REPORT TO CONGRESS ON
EVALUATION STANDARDS AND PRACTICES AT THE MULTILATERAL DEVELOPMENT BANKS

A Report to Congress

in response to

Section 7029(a) of the
Consolidated Appropriations Act of 2014

United States Department of the Treasury
September 2014

Evaluation Standards and Practices at the Multilateral Development Banks

Report to Congress

September 2014

TABLE OF CONTENTS

EXECUTIVE SUMMARY

This report analyzes the current evaluation standards and practices of five multilateral development banks (MDBs), responds specifically to key elements in Section 7029(a) of the Consolidated Appropriations Act of 2014, and identifies areas of focus for the Department of the Treasury's (Treasury's) continued engagement with the MDBs on this important topic.

In preparing this report, Treasury conducted a review of current evaluation policies and practices at the five major MDBs in which the United States is a shareholder: the World Bank, the African Development Bank (AfDB), the Asian Development Bank (AsDB), the European Bank for Reconstruction and Development (EBRD), and the Inter-American Development Bank (IDB). Treasury also consulted with a wide range of experts and specialists and found that their views converged in a number of key areas.

As this report will discuss, there is no single "best" standard, type, or system of evaluation; rather, there is a wide spectrum of evaluation methodologies that employ different degrees of rigor depending on the objective, purpose, and subject of evaluation (e.g., whether designed for accountability, learning, or assessing a project, sector, country, program or approach).

Good evaluation practices require input from multiple parties, which can include MDB staff, MDB senior management, independent offices of evaluation housed at the MDBs, beneficiaries, civil society, and other fully external organizations. Each of these has a vital role in strengthening the value of the overall evaluation systems at the MDBs. In this context, evaluation experts universally stressed the importance of making balanced choices about the evaluation architecture of the MDBs.

Given the breadth and diversity of MDB operations, evaluation methods, and levels of knowledge about what works in development, there was consensus among the experts whom we consulted that there is a strong case for using a diverse range of evaluation approaches at the MDBs, rather than promoting a single approach for all MDBs and for all projects. Indeed, most experts suggest that MDBs (and other organizations) should take dynamic approaches to evaluation, employing a wide range of methods based on their appropriateness to a given project and the specific question being asked.

Section 7029(a) states that none of the funds appropriated under the Consolidated Appropriations Act of 2014 should be provided to an MDB unless it has a policy and practice of "requiring independent, outside evaluations of each project and program loan or grant and significant analytical, non-lending activity, and the impact of such loan, grant, or activity on achieving the institution's goals." As a result of our review (described below), Treasury concluded that, while current MDB evaluation practices are already quite robust, none of them meet the criteria described in Section 7029(a). Moreover, another key conclusion of this report is that conducting independent, outside impact evaluations for every single loan, grant, or activity, as described in Section 7029(a), would not necessarily achieve the results at which Section 7029(a) appears to be aimed: namely, keeping the MDBs accountable to their shareholders, clients, and beneficiaries, allowing MDBs to determine whether development programs are in fact delivering on their intended objectives, and helping MDBs improve those programs going forward.

Finally, we found that the MDBs are, in fact, the standard-setters for evaluation in the development community. This is in part due to efforts by the United States and other shareholders to encourage the use of evaluations at the MDBs for accountability and learning purposes, recognizing that both are required for institutional success. (See Annex B for an overview of Treasury efforts.) But we also identified scope for improvement in MDB evaluation practices and specific gaps that should be addressed. These include lack of clarity or transparency about how MDBs determine when to apply different methods of evaluation, especially those requiring the greatest resources; insufficient feedback loops to ensure that lessons learned are incorporated into future programming; and lack of evaluation data or data that are difficult to use. We also agree that there is some limited scope to expand the role of fully external parties in MDB evaluation. Based on these findings, Treasury has identified six priority recommendations to the MDBs that could enhance evaluation standards and practice, and, ultimately, development effectiveness and impact. They are:

1) Adopt a risk-based, strategic portfolio approach to evaluation.
2) Increase the use of external and peer reviews as part of the overall evaluation process.
3) Institute regular external evaluations of the MDBs' independent evaluation offices.
4) Create incentives and feedback loops to strengthen monitoring and evaluation (M&E) and the generation, access to, and use of learning and knowledge in MDB programming.
5) Strengthen public disclosure of data in transparent and usable forms.
6) Bolster inter-MDB collaboration (and peer pressure) on evaluation.

I. INTRODUCTION

Treasury undertook this review of evaluation standards and practices at the MDBs in response to Section 7029(a) of the Consolidated Appropriations Act of 2014. Section 7029(a) states that none of the funds appropriated under the Act should be provided to an MDB unless the Secretary of the Treasury certifies to the Committees on Appropriations that such MDB has a policy and practice of "requiring *independent, outside* evaluations of *each* project and program loan or grant and significant analytical, non-lending activity, and the *impact* of such loan, grant, or activity on achieving the institution's goals." (Emphasis added.)

The review included consultations with 30 leading evaluation experts from think tanks, academia, and development institutions, which were conducted through bilateral interviews and an expert roundtable held on April 25, 2014. It also included an extensive literature review. (For a full list of experts consulted and a bibliography, see Annexes A and E.) Because Section 7029(a) specifically mentions "independent, outside evaluations," we also consulted with European evaluation agencies that most closely carry out this type of fully external evaluation, as discussed in Section IV.

Treasury has long prioritized the use of robust and independent evaluations at the MDBs, with a focus on promoting the status, autonomy, visibility, and resources of the independent evaluation offices. (A summary of these efforts is included in Annex B.) While MDB evaluation standards and practices have improved significantly in recent years, during the course of the review we have identified specific opportunities to expand our work with MDBs in the area of evaluation.

An immediate result of Section 7029(a) has been to bolster further the evaluation work stream within Treasury's Office of International Affairs, which oversees the United States' shareholding in the MDBs, including enhanced engagement with the MDBs' evaluation units and senior management this year.

The results of the review are presented in this report. Section II introduces evaluation and its purpose in international development. Section III provides a summary of evaluation practices at the MDBs. Section IV responds directly to the legislation by discussing three of its main elements: (i) the role of independence in evaluation; (ii) the importance of having each MDB project appropriately evaluated; and (iii) the evaluation of impact. The final section provides conclusions and recommendations for Treasury's continued engagement with the MDBs on evaluation.

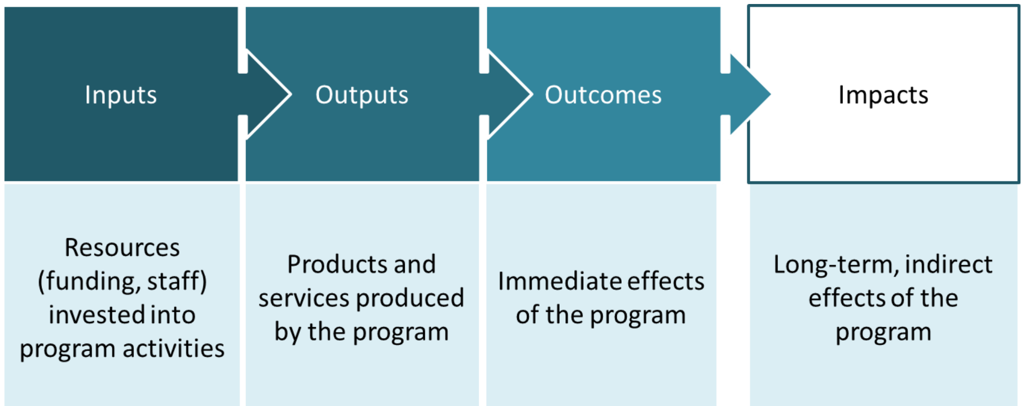
II. THE PURPOSE OF EVALUATION IN DEVELOPMENT

Evaluation is about discovering what works, what does not, and why. There are two main purposes for evaluation: accountability and learning. Evaluation for the purpose of accountability helps to "prove" whether development programs are delivering on their intended objectives, while evaluation for the purpose of learning helps to "improve" development programs by drawing out what works so that positive lessons can be identified, replicated, and scaled up.

Accountability is critical to good stewardship of MDB resources. If done effectively, evaluation can improve accountability, while recognizing that development effectiveness also requires scope for innovation, flexibility, and smart risk-taking. If evaluation tips too far in the direction of too heavily prioritizing accountability over learning, opportunities to innovate will be lost. But while fear of failure does not lead to effective development, neither do failures that are repeated over and over. Risk-taking and innovation need to take place within the construct of disciplined monitoring and evaluation so that mistakes are not repeated and successes are understood.

For example, at the project level, MDB funding may be used to purchase labor and piping (inputs) to construct a water distribution system (an output). The clean water source leads to health improvements for the community, which may result in fewer sick days from school or from work (immediate outcomes) and may, ultimately, help curb poverty by improving education and increasing household incomes (longer-term impacts). Figure 1 provides an illustration of this "theory of change." Evaluating for accountability would assess whether the inputs were properly purchased and the envisioned outputs and outcomes were achieved. Evaluating for learning would assess whether, how, and why the outputs did or did not produce the desired outcomes and impacts.

Figure 1. Theory of Change Flowchart



A. Accountability

The foundation of evaluation for accountability is monitoring, which helps track how inputs are transformed into outputs. Process evaluations often build on monitoring efforts to ensure cost efficiency and compliance with guidelines (e.g., related to procurement, fiduciary oversight, and environmental and social safeguards). As such, MDB evaluations for accountability often focus on projects' early, upfront compliance with safeguards and other policies that govern how inputs are organized and purchased. Process evaluations also monitor outputs to assess overall implementation performance. These types of evaluations are well suited to accountability because what they measure (inputs, outputs, and implementation) is within the MDBs' control.

Of course, MDBs aim to deliver more than outputs; they intend to have positive outcomes at the project level and, ultimately, to contribute to broad development objectives. Evaluating for accountability at the level of outcomes is challenging, however, because outcomes are often affected by multiple factors beyond an MDB's control. In the water distribution example, a community might have large reductions in water-borne illnesses following the project, but other efforts in the community—such as education about hygiene or the construction of a clinic nearby—might have contributed to these positive outcomes as well. Or, seasonal flooding could lead to an increase in water-borne illnesses, limiting the MDB project's positive outcomes.

The problem of attribution is even more challenging for impacts, for example reduction in poverty or increase in economic growth, because they are affected by even more factors, including the actions of the public and private sectors in the country, which often overlap with MDB efforts. It is far easier to measure and hold MDBs accountable for the outputs of projects, for example, the number of pipes installed and successful connections to households, than it is to determine the precise outcomes and impacts, for example, health outcomes and poverty reduction.

B. Learning

Evaluation for the purpose of learning aims to identify what solutions work best to achieve desired outcomes and impacts, how and why they work or do not, whether there are unintended consequences, and what can be modified to improve results. Impact evaluations—which are referenced in Section 7029(a)—are excellent tools for learning and can be conducted using diverse methodologies, as will be described in Sections III and IV. Impact evaluations are particularly well suited for assessing outcomes and impacts and provide crucial feedback that can help practitioners increase prospects for generating significant and sustainable outcomes.

Monitoring also has a role in the learning process. Collecting and monitoring data as projects are being implemented can flag problems and point to opportunities, allowing for real-time course corrections rather than waiting for the end of a project to gather lessons.

III. EVALUATION PRACTICES AT THE MDBS

Evaluation practices at the MDBs are already quite robust relative to other development organizations. Development experts agree that, in terms of evaluation, the MDBs are the standard-setters in the field and represent the vanguard among their peers.[1]

The MDBs covered by this review have many common features and modalities in their evaluation architecture. For example, at the heart of MDB evaluation systems are the independent evaluation offices, which report to the MDBs' boards of directors, and are charged with conducting independent evaluations of MDB projects, country programs, sectors, and policies—such as procurement, safeguards or gender. (The mandate and governance of each of the five independent evaluation offices are discussed in more detail in Section IV.) The MDBs' front-line operational staff also have an important role throughout the evaluation process. Involvement of MDB operational staff is critical to meeting the learning objectives of evaluation, as will be discussed below.

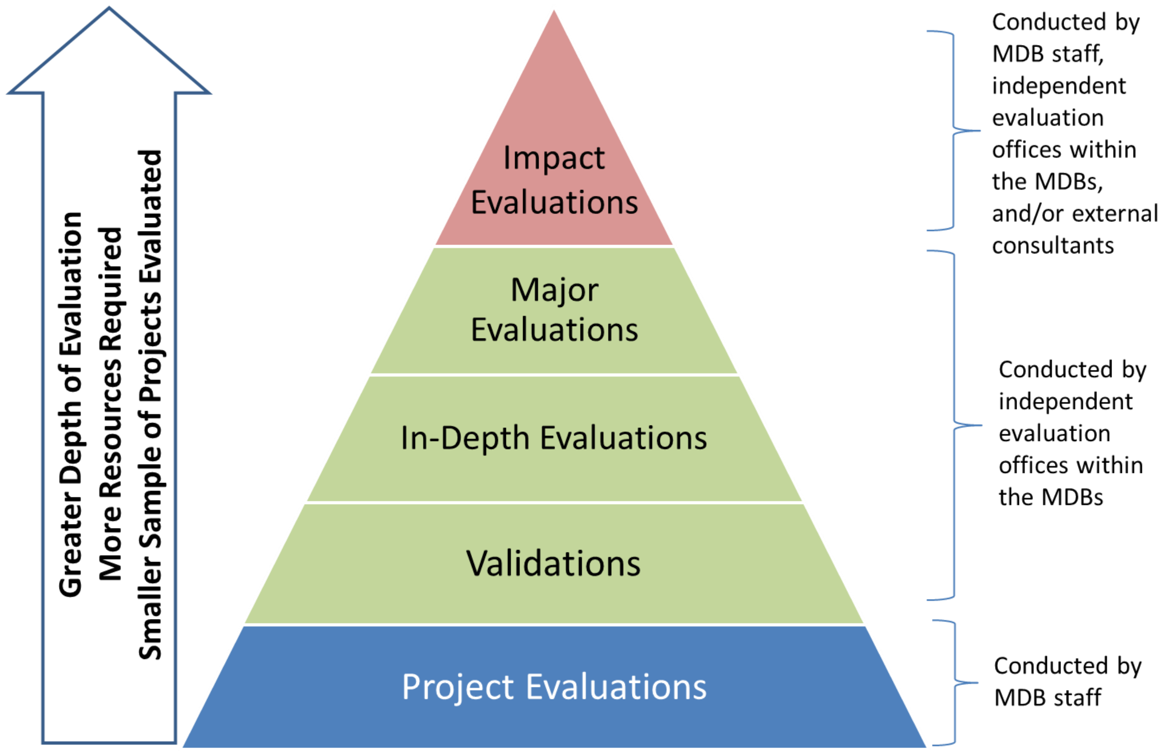
While each of the MDBs shares this general structure, each institution also has its own unique procedures and standards, and boasts its own strengths. This section takes a closer look at the MDB evaluation architecture while highlighting noteworthy variations from common practice. (See Annexes C and D for comparison chart of evaluation practices across the MDBs and additional information about the range of evaluation tools.)

A. The Evaluation Pyramid

Evaluations are not "one size fits all," and MDB evaluations can be conducted by a variety of offices and organizations, both within and external to the MDB. The building blocks of the evaluation architecture at the MDBs can best be understood as a classic pyramid, as illustrated in Figure 2. This stylized illustration is representative of the basic architecture of evaluation across all MDBs.

[1] According to the UK's Multilateral Aid Review (MAR)—a well-respected rating of 43 multilateral development organizations published in 2011 and updated in 2013—the five MDBs highlighted in this report score higher than other multilateral development agencies on the categories most closely aligned with evaluation: "strategic performance management" and "transparency and accountability." On strategic performance management, which looks at how organizations measure results, employ effective evaluation, and use evaluation evidence in decision making, the MDBs score on average almost 50 percent higher than other organizations. On transparency and accountability, which looks at organizations' disclosure policies, aid transparency, and commitment to publishing project documents and other data, the MDBs score on average 20 percent higher than other multilateral agencies.

Figure 2. The MDB Evaluation Pyramid



The broad foundation of the evaluation pyramid consists of staff-led ***project evaluations***. Project evaluations assess whether inputs were delivered and outputs produced, and occasionally cover outcomes, though they are generally not designed to measure impacts. They are typically conducted six to 12 months following project completion. These project evaluations are typically conducted for every project financed by an MDB.

The next level of the pyramid is the ***validation*** of project evaluations, which is conducted by the independent evaluation offices of the MDBs. These independent evaluation offices are generally staffed by evaluation specialists with expertise in economics, statistics, development, or specific sectors such as energy or social policy. Validations provide an independent, objective verification of the staff-led evaluations, thereby creating an incentive for honest reflection on the part of the project team. Validations are typically done through a desk review. The World Bank's independent evaluation office, the Independent Evaluation Group (IEG), validates every single project evaluation (see Box 1), while other MDBs' independent evaluation offices typically use established criteria to select a share of project evaluations to validate. The IDB's Office of Evaluation and Oversight (OVE), for example, validates about 33 percent of project evaluations, while the AfDB's Operations Evaluation Department (OPEV) validates about 89 percent.

The third level of the evaluation pyramid is the ***in-depth evaluation*** of projects. These evaluations are typically conducted one to four years after project completion and involve travel to the field site and interviews with project staff and beneficiaries. In-depth evaluations are resource-intensive and conducted on a subset of projects. Strategically selecting a subset of projects for in-depth evaluation is well accepted among the evaluation community as a prudent, risk-based, and smart approach to evaluation.

Box 1: Evaluation Product Mix at the World Bank's IEG

Of the total number of evaluations conducted annually by the IEG, approximately 90 percent are validations. While this is a high number, validations do not necessarily represent the lion's share of IEG's efforts. In fact, from a budget perspective, validations represent a much smaller proportion of what IEG does. In IEG's FY14 budget, for example, only 20 percent was allocated for validations while 53 percent was allocated for In-Depth and Major evaluations. The work streams of the MDB independent evaluation offices are set by the boards of directors and can include evaluations of country level strategies, the effectiveness of various systems and policies (including procurement and safeguards), and assessments of whether the institution is meeting its stated objectives at the corporate level.

The fourth level of the evaluation pyramid comprises ***major evaluations***, such as sector or thematic reviews and country program evaluations that go beyond a single project. The MDBs select various themes (e.g., fragile states, safeguards), sectors (e.g., infrastructure, agriculture), or regions (e.g., Central America, sub-Saharan Africa) on a rotating basis for this type of evaluation. While resource-intensive, major evaluations are more likely to result in operational changes than project level evaluations, as they provide high-level recommendations that apply more broadly to MDB operations.

At the tip of the evaluation pyramid are ***impact evaluations***, of which the MDBs conduct a small number each year. (See Annex C.) Impact evaluations are designed to attribute causal effects to a project and require extensive resources and specific expertise. However, not all impact evaluations are alike. One methodology that has received attention in recent years and has been applied to a range of development sectors, the randomized control trial (RCT), is comparable to a medical trial in that it requires a control group. (RCTs will be discussed further in Section IV.) For projects that cannot be randomized, there are other methods of impact evaluation, such as matching people affected by the project to people with similar characteristics who are not affected in order to construct a comparison group.

B. Evaluations Throughout the Project Cycle

Evaluation occurs at all stages of a project's life cycle, from design and implementation to completion and ex-post reviews several years later. While every MDB project receives some level of evaluation, not every project is subject to all the types of evaluations referenced in the evaluation pyramid.

Prior to project approval, evaluability should be carefully considered during *ex-ante* ***project design***. The teams in charge of designing projects incorporate evaluation plans into all project proposals; this is because monitoring a project or evaluating it after completion is not possible unless the project is first designed to be "evaluable." Specifically, this means clearly identifying the problem being addressed, using evidence to propose solutions, outlining a clear theory of change, and providing a framework of indicators and objectives that can be measured and evaluated. Project designers seek to establish indicators that are specific, measurable, attainable, relevant, and time-bound (i.e., the generally-accepted SMART criteria) and ensure that baseline data are collected at project inception. Up-front planning also helps embed evaluation into the institutional culture of the MDBs, creating buy-in among staff to take evaluation seriously and to recognize its value in strengthening the quality of projects.

The MDBs' ***internal control units*** are crucial stakeholders that are involved in evaluation early in the project cycle. Just as controls exist for fiduciary audits, the MDBs have policy and compliance departments that design, monitor, and adapt quality standards and safeguards for all projects. These units develop and monitor policies related to social and environmental safeguards, anti-corruption, risk management, and procurement. They also ensure that projects have been adequately prepared for evaluation. If projects do not meet threshold standards of preparedness for ex-post evaluation (e.g., having a clearly laid out theory of change, a sound monitoring and evaluation plan, and thorough risk assessment), project teams must address deficiencies before the project can be brought to the MDB's board for approval.

When a project begins, the next step in the sequence is ***monitoring and evaluation*** (M&E), which is conducted by field staff or consultants, and provides an ongoing record of project activities and outputs. Additionally, feedback from the MDB's client country governments, local implementing partners, and project beneficiaries serves an important role in the M&E function at all the MDBs. Better monitoring of projects helps implementers discover problems earlier and, ideally, allows for course correction while the project is ongoing. Strong M&E frameworks and diligent data collection during implementation are also tools for measuring development results after project completion and serve as the foundation for subsequent ex-post evaluation efforts.

After project completion, the project team that was in charge of implementation conducts an *ex-post* ***project evaluation***. As mentioned, these evaluations are conducted under management oversight, often with the support of external consultants, and assess whether the project (i) achieved its stated objectives; (ii) was implemented in accordance with MDB policies and safeguards; (iii) delivered outputs as anticipated; (iv) appropriately mitigated risks, both foreseen and unexpected; (v) followed its monitoring and evaluation protocol; and, (vi) remained within budget.

Of course, there is the potential for a conflict of interest when team members evaluate their own work, especially if there is a perception that honesty about failure will result in negative repercussions. Indeed, a key challenge is to ensure that any negative repercussions address failures caused by employee error or poor decision-making, rather than honest failures due to innovation and the willingness to explore new ideas. To address this risk, each MDB has in place mechanisms that incentivize credibility in the staff-led project evaluations and help deter the whitewashing of results, rather than punishing admissions of failures. For example, project staff know that all projects may have a likelihood of being validated by the MDB's independent evaluation office. MDBs periodically review and update the templates used by staff to conduct project evaluations to increase their relevance for measuring performance. (See Box 2.)

In Treasury's consultations, evaluation experts underscored how important MDB staff members are to strong evaluation standards. The MDBs' employees often have the best knowledge about the causes of project success and failure, which can be difficult for outsiders to discern. An important way to get key information about complex project outcomes is from the project designers and implementers themselves. Thus, it is important that they are not shut out of the evaluation process. Moreover, evaluation experts cite the importance of creating an "evaluation culture" within the MDBs. Staff involvement in the evaluation process can, if properly managed, make project designers and implementing teams more receptive to incorporating lessons learned into future projects.

Box 2: Improving Project Evaluation at IDB

In 2006 the IDB introduced a revised, simplified template for project evaluations in an effort to encourage better reporting. In 2012, the IDB's Office of Evaluation and Oversight (OVE) reviewed the process and found that while reporting had improved, more than half of project evaluations incorrectly used output indicators to measure outcomes. There were also large disparities between the ratings that project staff assigned to their projects and the findings of the independent evaluator, as well as insufficient evidence to support staff ratings. OVE recommended further changes that are now being implemented, including use of a new template, enhanced staff training, greater incentives for good project evaluations, and adequate resources for project evaluations in project budgets.

At this point in the sequence, evaluation becomes the mandate of the MDBs' ***independent evaluation offices***, which perform a wide range of independent, ex-post evaluations of projects, including evaluations conducted a number of years after project completion. These offices typically produce the evaluation products at every level of the evaluation pyramid except the initial level (project evaluations) and the top level (impact evaluations).

First and foremost, the independent evaluation offices review project evaluations to assess (by either validating or overturning) the initial staff-assessed ratings. As mentioned previously, these validations provide an independent verification of the analytical quality of the project evaluation, and an independent view on the project's performance and lessons learned. It is not unusual for there to be a discrepancy between the ratings given by MDB staff in their project evaluations and the ratings given by the independent evaluation offices in their validations. For example, on average, the World Bank's IEG typically gives positive ratings to about 10 percent fewer projects than the MDB staff's own ratings.

Beyond project-level evaluations, the independent evaluation offices conduct in-depth evaluations that involve field visits and interviews of project beneficiaries, local government officials, and MDB staff. For example, in 2012 the AsDB's Independent Evaluation Department (IED) conducted an in-depth evaluation of a $150 million project to strengthen the social services sector in Pakistan. The IED rated the project as "less than successful," because the AsDB did not coordinate effectively with the new government following Pakistan's 2008 elections. The evaluation provided eight lessons learned, disseminated the evaluation within the AsDB, and made the report available publicly online.

In addition, the independent evaluation offices conduct major evaluations, which can assess country programs, sectors, or themes. These typically look at a large number of projects over several years. In 2013, for example, the IDB's OVE conducted a thematic review of the IDB's engagement in the citizen security sector from 1998 to 2012. The evaluation found that the IDB played a pioneering role in addressing citizen security. It also found that the IDB faced significant challenges in managing the risks of this increasingly complex thematic area. The independent evaluation offices also conduct corporate evaluations, which provide MDB-wide assessments of program results and institutional effectiveness.

Impact evaluations, as will be discussed later in this report, are not typically conducted by the independent evaluation offices because they require extensive work before and during a project's implementation that may be more suited for operational teams and external specialists.

C. External Evaluators

The selective use of external evaluators, who do not fall under the purview of the independent evaluation offices, is another pillar of the MDB evaluation architecture. These assessments complement the types of evaluations already discussed, and they are typically commissioned by the MDB's board of directors to provide an appraisal of a particular aspect of its operations in order to guide policy reforms and operational strategies. There are multiple ways in which MDBs bring in external evaluators:

- ***Evaluations of the Evaluators***. Formal external evaluations of the MDB independent evaluation offices help strengthen the independence of the offices and bring in knowledge about the broader base of evidence that exists in the field of development evaluation. There is an external evaluation of the World Bank's IEG underway, but these are not annual occurrences: the last World Bank IEG external review was in 2004, and the EBRD and AfDB's independent evaluation offices have not yet received an external evaluation. One of the recommendations we put forward in Section V is that the MDBs should require evaluations of their independent evaluation offices more frequently and at fixed intervals.

- ***External and Peer Reviews.*** Some MDBs have established standing review committees of prominent academics and policymakers to provide advice to senior management (e.g., an Economic Policy Research Advisor Council at the EBRD, which commissions in-depth reports and evaluations of bank activities in particular sectors). There is also an active practice of retaining external experts for evaluation work, as well as frequent ad hoc peer reviews of independent evaluators' evaluations. Peer reviews and other engagement by external evaluators are facilitated by the work of professional groups such as the Multilateral Organization Performance Assessment Network (MOPAN) and the Evaluation Network of the Development Assistance Committee (DAC) of the Organization for Economic Cooperation and Development (OECD). MOPAN produces evaluations of MDBs' effectiveness based on shareholder and client surveys. The Evaluation Network of the DAC promotes joint and collaborative evaluations involving multiple development partners as a tool for mutual accountability and learning. We believe greater use of external and peer reviews would be beneficial, a conclusion reflected in our recommendations.

- ***External Impact Evaluations.*** The MDBs also tap into the growing number of organizations, think tanks, and academic institutions that specialize in conducting impact evaluations. For example, the World Bank, the AsDB, and the IDB have all partnered with the Abdul Latif Jameel Poverty Action Lab (J-PAL) to conduct impact evaluations of their projects. J-PAL, housed in the Economics Department at the Massachusetts Institute of Technology, is a prominent research center that conducts rigorous evaluations to test and improve the effectiveness of programs and policies aimed at reducing poverty. Additionally, the World Bank, the AfDB, and the AsDB all work with the International Initiative for Impact Evaluation (3ie) for research, training, and learning purposes. 3ie conducts impact evaluations and "systematic reviews," which distill the findings of multiple impact evaluations in a way that can usefully inform the direction of future development projects. The MDBs utilize 3ie's systematic reviews and "evidence gap maps"—which look at the existing completed evaluations and highlight areas where there is strong, weak, or non-existent evidence—to identify areas where there is an urgent need for more research.

- ***External Review of Publicly-Disclosed MDB Data.*** An often-overlooked component of MDB evaluation is the public disclosure of data, specifically the open and transparent availability of project-level data that independent academics and other outside stakeholders can scrutinize and use to assess performance and progress and to provide external recommendations on how to improve future projects. In general, project evaluations are publicly disclosed, though Treasury's review found that some MDBs are better than others in this regard. For example, even when the project evaluations are published online, often the data used in them are not available or easy to find. For this reason, one of our recommendations is that the MDBs strengthen the disclosure of data in a user-friendly format.

- ***Evaluation Cooperation Group***. All of the MDBs' independent evaluation offices are part of a network called the Evaluation Cooperation Group (ECG). Established in 1996, partly in response to a U.S. initiative on evaluation at the MDBs, the goals of the ECG are to harmonize performance indicators and evaluation methodologies to a high standard, to share lessons learned from evaluations across MDBs, and to enhance evaluation professionalism and collaboration. We believe that even stronger inter-MDB mechanisms would be beneficial, and have incorporated this as a recommendation as well.

IV. RESPONSE TO LEGISLATION

Section 7029(a) of the Consolidated Appropriations Act of 2014 raises specific issues about evaluation standards and practices at the MDBs. In the course of our review, Treasury has concluded that, while MDB evaluation practices are already quite robust, it does not appear that any MDB fully meets the elements of Section 7029(a), and there are serious questions about whether it is advisable, practical, or feasible for them to do so. The remaining sections of this report aim to explain why this is so and offer alternative approaches intended to meet the spirit of the legislation.

In this section, we address three specific elements in Section 7029(a) that all experts consulted agreed merit discussion, and their relationship to the ultimate goal of high standards of performance at the MDBs. These elements are:

A. the requirement for "independent, outside evaluation,"
B. the reference to "each project," and
C. the inclusion of "impact" evaluations in the provision.

A. Ensuring Independence

Independence enhances credibility and minimizes bias in the evaluation process. However, we believe that there are strong arguments that requiring evaluations to be both independent *and* outside will not lead to a more effective or efficient evaluation architecture for the MDBs. In addition, our review found that the MDBs' independent evaluation offices have consistently demonstrated an ability and willingness to issue strong, high quality, candid, and often hard-hitting and highly-critical evaluations of MDB activities.

Independence – a core pillar of evaluation

While the independent evaluation offices are not external to the MDBs, their governance structures have strongly embedded institutional safeguards to protect their independence. These offices grew out of the original evaluation functions at the MDBs and became increasingly independent over time. In 1974, the World Bank was the first MDB to make its evaluation department independent, through reforms that separated the department from management and placed it under Board authority. The regional MDBs made their evaluation functions independent more recently, in the late 1990s and early 2000s through reforms that were led and supported by Treasury. Treasury continues to push on this issue to this day. (See Box 3.)

The independent evaluation offices are organizationally independent: they report to the MDBs' Boards of Directors, not to management, which frees them to conduct objective evaluations without political pressure, allegiances to bank programs, or the fear of repercussions due to negative assessments. Independent funding sources and full control over the design and execution of their own evaluation work plans provide these independent evaluation offices with further protection from outside influences. There is no management involvement in the independent evaluators' staff recruitment, hiring, or firing. Additionally, the directors have term limits, and are prohibited from ever working at the MDB after their term expires.

Box 3: Evaluator Independence at the AfDB

To be effectively independent, MDB evaluation offices should report to the Board rather than management. All of the MDBs currently use this governance structure except the African Development Bank (AfDB), where the head of the independent evaluation office reports both to the Board and to management. Under pressure from key shareholders, including the United States, the AfDB is revising and strengthening its independent evaluation policy. Treasury is working to ensure that the AfDB evaluator is given strong organizational independence, including protecting the evaluation group's budget from political influence and ensuring that the head of the evaluation group is primarily accountable to the Board.

Some have expressed the concern that independent evaluation offices can never truly be independent if housed within the

MDBs. Externality or placement "outside" of the MDB is often equated with independence. However, in practice this tradeoff is more complex, and the perception of independence and its reality are not necessarily the same.

External evaluators are not immune from pressures or incentives that may cloud independence. For example, when evaluations are outsourced repeatedly to a select number of consulting firms or academic institutions, dependence on fees and interest in receiving future contracts can constitute a threat to the integrity of their evaluations. If consultants are retained by MDB management, this relationship also has the potential to impair their impartiality. Most of the outside experts consulted by Treasury were skeptical that firms relying on repeat business would be more objective than the independent evaluation offices. A major lesson is that wherever the evaluation function is located, strong safeguards and committed leadership are needed to preserve and nurture independence.

We believe that sufficient conflict-of-interest safeguards are in place at the MDBs to help ensure that staff members' personal and professional relationships and financial interests are not allowed to influence their judgments or impair objectivity. Together, these organizational features promote behavioral independence at the MDBs' independent evaluation offices.

Independence, but not isolation

The governance structure described above establishes enough distance to ensure independence, while housing independent evaluation offices at the MDBs provides them with close enough proximity to gain familiarity with and access to the projects and programs being evaluated. This produces deeper insights and expertise, and a better chance of eliciting cooperation from MDB employees, who have unique knowledge of projects and MDB systems. The major advantage that independent evaluation offices housed within MDBs have over fully external evaluation models is their deep knowledge of the internal workings and operational complexities of these institutions.

External evaluators are often perceived to have greater independence and credibility, and while they can certainly produce robust evaluations, they often suffer from inadequate understanding of the operating context. The learning curve is steep, as external consultants must be briefed on the projects and culture of the MDB. Additionally, investing MDB staff time to interact with consultants can be time-consuming and resource draining. These costs are magnified by the fact that external consultants eventually move on to other assignments, and thus do not provide a source of continuous knowledge that stays within the MDB. In addition, a critical component of evaluation is the response of management and staff; for these products to have real value, their lessons and recommendations need to be integrated into the MDB. The potential for follow up is weaker if the evaluator is an external contractor without a long-term relationship with the MDB.

Despite the risk of perceived lack of independence or institutional capture, the proximity to MDB activities is an important element in crafting and tracking actionable recommendations, establishing credibility with MDB staff, and promoting feedback loops. As evaluation expert Robert Picciotto noted in the book *Independent Evaluation at the IMF*, "having no connection or shared experience with the intended users of evaluations constrains evaluators' access to information, evokes resistance, and inhibits learning."

The external, outside evaluation model

As part of our review, Treasury proactively identified examples of fully external, outside evaluation models. We located three examples in the field of development, all within the context of European bilateral aid agencies: the United Kingdom's Independent Commission for Aid Impact (ICAI), Sweden's now defunct Swedish Agency for Development Evaluation (SADEV), and Germany's Institute for Development Evaluation (DEval).

These three examples were all funded by national governments, outside the budgets of the bilateral development agencies that they were reviewing, and reported to authorities outside of these agencies. They were established largely to complement the role of the national auditing organizations and to help perform the parliamentary role of scrutinizing the work of the bilateral development agencies for accountability purposes.

In general, the work programs in all three institutions were limited by a lack of access to and knowledge of the institutions being evaluated, as well as by their own capacities as small institutions with limited staff and budgetary resources. None of the three organizations conducted evaluations at the project level nor did they fund rigorous impact evaluations. They all tended to focus on sector and thematic reviews and compliance audits. Moreover, neither ICAI nor DEval (the two models still functioning today) has the ability or expertise to evaluate each project undertaken by the bilateral development agencies that they are charged with assessing.

While these models add another layer of accountability, the experience of these three institutions suggests that such a model in the multilateral setting would not significantly strengthen the current policy and practice of independent evaluation at the MDBs.

B. Evaluating Each Project Appropriately

Section 7029(a) envisions that "each" project be required to receive a full evaluation. Conducting a rigorous evaluation of every single project undertaken by the MDBs would lead to a misallocation of resources, as it would require evaluation of even very low-risk projects, diverting resources from evaluation of higher-risk activities. In addition, such an approach would be prohibitively expensive and would be unlikely to generate greater knowledge and lessons learned from evaluation. More importantly, the legislative provision does not attempt to distinguish between different types of projects that might require or benefit from greater or lesser degrees of evaluation.

Moving towards a risk-based and strategic portfolio evaluation approach

Treasury agrees that all projects must receive a basic level of monitoring and evaluation for accountability purposes. Beyond that, what should trigger the decision to perform various types of evaluations—not just at the project level, but also thematic or country evaluations? There is broad agreement that MDBs should carefully deploy a range of evaluations under a strategic portfolio approach that is risk based and uses clear guidelines and triggers to guide how the MDBs and the independent evaluation offices allocate funding and staff among the types of evaluation described in the previous sections.

Box 4: Impact Evaluation at the World Bank

The World Bank's Development Impact Evaluation (DIME) unit provides technical assistance to ensure high quality monitoring and impact evaluation. Every project assisted by DIME includes baseline and end-line data collection as well as a robust analytical strategy to help attribute results to a project, better understand causal links, and inform strategy. While the World Bank has significant evaluation resources from IEG and DIME, there is not an effective institution-wide strategy for selecting projects for impact evaluation. DIME, IEG, and the World Bank operations teams could more efficiently utilize evaluation through better coordination and by targeting impact evaluations to fill important knowledge gaps.

We found that a challenge to current practice is the lack of a clear methodology for selecting projects for higher levels of evaluation (using the pyramid analogy, the MDBs do not have a uniform or systematic method for selecting which projects will move up the pyramid). This is especially true for selecting projects for impact evaluation. (See Box 4.) Today, personal interest on the part of specific project team leaders is a major driver for which projects receive impact evaluations.

For the strategic portfolio approach to be robust, it is important that it not be haphazard. All projects should have a chance of being selected for scrutiny, but resources should be focused on projects where there are the greatest risks or opportunities for learning. Given the challenges in establishing a strategic portfolio approach, we have incorporated recommendations on this topic.

C. Using Impact Evaluations

Section 7029(a) suggests that all loans, grants, and other activities at the MDBs should undergo an "impact evaluation" to determine the ultimate effect of the activity on poverty alleviation and equitable growth. As noted, such an approach would be counter-productive and cost-prohibitive, and in many cases, the cost of the evaluation would exceed the cost of the activity itself.

Prioritizing impact evaluations

In the development community, there is no single approach to impact evaluation. As noted previously, there are several types of impact evaluations, each suited for different types of projects. The most robust type of project-level impact evaluation is the RCT, a method originally pioneered in medical science. RCTs identify the causal effect of a project by comparing outcomes for a large group of people receiving the benefits of a project (the "treatment group") to a large group of people who were not included in the project (the "control group"); whether a person is assigned to the treatment or control group is completely random, so the two groups are otherwise similar. RCTs are well suited for evaluations of targeted, micro level interventions (e.g., the impact of mosquito nets on reducing the incidence of malaria), but not for evaluating complex programs that aim to effect systems changes or policy work, such as the impact of introducing tariff reforms.

RCTs require an enormous amount of up-front planning, expert design, baseline data collection, the execution of complex evaluation methodology, and technical ex-post statistical analysis.

RCTs can also be quite expensive, with costs on par with the overall operational cost of small projects (e.g., between $200,000 to $2 million per project) and require teams of evaluators to collect large amounts of data in the field over the course of multiple years. In addition, RCTs require the consent of the participating partners to hold back project benefits from the control group to avoid contamination. Not all governments or civil society partners are willing to exclude certain households from receiving project benefits, citing ethical concerns. The challenge of consent is even greater with private sector partners who are not interested in withholding services where a potential market exists.

There are other methods of impact evaluation as well, such as comparing outcomes for project participants to a group that, while not a randomly selected control group, has similar characteristics to the project participants. These non-randomized methods, while useful in cases where there are ethical concerns about randomly choosing who receives the benefits of a project, still require extensive data collection and technical expertise and can be similarly costly.

MDBs face a trade-off between evaluating all projects and evaluating a thoughtfully selected sample of projects more rigorously. Impact evaluation experts generally agree that it is more valuable to conduct a few, very well done impact evaluations (selected strategically) than numerous impact evaluations on similar topics (which results in greater coverage across the project portfolio, but generates less-useful knowledge about effectiveness). A 2012 evaluation by IEG of the relevance and effectiveness of impact evaluations at the World Bank recommends the development of "mechanisms for strategic identification and prioritization of impact evaluations to balance learning and results measurement objectives." Further, experts do not support impact evaluations on topics for which large bodies of evidence have already been established. For example, there is now a significant body of knowledge around "conditional cash transfers," showing that cash transfers to the poor are most effective when tied to certain conditions for the recipients, such as sending their children to school. We also know a great deal about the effectiveness of certain kinds of health and education projects. These areas have benefitted from a substantial number of impact evaluations conducted around the globe over the past decade, and additional impact evaluations are unlikely to have as much value.

V. CONCLUSION AND RECOMMENDATIONS

This final section summarizes our major conclusions and provides priority areas for Treasury's continued engagement with the MDBs on strengthening evaluation standards and practices. These recommendations, if applied, could considerably improve evaluation at the MDBs. Our views, validated by nearly every outside party with whom Treasury consulted, are that:

- MDB practices are already quite robust, notwithstanding scope for continued refinement.
- Independence does not require that evaluators be completely external—indeed, establishing fully external evaluation systems may be counter-productive to the goals of strengthening the use and application of evaluations.
- Sound evaluation requires the involvement of many parties, including MDB staff.
- Impact evaluations should be deployed strategically, rather than applied universally to every MDB project, and are best suited for a relatively small subset of projects.

Treasury has identified six priority recommendations to the MDBs that we will press for as part of our enhanced engagement with them to continue to improve and refine their evaluation standards and practices, specifically in response to Section 7029(a):

1) **Adopt a risk-based, strategic portfolio approach to evaluation.** As noted in this report, there is no "one size fits all" solution to evaluation. It is important for development organizations to have a dynamic evaluation strategy that can guide decisions on what level of evaluation a project should receive. While it is good practice for a subset of projects to receive different types of evaluation, including impact evaluation, the selection of that subset must follow transparent, pre-established criteria.

2) **Increase the use of external and peer reviews as part of the overall evaluation process.** When used strategically, external and peer reviews strengthen the independence and quality of MDB evaluations. Professional external consultants or even ad hoc, unpaid peer reviewers contribute to evaluation quality at all levels of the evaluation pyramid. The independent evaluation office at each MDB is best-placed to establish an institution-wide strategy for utilizing external and peer reviews of evaluation products.

3) **Institute regular external evaluations of the MDBs' independent evaluation offices.** While some of the MDB independent evaluation offices are periodically evaluated, Treasury recommends that these evaluations be conducted at fixed intervals across all MDBs. External evaluations should be overseen by each MDB's Board of Directors, perhaps on a schedule that is aligned with the tenure of each independent evaluation office's director.

4) **Create incentives and feedback loops to strengthen M&E and the generation, access to, and use of learning and knowledge in MDB programming.** Ultimately, a high-quality evaluation is valuable only if it leads to a change for the better. While the current MDB evaluation systems do have mechanisms for providing feedback to operational teams, more can be done to tighten those feedback loops and provide incentive structures for MDB project staff to apply lessons learned into future projects, for example, rewarding the replication of success and offering incentives for data collection and mid-course corrections. Additionally, more can be done to incorporate real-time evaluation into project implementation, and to strengthen M&E systems. Potential failures can be avoided if implementers have the ability and the authority to monitor carefully and make mid-course corrections while a project is ongoing. Ultimately, actionable evaluation requires a robust institutional culture around evaluation.

5) **Strengthen public disclosure of data in transparent and usable forms.** Data disclosure is important for accountability and is also a public good for the development community, including academics, practitioners, and evaluators. Data can be used by external groups and individuals to derive additional lessons through their own independent analysis and evaluation. Each MDB is making strides to increase data transparency, but more can be done to ensure that project-level data is published in a user-friendly way and that MDB evaluation findings are more widely disseminated.

6) **Bolster inter-MDB collaboration (and peer pressure) on evaluation.** All of the MDBs' independent evaluation offices are active members of the Evaluation Cooperation Group (ECG), but the role of this consortium can be bolstered. As a network of peers, the ECG should explicitly commit to raising the bar for standards and practices across the MDBs. Specifically, the ECG could take a more proactive role to encourage, track, and have MDBs hold each other accountable for continual improvement of evaluation standards and practices. While the ECG does regular benchmarking exercises, the MDBs should be encouraged to make specific commitments for improvement following these exercises. Additionally, MDB management and operational staff should be included in cross-fertilization efforts on evaluation to share knowledge products throughout these institutions and promote stronger cultures of learning.

ANNEX A. LIST OF EXPERTS CONSULTED

NAME	ORGANIZATION
Rakesh Nangia	AfDB – Director, Operations Evaluation Department (OPEV)
Vinod Thomas	AsDB – Director General, Independent Evaluation Department (IED)
Homi Kharas	Brookings Institution
Scott Morris	Center for Global Development
Stefan Dercon	U.K. Department for International Development
Joe Eichenberger	EBRD – Chief Evaluator, Evaluation Department (EvD)
Alix Zwane	Evidence Action
Franck Wiebe	Georgetown Public Policy Institute
Rohini Pande	Harvard Kennedy School
Lant Pritchett	Harvard Kennedy School
Ruth Levine	Hewlett Foundation
Sandy Darville	IDB
Cheryl Gray	IDB – Director, Office of Evaluation and Oversight (OVE)
Dean Karlan	Innovations for Poverty Action
Annette Brown	International Initiative for Impact Evaluation (3ie)
Richard Manning	International Initiative for Impact Evaluation (3ie), former Chair of OECD Development Assistance Committee (DAC)
Howard White	International Initiative for Impact Evaluation (3ie)
Eva Terberger	KfW (German) Development Bank – Head, Evaluation Department
Bob Picciotto	Kings College, London
Sixto Aquino	Millennium Challenge Corporation
Margaret Kuhlow	Overseas Private Investment Corporation
Jeff Hammer	Princeton University
Eva Lithman	Swedish International Development Cooperation Agency
Lauren Platukis	U.S. State Department
Duc Tran	USAID – Development Innovation Ventures
Melissa Patsalides	USAID – Bureau for Policy, Planning and Learning's Office of Learning, Evaluation, and Research
Caroline Heider	World Bank – Director General, Independent Evaluation Group (IEG)
Richard Scobey	World Bank - IEG
Marvin Taylor-Dormond	World Bank - IEG
Nick York	World Bank - IEG

ANNEX B. TREASURY'S EFFORTS TO PROMOTE EVALUATION AT THE MDBS

Treasury and the U.S. Executive Directors' Offices at the MDBs have consistently advocated for stronger evaluation policies and appropriate resourcing in Board and committee meetings, and during negotiations for general capital increases (GCI) and concessional window replenishments.

In the 1990s, Treasury worked across the MDBs to press for the adoption of "results matrices," under which each project would be assessed to determine whether it achieved its intended outputs and outcomes. These matrices have now become standard. More recently, Treasury has pushed the MDBs to aggregate project results to enable shareholders to assess the extent to which MDBs are supporting development objectives across countries, regions and sectors. This effort contributed to the genesis of the World Bank Corporate Scorecard, adopted in 2011, an annual report that reviews both outputs and the extent to which Bank projects met its standards. Through an integrated results and performance framework that tracks both development results and corporate performance, the Corporate Scorecard is now the Bank's major tool for monitoring and reporting, at an aggregate level, the effectiveness of its operations and its overall efficiency and adaptability as an institution. Most of the other MDBs have since adopted a similar scorecard approach.

During negotiations for the 16th replenishment of the International Development Association (IDA), Treasury secured commitments from **World Bank** management to adopt a corporate strategic approach to the use of impact evaluations. This included increasing the number of projects with appropriate evaluation frameworks, improving the quality of impact evaluations, enhancing the learning process to strengthen feedback loops for project design, and convening a panel of experts to provide recommendations on strengthening the World Bank's framework for impact evaluations. During the 17th IDA replenishment negotiations, World Bank management committed to continue efforts to make more systematic use of impact evaluations, develop and mainstream a wider range of evidence-based tools and approaches to strengthen monitoring and evaluation, and provide real-time data to support project mid-course corrections. Treasury has continued to advocate that World Bank management find ways to better incentivize project staff to undertake mid-course corrections based on data from monitoring and evaluation.

Treasury has pressed the **African Development Bank** on several areas of results reporting and evaluation. We secured commitments from management during GCI negotiations in 2010 on improving quality-at-entry measures and readiness reviews of projects prior to approval and adopting guidelines on project completion reporting. We built on this during African Development Fund replenishment negotiations in 2013 by pressing for increased independence of the head of the independent evaluation unit and greater use of impact evaluations. Management also agreed to create a mechanism to record how it has responded to recommendations from the evaluations department, set up a database of projects completed since 2000 with information on development outcomes and monitoring and evaluation findings, and strengthen real-time results reporting.

Treasury has also pressed **Asian Development Bank** management (in the context of GCI and Asian Development Fund replenishment negotiations) to increase the number of impact evaluations and assess the results and sustainability of projects several years after completion. This resulted in management establishing an interdepartmental impact evaluation committee in 2011 to oversee implementation of a growing number of impact evaluation studies. Management also agreed to strengthen its focus on the latter stages of project implementation, including post-completion monitoring, and to increase staff training on managing for development results.

Treasury has been instrumental in pushing a number of changes in evaluation at the **European Bank for Reconstruction and Development** (EBRD). This includes creating a formal results framework to be included in new country strategies, which will improve the strategic focus and evaluability of the EBRD's work at the country level. Treasury has strongly supported a reform to the management-conducted quality-at-entry rating system (known as the "Expected Transition Impact rating"), which we hope will lead to more consistent achievement of ex-ante project objectives. Treasury has also supported the independent evaluator's decision to conduct a greater number of project evaluations in order to build a more comprehensive data set to generate learning and knowledge for future EBRD projects.

During negotiations on the **Inter-American Development Bank's** (IDB) GCI in 2010, Treasury emphasized that management should establish a strong ex-ante results measurement framework, and all projects should be evaluated during and after implementation to determine whether specific and tangible results are achieved. Management has responded to this requirement by updating it development effectiveness matrix (DEM), which measures a project's evaluability at entry, and introducing a pilot for a new project completion reporting system. The independent evaluator has also begun validations of randomly selected projects' DEM scores and staff-led project evaluations. The IDB has increased the number of projects with impact evaluations from 10 in 2008 to almost 60 in 2013.

ANNEX C. SUMMARY TABLE OF MDB EVALUATION PRACTICES

Evaluation Practices at the Multilateral Development Banks					
	African Development Bank	Asian Development Bank	European Bank for Reconstruction and Development	Inter-American Development Bank	World Bank
Independent Evaluation Office (IEO)	Operations Evaluation Department (OPEV), to be renamed Independent Development Evaluation (IDEV)	Independent Evaluation Department (IED)	Evaluation Department (EvD)	Office of Evaluation and Oversight (OVE)	Independent Evaluation Group (IEG)
Percentage of Projects Subject to Project Evaluation	70-90% of projects self-evaluated (PCRs); ultimate goal is 100%.	100% of projects self-evaluated.	100% of projects self-evaluated.	100% of projects self-evaluated.	100% of projects self-evaluated.
Percentage of Projects Subject to Validation by IEO	In 2012, 89% of PCRs were subject to independent desk review (PCREN).	The IED goal is to validate 75% of PCRs by operating units for public sector lending, and 100% of extended annual review reports (XARRs) for private sector lending.	A sample, designed to yield statistically sound and credible results, is subject to independent evaluation or validation by EvD.	Random sample (approximately 33%) validated by OVE. OVE validates 100% of non-sovereign guaranteed operations.	100% of self-evaluations validated by IEG.
Percentage/Number of Projects Subject to In-Depth Review by IEO	OPEV selects clustered samples of projects for independent performance evaluation (with field visit). Selection criteria are mainly: quality of PCR; lessons to be learned; importance to Country or sector; or bank corporate strategic priorities.	IED selects a purposeful sample of approximately 10 public sector projects and 3 non-sovereign operations for in-depth evaluation each year (PPERs).	EvD selects 5-10 transactions per year for in-depth evaluation.	OVE selects a sample of projects to receive in-depth review as part of "comparative project reviews," classified as "high level" (sector) evaluations.	IEG selects about 12% of projects for in-depth evaluation (PPARs).

Number of High-level Evaluations Conducted by IEO (Sector, Thematic, Geographic)	5 per year for the 2012-2014 work plan.	About 8 per year for the 2013-2014 work plan.	8 new evaluations for the 2014 work plan.	9-14 evaluations for the 2013-2015 work plan.	About 20 per year for the 2015-2017 work plan.
Number of Project Impact Evaluations Conducted by IEO	First launched in 2014; a second scheduled to commence January 2015.	1 per year for the 2013 – 2016 work program.	N/A	The IDB's first impact evaluations were conducted by OVE, but OVE subsequently recommended that IDB project staff is better suited to conduct them.	Impact evaluations are conducted by World Bank operational staff and external consultants.
Number of Project Impact Evaluations Conducted by Bank Staff or other non-IEO staff	N/A	20 ongoing, 5 of which are part of a special initiative to build the AsDB's technical capacity to implement impact evaluations.	N/A	67 ongoing impact evaluations.	324 ongoing impact evaluations; 170 of which the Development Impact Evaluation Initiative (DIME) is leading.
Budget	About 2.2% of the AfDB's administrative budget.	About 1.8% of the AsDB's administrative budget.	About 1.2% of the EBRD's Administrative budget.	About 1.3% of the IDB's administrative budget.	About 0.9% of the World Bank Group's administrative budget.
Budget Determined By	Board, not separate from administrative budget.	Board, separate from administrative budget.	Board, separate from administrative budget.	Board, separate from administrative budget.	Board, separate from administrative budget.
Head of IEO Reports to	Board, with an administrative link to the President.	Board	Board	Board	Board
Terms of Appointment for Head	5 year term, renewable once, selected by Board with involvement by Management.	5 year non-renewable term, selected by Board.	4 year term, renewable but not traditionally for more than 8 years, selected by Board.	5 year term, renewable once, selected by Board.	5 year renewable term, selected by Board.
Right of Return for Head	No, unless Board decides otherwise.	No	No, unless Board decides otherwise.	No	No

Consultants as proportion of IEO budget	47%	30%	16%	About 25%	23% of total full-time equivalent staffing.
Last external evaluation of IEO	None	2008	None	2012	2004; next external review will be completed in 2015.
Departments or Special Programs Supporting Impact Evaluation	None	The "Implementing Impact Evaluation" technical assistance project provides $1 million to conduct an impact evaluation in each of the regional departments, with support from the Economics and Research Department.	None	Office of Strategic Planning and Development Effectiveness (SPD) provides support for project impact evaluations.	DIME provides support for project impact evaluations; the Strategic Impact Evaluation Fund (SIEF) provides funding for impact evaluations in the areas of health, childhood nutrition, education and sanitation; the Africa Impact Evaluation Initiative provides support for impact evaluations in Africa.

ANNEX D. EVALUATION TOOLS

As noted in this report, the MDBs employ a diverse range of evaluation methods and tools to support accountability and learning. Two helpful ways of classifying these tools is by "category" (that is, evaluations that share similar themes and seek to answer similar questions) and by "level" (that is, which level of MDB operations is being assessed, from project-level evaluations to assessments of entire country programs or MDB strategies). A given level of evaluation might employ tools from a variety of different evaluation categories.

Categories of Evaluation

Process evaluations are designed to ensure that the project delivers the intended inputs on schedule, according to the procedures outlined in the project proposal, and in compliance with the institution's safeguards, procurement, and other policies. Process evaluations typically answer the following types of questions:

- Are tasks being completed and services being delivered?
- Is the project in compliance with MDB standards, policies, and safeguards?
- Is the project reaching the intended beneficiaries?

Performance evaluations focus on outputs and outcomes, and assess the degree to which the targeted objectives were achieved. The bulk of the evaluative work conducted during the MDB project cycle falls under this category. Performance evaluations typically answer the following types of questions:

- Did the project produce the planned outputs and outcomes?
- Which components of the project worked and which did not?
- Was the project completed on time and within budget?

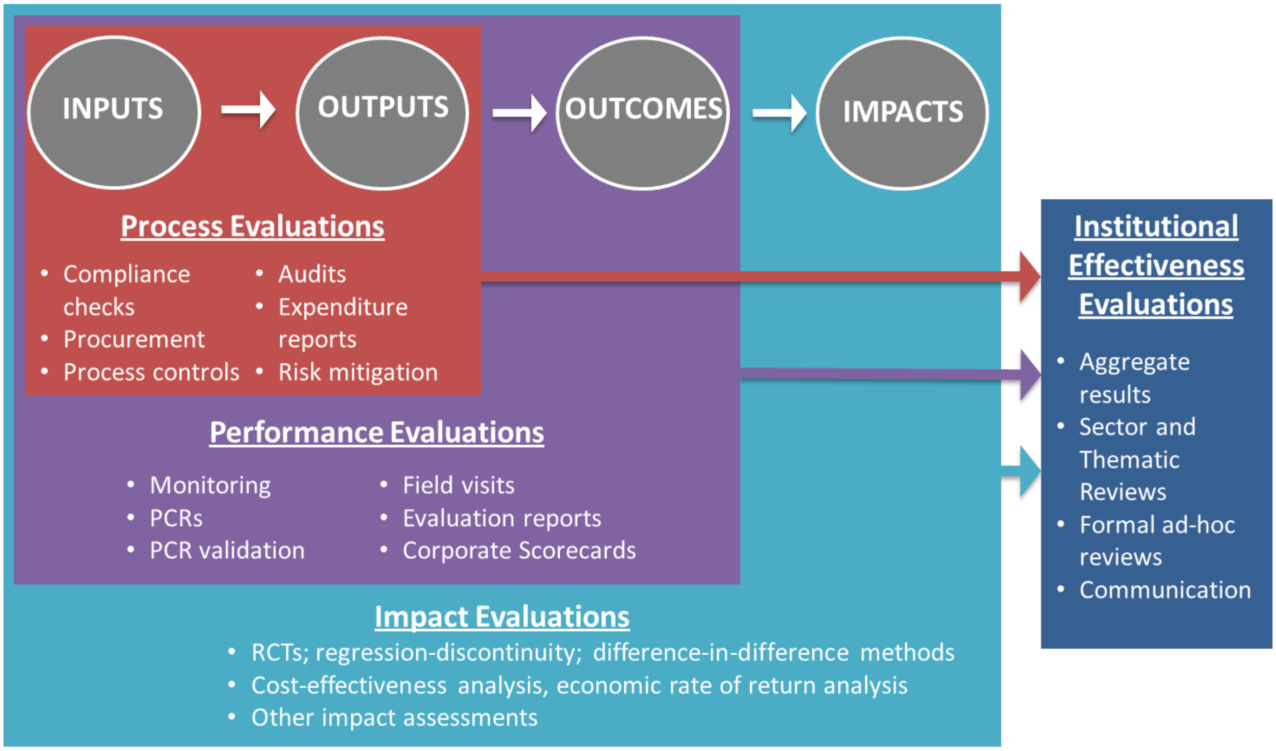
Impact evaluations go beyond inputs, outputs, and outcomes to measure the impact of a project and whether the impact is a result of the project itself. Impact evaluations typically answer the following types of questions:

- Would the observed impact have occurred in the absence of the project?
- Is the impact different for different types of beneficiaries?
- How, or through what causal channels, did the impact occur?

Institutional effectiveness evaluations seek to evaluate the overall effectiveness of a development organization. Many MDBs aggregate results from individual project evaluations to help determine institution-wide outcome indicators, which are often presented as "corporate scorecards" that report results. These efforts also typically focus on outputs, outcomes, and the share of projects rated satisfactory or better. Institutional effectiveness evaluations typically answer the following questions:

- Do program strategies support the institution's mission and priorities?
- Is the institution making the best use of resources, including relationships with other organizations?
- Are agency procedures and management policies conducive to effective work?

Figure A1. Categories of Evaluation and Theory of Change



Levels of Evaluation

Project-level evaluations assess process, performance, or impact at the project level and enable the MDBs to assess whether a project should be scaled up or replicated in another context.

Sector evaluations, also called "thematic evaluations," are conducted on a sample of projects to identify common lessons that can be applied to future projects in a sector or thematic area. Examples include evaluations of agricultural projects or projects focused on gender.

Country evaluations seek to describe and explain the performance of MDB programs at a country level. They evaluate progress towards country program objectives and seek to determine whether the mix of project design and implementation was the right fit for the country circumstances. Country evaluations are forward-looking and often influence future country strategies and operations. They provide valuable insights for improving coordination among development agencies, recipient governments, and civil society in borrowing countries.

Strategic evaluations are higher-level evaluations that seek to assess whether an organization's overall strategy—or a specific strategy in a particular region or sector—is effective. Often, when an MDB adopts a new multi-year strategy it will perform a strategic evaluation at the mid-point of the multi-year period to assess performance and offer recommendations for changes. Strategic reviews enable an MDB to refine and improve the organization's overall approach.

ANNEX E. BIBLIOGRAPHY

African Development Bank. (2007, March 23). *Independent Evaluation Policy and Functional Responsabilities of the Operations Evaluation Department (OPEV).*

African Development Bank. (2009, April 3). *Guidelines on Project Completion Reporting*. Retrieved August 1, 2014, from afdb.org: http://www.afdb.org/fileadmin/uploads/afdb/Documents/Policy-Documents/PCR%20Guidelines%20Approved%20by%20Ops%20%20Com%20%20on%2003-04-2009.pdf

African Development Bank. (2012, May). *Disclosure and Access to Information: The Policy*. Retrieved August 14, 2014, from http://www.afdb.org/fileadmin/uploads/afdb/Documents/Policy-Documents/Bank_Group_Policy_on_Disclosure_and_Acess_to_Infomation.pdf

African Development Bank. (n.d.). *History: Over 25 Years of Evaluation Heritage*. Retrieved August 1, 2014, from afdb.org: http://operationsevaluation.afdb.org/en/about-us/history/

Andrews, M., Pritchett, L., & Woolcock, M. (June 2012). Escaping Capability Traps through Problem-Driven Iterative Adaptation (PDIA). *Center for Global Development Working Paper 299*.

Asian Development Bank. (2008). *Review of the Independence and Effectiveness of the Operations Evaluation Department.* ADB Policy Paper.

Asian Development Bank. (2013, October 1). *Operations Manual Bank Policies (BP): Independent Evaluation.*

Asian Development Bank. (2014). *Portfolio Performance*. Retrieved August 1, 2014, from adb.org: http://www.adb.org/site/evaluation/portfolio-performance

Asian Development Bank Independent Evaluation Department. (2010, October). *Special Study on Post-Completion Sustainability of Asian Development Bank-Assisted Projects.* Retrieved August 1, 2014, from oecd.org: http://webcache.googleusercontent.com/search?q=cache:Q4MDJDUOV64J:www.oecd.org/derec/adb/47186868.pdf+&cd=8&hl=en&ct=clnk&gl=us

Asian Development Bank Operations Evaluation Department. (2007, March). *Independent Evaluation at the Asian Development Bank*. Retrieved August 1, 2014, from adb.org: http://www.adb.org/sites/default/files/pub/2007/Independent-Evaluation-ADB.pdf

Bank, W. (2013). *Project Performance Assessment Report: Tanzania Poverty Reduction Support Credits 1-8*. Retrieved from https://ieg.worldbankgroup.org/Data/reports/PPAR-78188-P074072_Tanzania_PRSCs_1-8.pdf

Bill and Melinda Gates Foundation. (n.d.). *How We Work: Evaluation Policy*. Retrieved July 31, 2014, from gatesfoundation.org: http://www.gatesfoundation.org/How-We-Work/General-Information/Evaluation-Policy

Blattman, C. (2008, February 14). Impact Evaluation 2.0.

Council of Inspectors General on Integrity and Efficiency. (2012, January). *Quality Standards for Inspection and Evaluation.*

Department for International Development. (2011, March). *Multilateral Aid Review 2011*. Retrieved August 13, 2014, from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/67583/multilateral_aid_review.pdf

Department for International Development. (2013). *International Development Evaluation Policy.*
Department for International Development. (2013, December). *Multilateral Aid Review Update 2013.* Retrieved August 13, 2014, from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/297523/MAR-review-dec13.pdf
Development Innovation Ventures. (2014). *Annual Letter2014.* Retrieved July 31, 2014, from http://www.usaid.gov/sites/default/files/documents/9276/USAIDAnnualLetter2014.pdf
European Bank for Reconstruction and Development. (2008, May 12). *Public Information Policy.* Retrieved August 14, 2014, from 2008. http://www.ebrd.com/downloads/research/policies/pip08.pdf
European Bank for Reconstruction and Development Evaluation Department. (2012, June). *Briefing Note: Evaluability - Is it Relevant for EBRD?* Retrieved August 12, 2014, from http://www.ebrd.com/downloads/about/evaluation/130214.pdf
European Bank of Reconstruction and Development. (2012). *Evaluation Department Work Programme 2012.*
European Bank of Reconstruction and Development. (2013). *Evaluation Policy.*
Evaluation Cooperation Group. (2004, March 17). Template for Assessing the Independence of Evaluation Organizations.
Evaluation Cooperation Group. (2009, March). *Review Framework for the Evaluation Function in Multilateral Development Banks.* Retrieved 2014, from ECGnet.org.
Evaluation Cooperation Group. (2010, June). *Good Practices Standards on Independence of International Financial Institutions' Central Evaluation Departments.*
Evaluation Cooperation Group. (2010, March 16). *Peer Review of IFAD's Office of Evaluation and Evaluation Function.* Retrieved August 12, 2014, from ECGnet.org: https://wpqr4.adb.org/LotusQuickr/ecg/Main.nsf/7e6e83714d63fac348257731002a960f/3f296435ce2c24c648257731002a9639/?OpenDocument
Evaluation Cooperation Group. (2011, November 8). *Good Practice Standards for the Evaluation of Private Sector Investment Operations: Fourth Edition.* Retrieved 2014, from ECGnet.org.
Evaluation Cooperation Group. (2011, September). *Third Benchmarking Review of ECG Members' Evaluation Practices for Their Private Sector Investment Operations.*
Evaluation Cooperation Group. (2012, February). *Good Practice Standards for the Evaluation of Public Sector Operations: 2012 Revised Edition.*
Evaluation Cooperation Group. (n.d.). *ECG Comparison Table.* Retrieved 2014, from ECGnet.org.
Evaluation Coordination Group. (2013, May 8). *Work Program and Budget (FY14) and Indicative Plan (FY15–16).* Retrieved 2014, from ECGnet.org.
Galindo, A., & Gray, C. (2014, April 11). IDB's Evolving Development Effectiveness Framework: Presentation to 3ie. Inter-American Development Bank.
Grasso, P., Wasty, S., & Weaving, R. (2003). *World Bank Operations Evaluation Department: The First 30 Years.* Retrieved August 1, 2014, from worldbankgroup.org: https://ieg.worldbankgroup.org/Data/reports/oed_30yrs.pdf
Hausman, R., Klinger, B., & Wagner, R. (2008). Doing Growth Diagnostics in Practice: A 'Mindbook'. *CID Working Paper No. 177.*

Independent Commission for Aid Impact. (2013, December). Triennial Review of the Independent Commision for Aid Impact.

Independent Evaluation Group. (2012). *Independent Evaluation Group Annual Report 2012.* Washington, DC: World Bank.

Indpendent Evaluation Group. (2012). *World Bank Impact Evaluations: Relevance and Effectiveness.* Washington, DC: World Bank.

Independent Evaluation Group. (2013). *Afghanistan Country Program Evaluation, 2002–11: Evaluation of the World Bank Group Program.* Retrieved from http://ieg.worldbankgroup.org/Data/reports/afghan_eval_full.pdf

Independent Evaluation Group. (2013, January 25). Guidelines to Manage Conflicts of Interest in Independent Evaluation.

Independent Evaluation Group. (2013). *Independent Evaluation Group Annual Report 2013.* Washington, DC: World Bank.

Independent Evaluation Group. (2013). *Results and Performance of the World Bank Group 2012: Volume I.* Washington, DC: World Bank.

Independent Evaluation Group. (n.d.). *Revised Mandate of the Director-General, Evaluation.* Retrieved July 31, 2014, from http://ieg.worldbank.org/Data/dge_mandate_tor.pdf

Independent Review Panel on Evaluation at the Inter-American Development Bank . (2011). *Strengthening Evaluation to Improve Development Results.* Washington, DC: Inter-American Development Bank.

Inter-American Development Bank Office of Evaluation and Oversight. (n.d.). *Project Completion and Evaluation Reports* . Retrieved August 1, 2014, from iadb.org: http://www.iadb.org/ove/engbook/evaiv.htm

Inter-American Development Bank. (n.d.). *OVE: History*. Retrieved August 1, 2014, from iadb.org: http://www.iadb.org/en/office-of-evaluation-and-oversight/about-us-history,6727.html

Inter-American Development Bank Working Group on the Board of Executive Directors on Oversight and Evaluation. (1999, June). *Final Report: Strengthening Oversight and Rebuilding Evaluation in the Bank.* Retrieved August 1, 2014, from iadb.org: http://idbdocs.iadb.org/wsdocs/getdocument.aspx?docnum=388290

International Fund for Agricultural Development. (2011, May 10). *Revised IFAD Evaluation Policy.* Retrieved August 1, 2014, from ifad.org: http://www.ifad.org/gbdocs/eb/102/e/EB-2011-102-R-7-Rev-1.pdf

International Fund for Agricultural Development. (2013, November). *Guidelines: Project completion report validation and project performance assessment.* Retrieved August 1, 2014, from ifad.org: http://www.ifad.org/evaluation/process_methodology/doc/pr_completion.pdf

Legovini, A. (2010). *Development Impact Evaluation Initiative: A World Bank-Wide Strategic Approach to Enhance Developmental Effectiveness.* Washington, DC: Development Impact Evaluation Initiative.

Millennium Challenge Corporation. (2012). *Policy for Monitoring and Evaluation of Compacts and Threshold Programs.* Washington, DC: Millennium Challenge Corporation.

Organization for Economic Co-operation and Development. (2010). *Evaluating Development Co-operation: Summary of Key Norms and Standards.* Retrieved July 31, 2014, from http://www.oecd.org/development/evaluation/summaryofkeynormsandstandards.htm

Organization for Economic Co-operation and Development. (2014, February 12-13). Summary Record of the 16th Meeting of the DAC Network on Development Evaluation.
Picciotto, R. (2012). "The Logic of Evaluation Independence and its Relevance." *Independent Evaluation at the IMF: The First Decade*, edited by Lamdany, R., & Edison, H. Washington, DC: International Monetary Fund.
Rossi, P. H., Lipsey, M. W., & Freeman, H. E. (2004). *Evaluation: A Systematic Approach (Seventh Edition).* Thousand Oaks: Sage Publications, Inc. .
Smith, K. F. (2006, March 31). *Quality in Design and Monitoring Frameworks.* Retrieved August 12, 2014, from Asian Development Bank Operations Evaluation Department: http://www.adb.org/sites/default/files/mar01-oed-working-paper.pdf
Thomas, V., & Luo, X. (2012). *Multilateral Banks and the Development Process: Vital Links in the Results Chain.* New Brunswick: Transaction Publishers.
United States Agency for International Development. (2011, January). *Evaluation: Learning from Experience - USAID Evaluation Policy.* Retrieved July 31, 2014, from usaid.gov: http://www.usaid.gov/sites/default/files/documents/1868/USAIDEvaluationPolicy.pdf
White, H. (2014, May 1). *Is independence always a good thing?* Retrieved July 31, 2014, from Evidence Matters: Improving development policy and practice: http://blogs.3ieimpact.org/is-independence-always-a-good-thing/
White, H., & Waddington, H. (2012). Why do we car about evidence synthesis? An introduction to the special issue on systematic reviews. *Journal of Development Effectiveness, 4*(3), 351-358.
World Bank Independent Evaluation Group. (2013). *Delivering the Millennium Development Goals to Reduce Maternal and Child Mortality.* Washington, DC: The World Bank.