

# 1. Assessing the World Bank Group's Self-Evaluation Systems

## Self-Evaluation in the World Bank Group

Self-evaluation, or the formal, written assessment of a project, program, or policy by an entity engaged in that activity (see complete definition in the Glossary), has been used systematically in the World Bank (the International Bank for Reconstruction and Development, or IBRD, and the International Development Association, or IDA) for 40 years and has recently been introduced in the International Finance Corporation (IFC) and the Multilateral Investment Guarantee Agency (MIGA) as well. For much of that period the systems used in the Bank Group have been at the forefront of efforts to improve the achievement of results by the world's development agencies (Appendix A).

Self-evaluation lies at the heart of the Bank Group's results measurement system. Since its introduction in the Bank in 1976, self-evaluation has evolved to include a wide range of tools and approaches to measuring and validating results. In the 1990s, the launch of results-based management in the Bank Group greatly expanded its systems, adding attention to country results to an existing focus on project results. For a while, the results of Bank sector strategies and policies were also assessed through retrospectives shared with the Board. Efforts to aggregate corporate results and track corporate commitments over the past 15 years have led to added demands on the self-evaluation systems, coming in part from pressure from donor nations around IDA replenishments, as donors need results measurement to make the case with their own governments, parliaments, and citizens on the value of IDA.

This evaluation seeks to assess whether and how the Bank Group's systems serve their expected purposes through a broad examination of the ways in which the systems operate, including analysis of the behaviors, incentives, and organizational culture surrounding the production and use of self-evaluation.

## Self-Evaluation Purposes and Uses

The closest the World Bank comes to a statement of purpose for self-evaluation is in Operational Policy 13.60, which frames the purposes of monitoring and evaluation (M&E) as follows:

*"The Bank's objective is to assist its borrowing member countries, individually and collectively, to reduce poverty and achieve sustainable*

## CHAPTER 1 ASSESSING THE BANK GROUP SELF-EVALUATION SYSTEMS

growth. To assess the extent to which its efforts and those of borrowers are making progress toward that objective, the Bank monitors and evaluates its operational activities. Monitoring and evaluation provides information to verify progress toward and achievement of results, supports learning from experience, and promotes accountability for results. The Bank relies on a combination of monitoring and self-evaluation and independent evaluation. Staff take into account the findings of relevant monitoring and evaluation reports in designing the Bank's operational activities."

IFC's "Operational Procedures – Portfolio and Supervision" stipulates three main purposes of self-evaluation as performance measurement, accountability, and learning.

In addition, the 2013 World Bank Group Strategy makes a number of statements about the role of (self) evaluation:<sup>1</sup>

- "supporting clients in delivering customized solutions that...encompass the complete cycle from policy design through implementation to evaluation of results lies at the heart of the...value proposition." (para 55)
- "the [Bank Group]...has made significant progress in helping clients focus on results [including] rigorous evaluations of program impacts... [it] needs to focus more specifically on how its engagements contribute concretely to reducing poverty and boosting shared prosperity, as well as how to monitor and measure results as a Group." (para 59)
- "The science of delivery centers on ensuring that the intended benefits of development solutions are realized in practice. [One facet to effective and efficient delivery is] evaluating whether [the promised] goods and services... benefit the targeted citizens and results in the intended outcomes." (para 60)
- "The [Bank Group] will... Develop an internal...results framework, with the ...Scorecard at the apex, and with the key elements being reflected down into [vice presidential unit] VPU/business unit and staff performance agreements....This framework is intended to strengthen the accountability for results." (para 61)
- "The [Bank Group] will establish a more evidence-based and selective country engagement model. [This model comprises three main elements including] Performance and Learning Reviews [that] will identify and capture lessons from implementation to determine mid-course corrections, end-of-cycle learning, and accountability, as well as to help build the [Bank Group]'s knowledge base...." (para 68).

Based on the elements of the systems as they exist today and as described in the various cited documents, the implicit purposes of the Bank Group's self-evaluation systems are to *measure performance, verify progress toward the achievement of results, promote accountability* (including by providing information that supports the Independent Evaluation Group's [IEG] evaluations), and *support learning* that leads to enhancements of operational quality. This compares quite closely with the purposes as presented in the literature on evaluation according to which an ideal self-evaluation system should be able to support:

- Performance management internally via data and information that can assist evidence-based project, portfolio, policy, and strategy decision-making
- Reporting on project and portfolio results suitable to support internal and external accountability mechanisms
- Learning about challenges to managing for and achieving results.

### **Self-Evaluation Coverage in the Bank Group**

The Bank Group's self-evaluation systems cover many different operational product lines and their scope, processes, and methodologies have important differences (box 1.1, figure 1.1). Some reports are validated by IEG and feed into organizational scorecards and results measurement systems, others do not. An important distinction can be made between the mandatory self-evaluation products and voluntary evaluation studies such as impact evaluations and occasional programmatic evaluations or retrospectives commissioned by business units.

#### **THE WORLD BANK SELF-EVALUATION SYSTEMS**

Project or program design documents should describe the expected results, and monitoring systems should regularly collect data on those results. Self-evaluations — written by or for the operational department in charge of the activity — use the design documents, monitoring data, and other information to describe what happened, what was achieved, identify lessons, pass evaluative judgments, and assign ratings. The resulting report becomes an important permanent record of the activity. Information from self-evaluations and IEG's independent validation of them is used to report aggregated results and for accountability purposes. Learning from self-evaluation helps improve performance over time, at least in theory. Figure 1.1 shows what is covered by self-evaluation and how processes work.

The Bank began requiring all operating departments to prepare self-evaluation project completion reports in 1976. Those early reports were subject to review by the evaluation department (now known as IEG) before being submitted to the Board. Their

## CHAPTER 1 ASSESSING THE BANK GROUP SELF-EVALUATION SYSTEMS

variable quality resulted in a tightening of evidentiary standards in the late 1970s. A brief attempt to mandate self-evaluation by borrowers in 1980 led to a decline in report quality and timeliness and eventually Bank staff resumed preparation of completion reports, with an option for borrowers to provide their comments, which go on the record as an appendix to the reports.<sup>2</sup>

### Box 1-1. Bank Group Self-Evaluation Instruments

Self-evaluation in the Bank Group covers most operational activities. Primary, mandatory self-evaluation systems include:

- ICR (Implementation Completion and Results Report) for Bank lending at closing
- ISR (Implementation Status and Results Report) for Bank lending in implementation
- Country Partnership Framework Completion and Learning Reviews for country programs
- XPSR (Expanded Project Supervision Report) for IFC investments at maturity
- PCR (Project Completion Reports) for IFC advisory projects at closing
- PSR (Project Supervision Reports) for active IFC advisory projects
- PER (Project Evaluation Reports) for MIGA guarantee projects.

There are also voluntary self-evaluations:

- Impact evaluations
- Evaluative studies, such as IFC's program performance evaluations.

Data from self-evaluations feed into corporate results measurement:

- World Bank Group corporate scorecard; IFC, MIGA, and World Bank scorecards
- IDA's results measurement system
- The website of the President's Delivery Unit
- Various internal portfolio monitoring reports.

Some activities are **not currently covered** by self-evaluation, such as:

- Board operations
- Control and Treasury functions
- The Bank's Reimbursable Advisory Services
- Country programs under country engagement notes
- Various assessment tools such the Country Financial Accountability Assessment.

Figure 1.1 and the Approach Paper for this evaluation<sup>3</sup> present a more detailed inventory.

A decline in the development effectiveness of Bank projects in the early 1990s spurred a number of changes. The ICR was introduced with validation by IEG after submission to the Board rather than before, resulting in ratings differences between the ICR and IEG's validation (Appendix A). The Quality Assurance Group was established (and later disbanded) to evaluate the quality at entry, quality of supervision, and overall portfolio

performance. Over time, the independent evaluation function has worked closely with Operations Policy and Country Services (OPCS) to harmonize rating systems, adjust ratings criteria, and introduce new self-evaluation products.

Currently, World Bank carries out self-evaluations which IEG validates for all IBRD/IDA operations regardless of funding size and all recipient executed trust funds above \$5 million (with a few exceptions).<sup>4</sup> The evaluations assess the project against the original project objectives and any subsequent formal revisions and rate outcomes based on criteria for relevance, effectiveness, and efficiency. Risk to development outcome, Bank performance, and borrower performance are also assessed and rated. In addition, IEG separately assesses and rates project M&E and the quality of the self-evaluation. IEG also writes Project Performance Assessment Reports on a purposefully selected share of projects, currently around 15 percent.

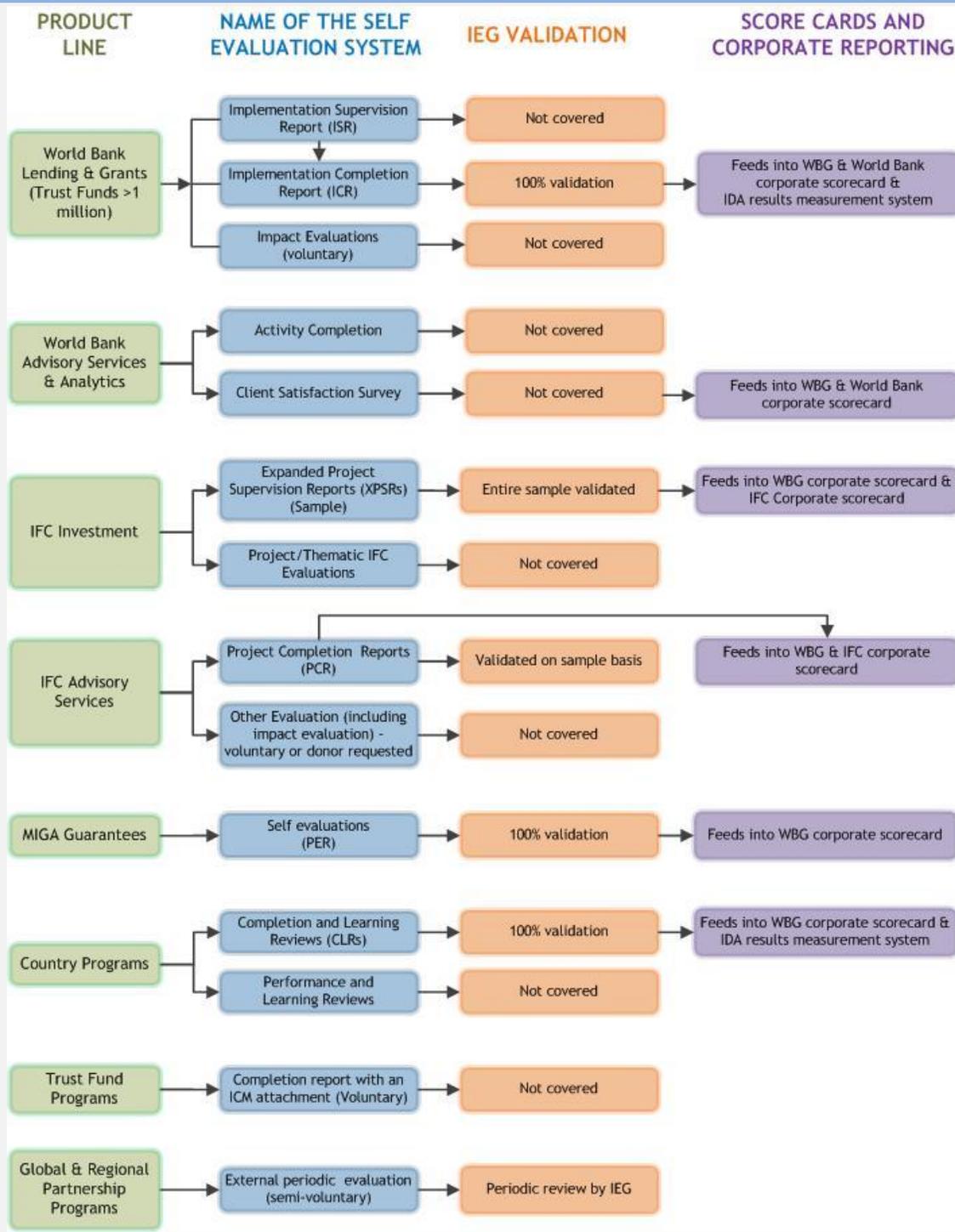
#### **THE IFC INVESTMENT AND MIGA GUARANTEE SELF-EVALUATION SYSTEMS**

IFC uses self-evaluation to assess performance, results, and effects on private sector development. It and started conducting formal self-evaluations of its projects in 1987. The system changed a number of times, and, in response to a 1995 review, was modified to focus on accountability for corporate objectives and identification of lessons.<sup>5</sup> IFC began preparing XPSRs in 1999 for a random representative sample selected by IEG (recently lowered to 40 percent) of investment projects, all of which are then validated by IEG. The evaluation criteria cover project or program effects on stakeholders and include financial, economic, environmental and social, and private sector development dimensions, along with IFC's investment return, work quality, and additionality. The evaluation standards and guidelines, developed in collaboration with IEG, reflect a combination of benchmarks, qualitative criteria, and performance standards. To protect commercially sensitive client information, IFC self-evaluations are restricted and are not shared with the Board. IFC also runs various quality assurance programs such as DOTS, client surveys, credit risk rating, research, and knowledge management activities most of which fall outside the scope of this report.

MIGA started selective ex post self-evaluations in 1996, and, after a period in which IEG evaluated MIGA projects, resumed self-evaluations of projects in 2010 on a pilot basis and using an approach that resembles that of IFC in many ways.<sup>6</sup> MIGA has collaborated with IEG on the development of its self-evaluation system, which is now fully operational but still at an early stage.

**CHAPTER 1**  
**ASSESSING THE BANK GROUP SELF-EVALUATION SYSTEMS**

**Figure 1.1. The Scope and Process of Self-Evaluation Differ across Operational Product Lines**



## ADVISORY AND KNOWLEDGE SERVICES

The Bank has put in place a reporting infrastructure for advisory and knowledge services but has yet to develop a reliable way to evaluate the effectiveness of this work. It does, however, collect client feedback through a new World Bank Satisfaction Survey, the results of which are used in different ways. Creating a system to reliably measure results of Bank knowledge work has proved difficult for two reasons. First, the number of products and their relatively small size make an elaborate, activity-level results architecture impractical, unless done on a selective basis.<sup>7</sup> Second, it has proven difficult to establish the necessary conditions for evaluability.<sup>8</sup>

Corporate guidelines for Bank ASA require statements of a development objective and intermediate outcomes. A large number of guidance documents and intranet pages are available to lead users through the process. IEG's review of the guidance indicates there is greater attention to the transactions involved (for example, how to enter the required information in the Operations Portal) than to the design of ASA and the attendant planning for data or observations that would signal that an ASA has been successful in achieving its objectives. Assessing results of ASA is made complicated by the dynamic nature of policy dialogue and policy change: results may not have materialized when the ASA closes, and attribution of results is often not possible.

IFC pays closer attention than the Bank to demonstrating results in its advisory services using a centralized results-monitoring system that has been in place since 2005. It has a comprehensive self-evaluation system, designed in consultation with IEG, that assesses all projects and is embedded in the project cycle from design to completion. M&E specialists sign off on results frameworks and M&E plans, and, on completion, review reported results and evidence, a sample of which are then validated by IEG. Advisory Services generally involve IFC helping to implement IFC-funded investments, resulting in more comprehensive depth and coverage of PCRs than the Activity Completion for Bank knowledge products.

## IMPACT EVALUATION

The World Bank Group has expanded and deepened its impact evaluation work over the past decade. Between 2004 and 2008, the number of Bank Group-supported impact evaluations increased sevenfold, starting with the creation of the Development Impact Evaluation Initiative (DIME) in 2005, followed by the Africa Gender Lab, the Strategic Impact Evaluation Fund (SIEF), and the Health Results Innovation Trust Fund. At the IDA replenishment in 2010, donors called on World Bank management to further strengthen the Bank's impact evaluation program, which management subsequently scaled up.<sup>9</sup> In FY15, 82 impact evaluations went through concept review at the Bank. Impact evaluations are financed mostly by trust funds provided by donors for this

## **CHAPTER 1**

### **ASSESSING THE BANK GROUP SELF-EVALUATION SYSTEMS**

purpose (Appendixes C and F). The fact that the selection of projects for impact evaluations is not carried out according to transparent rules and depend on the team's self-selection and the interest of trust fund donors makes this instrument less suited for corporate accountability purposes. It also cannot be used for aggregated reporting, and is more akin to research. For these reasons, IEG does not validate impact evaluations.

#### **COUNTRY PROGRAMS**

The Bank Group strengthened its focus on strategic country-level engagements in the late 1990s as aid perceptions changed and as part of the evolving results agenda. Self-evaluations of country strategies started in 2003. Today, Country Partnership Frameworks are prepared jointly where relevant and the Completion and Learning Reviews assess and rate achievement of program objectives and the Bank Group's contributions. They evaluate Bank Group performance with regard to program design and implementation and attempt to separate the efforts of the Bank Group from exogenous factors. They are meant to fulfill a learning function through lessons and findings to guide future country programs. The assessments are also validated by IEG.

#### **THE CORPORATE SCORECARDS**

The Bank Group corporate scorecard, created in 2011 and under continuous revision, is a major element of the organization's external accountability framework. The stated purpose of the scorecard is to "provide a high-level and strategic overview of the World Bank Group's performance toward achieving the institution's goals. It is the apex from which indicators cascade into the monitoring frameworks of the three World Bank Group institutions."<sup>10</sup> There are three tiers:

1. Development context – Reports the long-term development outcomes that countries are achieving.
2. Client results – Reflects the results of clients implementing Bank Group-financed operations.
3. Performance – Covers operational and organizational effectiveness.

#### **THE MANAGEMENT ACTION RECORD**

The Management Action Record (MAR) is a tool that tracks follow-up on the adoption of recommendations made by IEG in its major evaluations. It was most recently reformed in 2012. For each IEG recommendation to which it agrees, Bank Group management formulates an action plan. It then reviews and rates its adoption status annually, a form of self-evaluation. IEG also comments and rates the status of adoption. The MAR is updated annually. Recommendations are tracked for four years, after which they are retired.

## Costs of Producing Self-Evaluation

There is no consistent method of budgeting for self-evaluations and tracking expenses involved in producing them. Expenditures on World Bank ICRs are charged against a general project budget code, and mission and other work done for ICR activities could be mixed with other purposes, limiting the ability of this evaluation to accurately measure costs. ICRs on average cost \$40,000-\$50,000 each to produce, according to interviews with resource management staff, and around 300 are done each year, yielding a total cost of approximately \$13 million. This is a highly imprecise and lower-bound estimate and does not include monitoring, ISRs, quality reviews, interaction with IEG during validation, IEG's own costs, and the costs to clients to provide data and their own responses. Considering also Country Completion and Learning Reviews, XPSRs (of which 76 were done in 2015), PCRs, and PERs brings the estimated total cost of producing self-evaluations to at least \$15 million or around 0.6 percent of the Bank Group's total annual administrative budget (excluding quality reviews and IEG's budget). See Appendix C for details.

## Evaluating Self-Evaluation

### WHY EVALUATE NOW?

The ongoing reform of the Bank Group is a good time to assess how well the self-evaluation systems support the mission of the institution. This report links closely to ongoing efforts to enhance operational and organizational effectiveness, promote a learning and "development solutions" culture, simplify internal processes, and promote evidence-based decision-making. Strong M&E is closely associated with high performance and contributes to the learning and mid-course correction emphasized by the 2013 Bank Group strategy and the results focus of IDA. Further, the recent external review of IEG has noted that IEG is only one component of a broader system that is not well-articulated or working optimally.

### EVALUATION SCOPE

The report covers self-evaluation of Bank and IFC projects (and, very selectively, MIGA guarantees), as well as knowledge and advisory services, country programs, and impact evaluation. While "appraising the World Bank Group's operations self-evaluation...and attesting to their adequacy to the Boards" has long been part of IEG's mandate, and earlier reports did this separately for the Bank and for IFC and MIGA, this is IEG's first review of the entire self-evaluation system.<sup>11</sup> The report complements or builds on other IEG reports, most notably:

## CHAPTER 1

### ASSESSING THE BANK GROUP SELF-EVALUATION SYSTEMS

- The annual Results and Performance Report; where trends in results are assessed. In contrast, this report reviews how results are measured.
- Learning and Results, Volumes I and II; where the ways in which Bank staff learn are assessed. This report reviews how self-evaluation fosters learning.
- IEG's Biennial Report on Operations Evaluation (BROE), which covered IFC (up to 2008) and IFC and MIGA in 2013. This report pays closer attention to Bank systems and more selective attention to certain aspects of IFC's systems (because MIGA's systems are still relatively new and evolving, their coverage is limited) (IEG 2013).

Evaluation systems can be analyzed at various levels. Three levels are considered in this report: (a) templates, guidelines, and information technology; (b) business processes, data streams, reporting lines, and roles and responsibilities; and (c) behaviors influenced by motivations that are both extrinsic (incentives) and intrinsic (norms and values) as well as organizational culture. The report also examines the interfaces between self-evaluation and IEG's independent validation and evaluation functions, recognizing that IEG validation influences processes, behaviors, and incentives.

The report focuses on the production and use of mandatory self-evaluation and impact evaluation. One limitation of this report is that it does not cover occasional voluntary self-evaluations apart from World Bank impact evaluation such as retrospectives because no inventory or database tracks them (in the Bank) and they can be hard to distinguish from broader studies.<sup>12</sup> IFC impact evaluations are also outside the scope. The report selectively assesses how monitoring information feeds into the production and use of self-evaluation. The report does not cover self-evaluation of the Program-for-Results and the MAR and does not consider whether there should be (self) evaluation of areas such as governance arrangements and Board functions, human resources, and back office functions.<sup>13</sup>

#### EVALUATION QUESTIONS

The report seeks to answer the following questions:

- Are the Bank Group self-evaluation systems adequate to inform decision-making as it relates to operational performance management? See chapter 2.
- Are the Bank Group self-evaluation systems adequate to verify achievement of results and promote accountability for results? See chapter 3.
- Are the Bank Group self-evaluation systems adequate to support learning from experience? See chapter 4.
- How are organizational incentives, norms, culture, and practices shaping the production and use of self-evaluations? This is discussed in chapters 2-5.

## Methodology and Data Sources

This evaluation relied on diverse methodological approaches targeted to answer particular evaluation questions. Data collection methods were purposefully eclectic to match particular questions and to triangulate information. The report integrates findings that have been triangulated across all of these approaches. A range of information sources was used:

- A study of the constituent systems' historical evolution based on a desk review of relevant documents.
- A study of self-evaluation in five multilateral and bilateral development agencies, joint initiatives assessing the development effectiveness of the World Bank Group and comparator organizations; and good practice standards for self-evaluation. This was based on a desk review of documentary evidence from comparator organizations supplemented with 14 interviews with staff from the African Development Bank (AfDB), Asian Development Bank (ADB), United Kingdom Department for International Development (DFID), the European Commission, and Inter-American Development Bank (IADB).
- Analyses of how gender and citizen engagement – both areas with prominent corporate goals – are covered in self-evaluation reports and how that information is used, based on a review of a sample of reports and key informant interviews.
- A study of how impact evaluation is produced and used by the Bank based on academic studies, IEG's 2012 evaluation, databases, and triangulated interviews.
- A review of country program self-evaluations based on experience validating them, a desk review of key documents, and interviews with authors.
- A study of systems for learning lessons in IFC based on interviews, a review of lessons and related architecture, and an electronic survey of IFC staff to assess the collection, use, and incentives for using lessons from self-evaluation.
- An assessment of the quality of ICR lessons based on a random sample of ICRs and qualitative analysis of ICR review sections on lessons quality.
- Quantitative and qualitative analysis of the quality of project M&E, including econometric analysis of the links between project M&E and outcomes based on IEG's ICR reviews and the ICR review database.
- An estimate of the costs of running the systems.
- Reviews of: (a) IEG reports on learning, project M&E, the matrix organization, impact evaluation, self-evaluation, and select similar evaluations from other agencies; (b) self-evaluation reports and IEG validations; (c) scorecards and

**CHAPTER 1**  
**ASSESSING THE BANK GROUP SELF-EVALUATION SYSTEMS**

indicators tracked by the President’s Delivery Unit and by regular business monitoring reports; (d) guidance documents; (e) the reporting architecture for ASA; (f) select Board documents and presentations related to results measurement and M&E; and (g) correspondence between IEG and the Global Practices.

- Semi-structured interviews with 110 Bank Group managers, staff, and consultants.<sup>14</sup> Interviewees were stratified among self-evaluation authors; managers, directors, and regional chief economists (representing both the Regions and the Global Practices/Cross-Cutting Solutions Areas); M&E, gender, and social specialists involved in self-evaluation; operational quality staff; and key informants with knowledge in specialized fields such as gender or impact evaluation (see table 1.1 and Appendix F). Respondents were selected using a mix of random sampling (self-evaluation authors) and purposeful stratified sampling (most other categories). Seventy-eight of these interviews had the broad purpose of gathering data from people with first-hand experience of using or producing self-evaluations from diverse roles and perspectives (the remainder had more narrow purposes). The 78 broad interviews were semi-structured, used templates tailored to the role of the interviewee, and focused on specific systems, barriers to producing good self-evaluation, use of self-evaluation information, incentives, and more. The interview transcripts were coded using content analysis software (MaxQDA), resulting in a dataset that the team used to write this report.

**Table 1.1. Number of Interviewees and Workshop Participants**

	Bank	IFC	MIGA	Totals
<b>Interviews-Core Team</b>				
Managers	22	8	1	31
Staff	54	23	2	79
<b>User-Centric Workshops and Interviews</b>				
Managers	1	2		3
Staff	20	8	1	29
<b>IFC Lessons Learning Interviews</b>				
Managers		4		4
Staff		5		5
<b>Total World Bank Group</b>	<b>97</b>	<b>50</b>	<b>4</b>	<b>151</b>
<b>Total Partner Agencies</b>				<b>14</b>

- Four professionally facilitated user-centric design workshops drawing on design thinking and aiming to diagnose user experiences, motivations, and perceptions, and to develop prototypes of highly functional systems were held. The 32 participants were self-selected in that they chose to sign up after being invited.<sup>15</sup>

- Three game-enabled workshops were held with about 45 participants, also self-selected, and were followed by a facilitated discussion on incentives, motivations, and challenges underlying self-evaluation. The game simulated project planning, implementation, and evaluation.
- One focus group was held with eight experienced IEG validators, complemented by interviews and conversations with other IEG staff.

The findings on behaviors, motivations, incentives, and culture were triangulated using systems thinking (box 1.2) to diagnose the various pressure points and how they relate to each other (including pressure points associated fully or partly with IEG validation). Using an iterative process of analysis, mapping, and calibration, the team produced a simple analytical representation of how systems operate that this report draws upon extensively.

### **Addressing Potential Biases and Conflict of Interest**

IEG is an actor and stakeholder in Bank Group self-evaluation. It confirms or overrules ratings (see figure 1.1), has contributed to the design of systems, and is a frequent user of data from systems in its macro-evaluations and learning products.<sup>16</sup> This creates a perception of potential conflicts of interest that the team managed by:

- Being clear from the outset that this evaluation examines self-evaluation by operational staff and is not an evaluation of IEG or how IEG performs its validation functions (an external review of IEG commissioned by the Committee on Development Effectiveness [CODE] was released in August 2015).<sup>17</sup>
- Paying close attention to IEG's role in all its data collection and examining the interfaces between self-evaluation and IEG to mitigate any real or potential concerns that IEG in this evaluation would be blind to how its own work shapes incentives.
- Employing outside consultants for certain roles, such as analysis and coding of interview data and facilitation of focus groups.
- Presenting only findings that could be triangulated from multiple independent sources and did not appear to represent bias or self-serving positions by individuals.

Summing up, the methodologies used in this evaluation were geared to assess complex systems, their history and evolution, how they compare to systems in other development agencies, how they are being used, the quality of data, and to understand the perspectives and concerns of a wide range of people using or interfacing with those systems.

### Box 1-2. Applying Systems Thinking

Systems thinking and complexity science have made their way into evaluation approaches and methodologies with the realization that linear ways of thinking about processes of change have little relevance for assessing dynamic systems (Williams and Hummelbrunner 2011; Forss and others 2011; Befani and others 2015; Bamberger and others 2015). Complex systems are made up of numerous components and animated by the interactions of many actors. As practices of monitoring, self-evaluating, and validating have become commonplace in the Bank Group, these practices have become embedded in organizational processes, norms, routines, and belief systems. Self-evaluation systems in the Bank Group qualify as complex systems and understanding how they work and diagnosing why requires a systems perspective (Leeuw and Furubo 2008; Rist and Stame 2006; Hojlund 2014).

Bob Williams (2015) recommends looking at three aspects of complex systems:

- How the relationships between people engaged in a system affect behaviors and how these relationships are affected by context.
- How the range of perspectives that people bring to a particular system promote behaviors that influence how a situation unfolds.
- How people draw boundaries between what they consider valuable and what they consider invaluable and therefore tend to marginalize.

Peter Senge (2006) underscored that the voice of the practitioner is central to understanding complex systems. Jody Kusek and Ray Rist (2004) emphasized organizational, political, and cultural factors, and the imperative of understanding the need of end-users when building and sustaining results-based M&E systems. This evaluation was designed to elucidate some of the fundamental issues related to norms, implicit and explicit rules, values, and incentives. To this end, the evaluation engaged users and producers of self-evaluation through:

- Semi-structured interviews to discuss specific issues in-depth. Separate interview templates were used for team leaders/authors; quality reviewers/M&E specialists; and managers/directors geared to their respective roles.
- Games to engage users in a low-stakes, dynamic environment where their behaviors and attitudes could be observed in action, rather than discussed in the abstract (as in interviews).
- User-centric workshops to understand users' experiences and motivations and brainstorm with them on how to craft elements of highly functioning systems.