned/not aligned)?
- Does the project contribute to IDB's country level development objectives (aligned/not aligned)?
- Does the project generate indirect positive improvements in management standards to the public sector entity (additionality)?

The strong role given to evaluability at IDB is a direct consequence of evaluability assessments carried out by the Office of Evaluation and Oversight (IDB's independent evaluation department, the equivalent of EvD) in 2001 and 2005.

3.3 International Finance Corporation, World Bank Group

IFC has a strong results focus. At the project level, aspects of evaluability are deeply integrated into IFC's approval process for new operations although it does not use the term. Use of a Development Outcome Tracking System (DOTS) is now well embedded, having been introduced in 2005 for investment operations and later for advisory services. This is designed to ensure that investment projects and advisory services are evaluable and can be monitored among other things.

Entry into DOTS is mandatory for all new projects. The operational departments are responsible and accountable for selecting the relevant indicators at project appraisal from a set of standard indicators that aim to capture typical development impacts in a manner which allows aggregation, and then for monitoring them throughout implementation. Baselines and targets must also be set. As stated in IFC's 2011 Annual Report, DOTS "gives IFC a key competitive advantage, because it is critical to understanding how well our strategy is working and whether we are reaching the people and industries that most need our help." [14] The report goes on to note "IFC also introduced mechanisms to link incentives to project results through performance awards."

Quality checks on the project design and its evaluability are required during project due diligence. In the case of investment operations, this is first done by the operational department, and then by the Development Impact department (see Figure 3). The World Bank Group's Independent

14 IFC, 2011, *IFC Annual Report 2011*, IFC, Washington, DC.

Evaluation department (now encompassing the formerly separate evaluation departments of IFC, the Multilateral Investment Guarantee Agency and the World Bank) assesses quality-at-entry of IFC operations based on completion reports (see Figure 3).

IFC has a scorecard to assess its performance against its strategic priorities. The first two indicators are the percentage of investment projects and advisory services that are rated “high” by the development objectives tracking system. IFC presents its performance against this scorecard in its annual report.

Figure 3: Comment on quality-at-entry at IFC

> “For investments, our quality control has ensured that all projects had appropriate monitoring and evaluation entries (for example, including baselines, targets, timelines, and the use of standard indicators). This initially enabled IFC to report on the expected results of new business, which we have been doing since fiscal year 2007 for investments, and which was clearly appreciated by management. This has also allowed us to subsequently introduce IFC's Development Goals, which also require completeness and high quality of entries. The IFC Development Goals would not have been possible without the discipline of clear, standardised results frameworks with baselines and targets – and quality control.
>
> For advisory services, one of the problems we faced at one point was that objectives were either ill-defined or not realistic, and the monitoring and evaluation framework did not support the project design. This led to problems at the time of evaluation, where the apparent "failure" of problem projects was at least in part attributable to a poor ex ante results framework. Management asked us to help improve the quality of at-approval documents, in order to reduce this shortfall, and more recently specifically asked for help in using monitoring and evaluation for improvements in project design.”
>
> Nigel Twose, Director Development Impact Department, IFC
> Comment provided for this Evaluation Brief

The experience of another member of the World Bank, the Multilateral Investment Guarantee Agency (MIGA) is also highly relevant. This is summarised in Figure 4.

Figure 4: Experience of using quality-at-entry at MIGA

> “At MIGA we have conducted four quality-at-entry assessments starting in fiscal year 2003 as part of Independent Evaluation Group-MIGA's Annual Reports. The quality-at-entry assessments served three purposes: (i) to supplement the small sample of ex post evaluations; (ii) to help improve MIGA staff's assessment of project development impact; and (iii) to accommodate MIGA management's interest in the Independent Evaluation Group’s assessment of more recent projects rather than the mature projects selected for ex post evaluation.

A formal quality-at-entry guideline was adopted in 2005. It mirrors the dimensions and indicators in the Project Evaluation Review methodology and does not review the specific evaluation objectives. A streamlined version of the quality-at-entry questionnaire was also developed based on the guideline and this was presented to the Committee on Development Effectiveness as part of the fiscal year 2007 Annual Report.

We have not yet conducted a formal analysis of the relationship between quality-at-entry review and the Project Evaluation Review results. However, we have observed a striking pattern of early cancellations of MIGA contracts involving projects rated low for quality-at-entry and in a few cases, with investor and host government dispute or even pre-claim. This pattern hindered the Independent Evaluation Group's evaluation of these projects' actual development results because before 2008, cancelled projects were excluded from the Project Evaluation Review sampling framework and projects with disputes/pre-claim were dropped from the final sample for evaluation.

With respect to work quality, a dimension assessed for all completed Project Evaluation Reviews, we have not yet conducted a statistical analysis of the relationship between these indicators and development outcomes. The reason is that the sample size was, up to last year, too small. However, with new observations, by next fiscal year we will be in a position to conduct such analysis."

Marvin Taylor-Dormond
Director, IEG Private Sector Evaluation, World Bank Group
Comment provided for this Evaluability Brief

4. Current practice at the EBRD

There are various strands of evidence to support the view that the EBRD's operations and strategies are not particularly evaluable using the standards outlined in Chapter 2. A part of this evidence is cited in the following sections—this is not a comprehensive assessment of the *status quo* but it does give an indication of the existence of a number of issues surrounding the evaluability of the EBRD's operations.[15]

4.1 Structure of Final Review Memorandum

A review of the structure and content of the Final Review Memorandum (and hence information going to the Board for approval of the operation) shows that a number of the elements of evaluability (as outlined in

15 This chapter is not presented as a comprehensive review of the current situation on evaluability of the EBRD's operations. Rather, it provides some strands of evidence that are considered indicative of the situation. Further work is planned by EvD, including a study in the 2012 work programme; *Performance Metrics in Selected Projects*.

section 2.1 above) are covered by guidance given in the Operations Manual. However, the following observations can be made on the extent and nature of coverage:

Final Review Memorandum Guidance	Comment
Expected results are described through the use of a variety of terms including: business purpose, objectives, purpose of technical cooperation, conditionalities, covenants, transition impact, target groups/market segment, environmental objectives, projected profitability of the Bank, and projected financial performance.	There is no consistent terminology or guidance given for describing results. There is no requirement to establish a hierarchy of results such as outputs, outcomes and impacts.
There is a requirement to ensure that the expected results are achievable— "the description should not oversell the operation and [it] should be consistent with the ratings on transition impact potential and risk…where practicable, it should take the form of checkable stories which allow subsequent monitoring of success. Thus, avoid asserting project benefits which are too ill-defined or general to allow assessment of their achievement at a later date by the Bank's Post-Evaluation department."	EvD's observation is that this provision is frequently not heeded to the extent intended as shown in the sections that follow.
A summary of expected results should be "provided in a table summarising key objectives which emerge from the above sections, and indicating what specific parameters or benchmarks will be used to evaluate the Bank's success in meeting these. While the financial success of the project will be an important indicator…there will usually be other indicators to confirm whether a checkable story relevant to transition has been fulfilled" but no guidance given on what an acceptable summary comprises or how indicators should be chosen and what constitutes a good indicator. There is no requirement to establish baseline levels, targets or timeframes for the indicators identified.	Again, EvD's experience is that practice is not fully aligned to this guidance. Although the section containing the quote is entitled "Measuring/ Monitoring Success" there is no requirement to establish the basis for monitoring—responsibility, resources and data sources.

Final Review Memorandum Guidance	Comment
Risks are required to be dealt with the Operations Manual stating a requirement to: "demonstrate to Operations Committee and Board that the risks associated with the operation have been addressed and why risks which are not mitigated are acceptable to the Bank." There is also a requirement to demonstrate the likelihood of each risk occurring and the extent of the EBRD's control over the events giving rise to the risk.	The treatment of risks has not been comprehensively assessed by EvD.
In the guidance on additionality, the Operations Manual advises: "each section should be strengthened by a description of the counterfactual of whether the project, or how a different project, would proceed in the absence of the Bank."	EvD observes this is almost never done.

Although evaluability is dealt with to some extent in guidance material, the extent to which this flows through into practice is variable and generally low as evidenced by the work described below.

4.2 EvD evaluation of the Early Transition Countries Initiative

An EvD evaluation on the Early Transition Countries Initiative (forthcoming) looked specifically at project design and the issue of evaluability. The relevant findings, which were based on 10 case studies, include:

– There is a lack of an overarching project design methodology for integrating the different project objectives—there is no apparent hierarchy of results.

– There is a lack of consistency and often a misunderstanding in the difference between impacts and outcomes, and between outcomes and outputs.

– Baseline data are not systematically presented.

– Policy dialogue and technical cooperation are very often not effectively connected to investments, nor accompanied by effective performance metrics and monitoring plans.

Figure 5: Conclusions of these findings as identified in the EvD Special Study

> "On the basis of these findings the study suggests that the effectiveness of the initiative could be increased by: strengthening the project design and review process to ensure improved "quality-at-entry" in the form of baseline metrics, greater clarity in measurable criteria for performance impact, and a more systematic focus on transition performance monitoring and reporting."
>
> Special Study
> Early Transition Countries Initiative
> EvD Work Programme 2011

4.3 EvD synthesis of findings on a decade of evaluations of technical cooperation

An EvD synthesis paper (forthcoming) summarises the findings of the 60 evaluation reports on technical cooperation prepared by EvD between 2000 and 2010. These 60 reports covered 309 individual technical cooperation operations with a total value of €123.56 million. The findings with respect to evaluability include:

– The great majority of evaluation findings (about 92 per cent) concern process, design and delivery issues relating mainly to the instrument itself rather than to specific sectors, themes or countries of operations.

– Findings related to specific sectors, such as small and medium-sized enterprises, transport, energy, or municipal infrastructure, comprise only about 8 per cent of the total (namely, 20 findings).

– Sixty-five findings (around 25 per cent) concern shortcomings in the specification of outcomes, outputs, and related indicators.

– Findings and lessons stress the importance of clearly spelling out expected results at the design stage in order to establish the observable benchmarks and targets needed for effective project monitoring and supervision.

– Operations built on a reasonably clear model of cause and effect are much more likely to produce the intended results than those lacking this rigour, not least because by design they have identified critical performance indicators that support effective execution.

– More rigorously designed technical cooperation operations appear to be more generally successful than those that are not.

– There is no specific requirement for the identification of the objectives of the technical cooperation in submission templates and it is not possible to distinguish a hierarchy of inputs and results (impacts, outcomes, outputs).

– Success indicators are not required to be specific, measurable, achievable, relevant and time-bound (so-called SMART indicators).

Figure 6: Importance of adoption of a results-based management approach

> "The adoption of a results-based management approach, as already identified by the ongoing Grant Co-financing Strategic Review, will be crucial in tackling the most frequent findings related to the Bank's TC management, from design to reporting, and ultimately will serve to enhance the Bank's performance and the perception of this. As a condition of approval, technical cooperation project submissions should include clear and measurable success indicators. Baseline data should be provided. Targets should be specified for each indicator, as well timelines for expected accomplishments."
>
> EvD Synthesis Paper
> TC evaluations – Insights and Findings
> EvD Work Programme 2011

4.4 Grant Co-financing Strategic Review

The interim report of the review had managing for results as one of its focus areas.[16] The report findings on results were that:

i) [Current processes] deliver products of variable quality, which provide partial insights into different aspects of the effectiveness of these operations.

ii) The transition impact monitoring system can only offer a very partial solution when it comes to technical cooperation, as it only captures a portion of investment-related technical cooperation, and even then it only includes a subset of the objectives and indicators that are needed to effectively manage a technical cooperation operation.

iii) Enhancements are necessary to provide reporting that more closely meets donor expectations and demands, with a focus on outcomes and indicators that can be analysed across strategic areas. Such reporting will also help operational teams to understand better the effectiveness of technical cooperation and investment grant

16 EBRD, 2012, *Grant Co-financing Strategic Review: Interim Report*, CS/BU/12-10 (Final).

products (what works, what does not and why), thus contributing to improvements in the Bank's ability to deliver transition impact. This will require a more systematic structuring of output and outcome indicators and risks in technical cooperation submissions coupled with improvements to reporting on these…Narrative reporting on technical cooperation operations more generally also needs to be strengthened, enabling the Bank to better "tell the story" of the results being produced by technical cooperation and investment grants.

In light of the identified issues the report concluded that:

i) With respect to managing for results, clearer objectives will be required at the design stage of technical cooperation proposals. These will include output and outcome indicators, and risks. A compendium of good practice indicators is being established to support this. Work on this is at an advanced stage and will be piloted during 2012 for comprehensive roll out from January 2013.

ii) Monitoring against these indicators will extend beyond contract end to capture more outcomes. Reporting will where possible be integrated with the reporting on investments. It will focus on achievement of results, allowing the Bank to better "tell the story" of what grants and technical cooperation are achieving. The new approach will enable enhanced "aggregated" reporting, for example to the Board, as well as at the level of individual assignments. While there is clearly a time lag between introduction of the new project design tools and the reporting that will follow, management will make every effort also to strengthen reporting during the transition period.

4.5 The findings of the Besley Report

An expert panel was convened in 2009-10 to look into the concept of transition impact at the EBRD. Its report (the so-called Beseley Report)[17] drew a number of conclusions relevant to the issue of the evaluability of EBRD's operations, including the following:

– No measurement framework is at hand that would allow an assessment of transition progress.

– There is scope for further work on rethinking the conceptual basis of the core transition indicators as a framework for monitoring broad goals.

– Existing transition indicators seem poorly integrated into the

[17] Besley, T. Dewatripont, M. and Guriev, S. 2010, *Transition and Transition Impact: A Review of the Concept and Implications for the EBRD.* http://www.ebrd.com/pages/research/publications/special/besley.shtml.

process of assessing transition impact at the project level.

– Transition impact assessment misses out an explicit measurement of the institutional preconditions for development, legitimacy and resilience of market-supporting institutions.

– The current framework seems less robust on how transition impact is to be assessed in more contentious areas such as public utilities.

The report concluded by saying:

> "we recommend that the EBRD collate[s] available measures of performance [including] those created by…other organisations. This could be used as a means of assessing how such measures could be put alongside existing transition indicators and used in…monitoring of transition progress."

An examination of the Final Review Memorandum structure and guidance, together with the evidence provided by recent management reviews and EvD studies all point to a reality that the EBRD's operations lack evaluability and that this has negative consequences for the Bank. An opportunity exists to address this problem. Evaluability assessment is a means of doing so.