

Review of Results Frameworks from Donor Agencies

Synthesis Note

The purpose of this paper is to provide a short summary review of how Multilateral Development Banks (MDBs) and selected bilateral donor organisations present their results (the form) and the focus and types of indicators typically used at the different tiers of their Results Frameworks (the substance). Reviewed organizations are AfDB, ADB, World Bank, IFAD, IDB, Australia, United Kingdom, Denmark and Netherlands. Of the bilateral agencies reviewed only Australia and DFID operate with multi-tiered results frameworks with targets and indicators similar to the multilaterals.

These papers are intended to help ensure a high level of knowledge about comparable frameworks and to identify good international practices.

In terms of the context for writing this paper, it is important to recognize that a significant aspect of the international partnership on aid effectiveness is a focus on results, and by extension the cost to and difficulty for development partners to measure, track and monitor results. UNDP clearly shares the commitment to reducing the cost to developing countries of producing data to feed the reporting needs of various institutions. Therefore, better knowledge of other institutions' indicators may help the harmonization efforts. In this respect it is worth noting that the MDBs are converging toward a common MDB methodology that tracks progress along the entire results chain, from agency institutional inputs to high-level country outcomes. And donors are under the heading "Multilateral effectiveness" reflecting on ways to measure the results of their multilateral cooperation.

The following is structured with a section on terminology followed by a section summarizing findings on Reporting and Presentation. Thereafter, individual sections summarize key findings for each of the four tiers, their focus and key indicators. Lastly, ten Good Practice Highlights are presented.

A Word on Terminology

While everybody would agree that a results chain describes a causal relationship from input to impact, deciding where exactly in this chain an output or an outcome starts and ends is, however, a different matter. Inconsistent use of the terms output and outcome with respect to what degree of change is included confuses the understanding and makes comparison across peers difficult. At the output level, for example, some would include completed activities as an output (people trained), while at other times and for other agencies the term output is used about the *change*

that results from the completion of activities (new skills or abilities). In other cases, it is the exact level of change that is required for an output to become an outcome, that draws the line between the two.

Some of the confusion may be ascribed to the definition of Outputs (in English) provided in the OECD/DAC glossary, which has two elements: “The products, capital goods and services which result from a development intervention; may also include changes resulting from the intervention which are relevant to the achievement of outcomes.” It is the “changes” in the latter part of the definition that blurs the distinction between products or completed activities on the one hand, and results on the other.

Many of the bilateral donors reflect this same ambiguity in how they define Outputs, some specifying them as just completed activities or products, while others include the additional possibility of near-term change. Irrespective of the definition, however, many bilaterals report both on output and on outcome. Output is recognized as a useful measure, easily communicated to the public, but also recognized as insufficient to constitute results reporting in results based management terms.

Outcome

Outcomes represent changes in the institutional and behavioral capacities for development conditions that occur between the completion of outputs and the achievement of goals.

Outputs

Outputs are changes in skills or abilities and capacities of individuals or institutions, or the availability of new products and services that result from the completion of activities within a development intervention within the control of the organization. *Outputs are the level of result in which the clear comparative advantages of individual agencies emerge and accountability is clearest.*

UNDG RBM Handbook 2011

Reporting and Presentation

In terms of reporting and presentation, most MDBs produce both a short 5-8 page “scorecard” and a more comprehensive annual results or effectiveness report, and aim to answer the following four questions in one form or other:

- What is the development progress in the region covered/what is the global context?
- What are the organization’s key operational outputs and intermediate outcomes/how is it supporting countries or contributing to country outcomes?
- Is the organization managing its operations effectively?
- Is it managing itself efficiently?

These questions are answered through narratives and a range of indicators, typically presented in a three to five tier framework with explicit targets and using colour-coding to show progress.

Often references are made to additional documents, where results chains and causality is documented, such as the briefs that both ADB and AfDB produce.

Common for all reviewed organizations is that they state that they have comprehensive data and information systems that “feed” the Results Framework but that they are selective in what they report on, and that the Scorecard and the Report are *representative* of results and performance but do *not cover the full range of results at all levels*. No organization specifically state the share of results reported in the framework.

Targets or baseline values are used by all organizations in some, but not all, of the tiers and most organizations use color-coding to indicate progress. The World Bank does not use colour coded traffic lights for Tier 1 as this tier provide overall development context, and no ratings are assigned to individual indicators in Tier 2 as results are country results. The number of indicators varies from 109 (AfDB) to 57 (IDB).

There is general recognition among all organizations that targets and indicators need to be reviewed regularly and will change over time as experience with the framework grows and as more data become available

All MDBs, and bilaterals, struggle with aggregation issues and most have developed standard indicators and/or some kind of organization specific ratings or surveys to facilitate aggregation, often based on project data, and project completion reports. Danida for example simply report on the aggregated success-ratings of hundreds of Programme-level sub-goals. The MDBs are trying to harmonize the use of standard core sector indicators through COMPAS, the Common Performance Assessment System.¹

Key Findings on Tier 1

Tier 1 typically aims to answer the following question: ***“What is the global context?”*** Or, for organizations with a regional mandate: ***“What is the development progress in the region?”***

The question is about the context in which each organization works and is usually answered through a range of indicators and a more analytical narrative, sometimes highlighting specific country performances and providing explanations for, and more nuanced background to, the data. All organizations explicitly acknowledge that this level of results clearly cannot be attributed to any one organization alone.

MDG or similar globally agreed goals and targets are used by most organizations, bilateral and multilateral. While Australia and DFID use all the MDGs for Tier 1, none of the multilateral organizations include all MDGs in their results framework, but have a range of specific indicators that relate closely their core areas of focus or strategic goals.

¹ <http://www.mfdr.org/Compas/index.html>

The number of indicators ranges from seven (IFAD) to 28 (WB) and several use different indexes (Competitiveness) and scoring systems (Governance, CPIA, EITI) to capture complex result areas. Some MDBs provide data broken down by country typology, i.e. the Banks provide data separately for the countries with access to concessional resources.

In terms of targets, practices vary. Some MDBs track Tier 1 indicators against targets or progress against a baseline using a traffic light method². Others, like the World Bank and IDB, state that Tier 1 is seen as a context to the bank's performance rather than goals to be achieved and therefore no colour-coding is assigned.

Practices also vary in terms of how change is tracked. Some MDBs track relative performance (of other regions), some track direction of change, i.e. improving or deteriorating against a target or a baseline, and some track both. Colour-coding is used by most of the organizations as an easy to understand way to show performance, as is some way (arrows) to show direction of change.

Key Findings on Tier 2

Tier 2 typically aims to track core output and outcomes in different ways and answer the question: ***“What is the contribution to development results?”***

The complex questions of output and outcomes, the link from country to corporate results, and the added value/comparative advantage are addressed in this tier, to different degrees. It is the tier with biggest differences among the organizations in terms of indicators chosen, and the methodological reflection behind. It is also at this level that terminology confusion is greatest.

No framework aim to capture all aspect of an organisation's work, so most often strategic areas are identified and measured through “core” indicators. DFID for example explicitly states that “DFID indicators reflect those outputs where it is possible to aggregate results across different countries. They do not reflect all the results that DFID is delivering. Results that are vital to each country's development may not be covered simply because they cannot be aggregated across countries.”

Some key methodological issues discussed in the reports or methodology papers include:

The demand driven nature of the organizations' work: Some organizations (IFAD, World Bank, IDB) stress the demand driven nature of projects, and as a consequence that targets cannot be established at this level.

Explaining the results chain: Scorecards are too short to show and explain the full results chain, and Tier 2 often includes a mix of output and outcomes. Therefore showing the causality between the different tiers is done either in separate documents or through an analytical narrative that

² The World Bank explicitly does not use traffic-light coding and targets as they see Tier 1 as providing context rather than goals.

accompanies the scorecard indicators, explaining performance and discussing the links between tiers.

Aggregating output and outcomes: Several organizations (AfDB, ADB, IFAD) use Project Completion Reports (PCR) as the foundations for the results reporting system. PCRs are however produced upon project completion, and as operations take several years to complete, many development outputs are produced at different stages of their implementation, meaning that a significant portion of the results achieved in any one year, is not covered by annual corporate reporting. Another method to facilitate aggregation is use of Sector Indicators (World Bank, AfDB, ADB), which also means that annual reporting covers both on-going and closed operations, providing a more accurate picture of results at any point in time.

Measuring Capacity Development: This is stressed by several organizations as particularly difficult, especially in an aggregate form across countries. Both multilaterals (World Bank) and bilaterals (DFID) chose to use as an indicator, numbers of countries where they have supported key institutions, such as procurement system reforms, or strengthening of improved management of government budgets and public financial management systems.

Setting Targets: Targets are used by most organizations, bilateral and multilateral. While Australia and the United Kingdom use output/access targets, the multilateral organisations tend to mix outcome and output/access targets. Some use output and beneficiaries as proxy for outcomes (ADB), and several use sector outcomes. The World Bank, IFAD and IDB state that results at this level are demand driven, and therefore targets are not established.

Country typology: several of the MDBs report indicators separately for their concessional window.

Tracking change: Practices vary in terms of how change is tracked. Some organizations track performance in meeting targets, some track change vis a vis baselines. Colour-coding is used by several of the peer organizations, as is some way (arrows) to show direction of change. The number of indicators ranges from 22 in World Bank scorecard to 51 in AfDB, and tend to be quantitative for the MDBs. Bilaterals have both quantitative and qualitative output/access targets or indicators, often linked to an outcome, although the latter tend to be at a very general level.

Several organisations (World Bank, ADB, AfDB, DFID) stress that they continue to work with their results frameworks.

Key Findings on Tier 3

The question that Tier 3 tries to answer is: *“How effectively are operations managed?”*

The question, and how to measure it, has in many MDBs been the subject of discussion in the governing bodies and all peers measure and report on effectiveness in one form or another.

Measures often relate to the health and performance of the portfolio, and to other aspects of aid effectiveness, for example related to Paris Declaration commitments. The performance of operations is measured at different stages, includes quality at entry, projects at risk, disbursement rates, or project completion. Some peers also measure in this Tier, how well operational activities are focussed. Numbers of indicators vary from 24 in the World Bank to 17 in IDB.

Much results tracking and reporting in this tier rests on program/project level M&E systems; almost all peers aggregate scores measuring degree of success in achieving stated project or programme objectives, sometimes broken down on country typology. DFID and Danida³ both track success rates in terms of achievement of objectives as a key results metric.

In terms of targets, these have generally been agreed in the organization's governing bodies and are often monitored on a regular basis, with frequent reporting to management; the scorecard or annual report presents a selected number of key indicators, extracted from a much wider set of metrics.

Aggregation is facilitated in this tier by ratings or scores in some form or another, often satisfaction ratings for various dimensions of portfolio performance. Perception surveys are also commonly used to monitor and report partnerships or knowledge management. The Paris Declaration commitments still constitute a key reference point for aid effectiveness, and several MDBs and bilaterals used Paris Declaration indicators, or variants thereof, to measure effectiveness.

Qualitative aspects of effectiveness such as knowledge management, gender, and environment are captured in different ways. Knowledge management and partnerships through different kinds of satisfaction surveys, and gender and environment often through project-related measures such as number or share of climate proofed projects, or number of projects supporting environmental sustainability, or based on compliance with internal policy, such as the extent of use of gender disaggregated data.

Various types of reviews are used to ensure solidity and credibility of the data and assessments, for example by DFID where reviewers assess whether the programme still represents value for money. Often a mix of self-evaluations and assessments and more independent assessments are used. There does not seem to be a commonly agreed standard or "best practice" in terms of how to ensure the quality and reliability of the data.

Key Findings on Tier 4

Tier 4 typically aims to answer the following question: *"How efficient is the organization?"*

³ Not reviewed in this note

The focus is clearly on internal management processes and indicators over which each institution has full control and accountability. All organizations stress the importance of efficiency for achieving the higher order results, and the AfDB gives specific examples on how the different tiers are linked, including how efficiency contribute to the overarching goals.

Most MDBs have targets and most report on these regularly for management purposes, and publish a select set of metrics annually in a scorecard and/or annual report. Colour coding is used to indicate performance, and trends are often shown through use of arrows.

Indicators track performance in areas linked to Human Resources, business processes, modernization and reform, various aspects of budgeting including allocations and adequacy, and transparency and oversight (auditing). Several dimensions tracked s are comparable across the organizations, while the dimensions tracked by the bilateral agency reviewed (DFID) are somewhat different. Numbers of indicators range from 11 to 25

Ten Good Practices from the Review

MDBs and bilateral donor institutions have all invested in reflection and experimentation with the development of their results frameworks, both in terms of the architecture, the individual indicators, and the measurement techniques chosen. There is therefore a lot of experience to draw from. The following draws out ten key “Good practices” from the reviewed organizations.

Good practice highlight 1: Both AfDB and ADB have introduced specific briefs that aim to provide “a more qualitative explanation of the link between Bank deliverables and long-term country results.”⁴ In AfDB Result Briefs are newly introduced and will use case studies to illustrate the links between tier 1 and 2 of the results framework and will “address some of the limitations of the output and intermediate outcome aggregation approach”.⁵ In ADB such briefs summarize “how the partnership of the ADB and its developing member countries has been instrumental to the development of each country. Each report includes a section on ADB’s contribution to development and poverty reduction, operational effectiveness, and future challenges.”

Good practice highlight 2: IFAD clearly demonstrate how performance against targets guides management action; for example, performance of projects that closed in the 2009-10 period showed that project efficiency was the poorest performing area. This led to an in-depth examination of the underlying causes and solutions required to accelerate progress, something that will now be addressed by a corporate-level evaluation of the Fund’s efficiency

Good practice highlight 3: The Netherlands in its reporting every two years reports based on specific, explicit results chains, linked to the MDGs, although without targets and quantitative indicators at lower levels of the results chain.

⁴ AfDB: Bank Group Resultmeasurement Framework

⁵ Op cit.

Good practice highlight 4: The IDB clearly state the criteria for selecting indicators: (i) relevance to the Bank's priorities; (ii) whether they are, or have links to MDGs, and (iii) whether they are included in other MDBs results frameworks or used by other international cooperation agencies.

Good practice highlight 5: Several indicators have been developed jointly by the MDBs and are included in the Common Performance Assessment System (COMPAS) initiative established by the Multilateral Development Bank Working Group on Managing for Development Results (MfDR); there could be some potential gains for UNDP in seeking a harmonized set of indicators with other peers.

Good practice highlight 6: AfDB and the World Bank (in the IEG annual report) clearly state what is understood by the different terms used, to avoid terminology confusion.

Good practice highlight 7: Most peers used project level data and ratings/Project Completion Reports as the basis for their results measurement. (AfDB, ADB, World Bank, IFAD, Danida, DFID), hence a strong project level M&E system can be key as the basis for developing aggregate data sets that can be broken down on country typology.

Good practice highlight 8: In recognition of the temporal dimension of results, the World Bank has established a range of indicators, where Output/Access indicators will be replaced by outcome indicators as country-level data become available. These are clearly marked in the results framework.

Good practice highlight 9: As the full results chain is not presented, the AfDB makes a point in highlighting some of the links between the different tiers in their annual report.

Good practice highlight 10: DFID tracks both its international efforts in terms action to combat climate change, and to improve governance, but also its own efforts, for example tracking its own energy consumption and carbon footprint, and performance in dealing with parliamentary questions and ministerial correspondence, an important aspect of accountability.