# Overview

# Behind the Mirror: Report on the Self-Evaluation Systems of the World Bank Group

### About this Evaluation

Self-evaluation—the formal, written assessment of a project, program, or policy by an entity engaged in that activity—lies at the heart of the World Bank Group's results measurement system and has been used to assess the outcomes of investments for 40 years. This evaluation seeks to assess how well the Bank Group's self-evaluation systems serve their expected purposes.

The Bank Group's self-evaluation systems cover most operational activities and include the:

- 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
- PCRs (Project Completion Reports) for IFC advisory projects
- PERs (Project Evaluation Reports) for MIGA guarantee projects.

These systems should be able to support:

- *Performance management* via data for evidence-based decision-making about projects, portfolios, policies, and strategies
- *Reporting* on project and portfolio results to support internal and external *accountability*
- *Learning* that leads to enhanced operational quality

Evaluation systems can be understood and analyzed at various levels; three levels are considered in this report:

- Templates and guidelines
- Business processes and data streams
- 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 the Independent Evaluation Group's (IEG) validation and evaluation functions, recognizing that these influence behaviors.

The evaluation aims to support ongoing efforts to enhance effectiveness, promote learning, foster the move toward a "Solutions Bank," and simplify processes. It complements and builds on other IEG reports, most notably *Learning and Results in World Bank Operations: How the Bank Learns* and *Learning* and *Results in World Bank Operations: Toward a New Learning Strategy*.

### Main Findings

The World Bank Group's self-evaluation systems have expanded since they started 40 years ago, and compliance with requirements is mostly strong. The systems mesh well with the independent evaluation systems for which they provide information and the systems have been emulated and adapted by other development agencies.

However, the self-evaluation systems primarily focus on results reporting and accountability needs and do not provide the information necessary to help the Bank Group transform into a "Solutions Bank" or develop learning to enhance performance as emphasized in its 2013 strategy. Information generated through the systems is not regularly mined for knowledge and learning except by IEG, and its use for project and portfolio performance management can be improved. The systems produce corporate results measures but need to produce value to staff and line management and to the primary beneficiaries of the "Solutions Bank"—client governments, implementing agencies, firms, and beneficiaries and citizens.

### Performance Management through Self-Evaluation

Bank management has put processes in place to monitor and manage operational quality and portfolio performance using a comprehensive system of cascading indicators, some of which draw on information from Implementation Status and Results Reports (ISRs) and Implementation Completion and Results Reports (ICRs). The information produced by this system is used in regular processes for performance management. Thus, management has access to, and makes use of, data that can track performance, identify problem areas, and foster corrective action.

The Bank's performance management system, built around the ISR, serves its purpose but can be used better. When the ISR works as intended, warning flags are raised at the right time, and teams and managers act on these flags, problem projects can be turned around and deliver results. Yet ISR ratings and indicators derived from them are not always precise because of weak project monitoring and optimistic reporting. The ISR would be more effective for early warning if team leaders had incentives to proactively acknowledge issues and raise risk flags. Better early warning needs to be combined with timely action. Many mid-term reviews occur late, as does remedial action to address identified problems because Bank and client procedures complicate and delay restructuring of Bank projects. The Bank may want to move toward more adaptive project management in which course corrections occur as frequently as needed, informed by relevant and timely monitoring data.

#### Evaluation Scope and Evidence Base

The report covers self-evaluation of World Bank operations (investments, policy-based support, knowledge and advisory services, impact evaluations, trust funds, and partnerships); International Finance Corporation (IFC) investment and advisory services; country programs; and, very selectively, Multilateral Investment Guarantee Agency (MIGA) guarantees.

The evaluation relies on diverse data sources and methodological approaches geared to assess complex systems. Data collection and analyses aimed to generate perspectives on the architecture and history of the systems, review specific constituent parts, and analyze behaviors, motivations, and incentives.

The team conducted semi-structured interviews with 110 Bank Group managers and staff, and 14 interviews with staff in partner agencies. Focus group discussions and game-enabled workshops also provided data for the evaluation. A number of background studies, including quantitative and content analyses of project performance data, a review of academic and evaluation literature, and institutional benchmarking, formed the backbone of the analysis.

The incentives in the Bank and IFC need to shift so as to reward teams for good M&E and identification and fixing of problems rather than pressuring teams on rates of disconnect and other quantitatively tracked indicators.

### Verifying Results and Promoting Accountability

Thanks to self-evaluation frameworks and data, the Bank Group is able to produce holistic and high-level corporate results reporting to the Board and externally that are easy to compare across time, contexts, and sectors. The design and operation of the systems adhere to relevant good practice standards, coverage is comprehensive, and many evaluation experts consider the Bank Group's systems to be as good as or better than those in comparable organizations.

Shortcomings remain in the project M&E systems that generate results evidence despite various initiatives to strengthen M&E and results orientation. For example, inadequate evidence on results is a factor in 70 percent of all downgrades, implying that, for some projects, weak M&E affects the degree to which ratings are an accurate measure of results.

Self-evaluation frameworks direct attention to impacts on citizens, but their implementation often results in mechanical tracking of citizen "participation" and gender "flags" but not of broader social outcomes and beneficiaries' voices.

IFC has sought to reform and reduce the scope of its results measurement and self-evaluation. Some stakeholders perceived a risk of erosion of the accountability function under the proposed reforms and arbitrating between different positions proved difficult in the absence of a policy or other guiding principles. There has been only limited progress toward systems that better meet learning and business needs yet maintain a credible level of accountability and the tone at the top of the institution has not been supportive of self-evaluation.

### Learning from Self-Evaluation

Having all operational units write substantive end-of-project reports is a noteworthy accomplishment that not many other organizations afford themselves, opening a vast potential for individual and organizational learning. In practice, however, knowledge from the Bank Group systems is rarely valued or used and there is little effort to extract and synthesize evidence and lessons or to inform operations. Staff are more likely to rely on tacit knowledge than on written information from the self-evaluation systems. There is some individual learning but few benefits of this learning accrue beyond the authors and, hence, the potential of the systems for organizational learning is unfulfilled.

Learning has taken a backseat to accountability. The systems' focus on accountability drives the shape, scope, timing, and content of reporting and limit the usefulness of the exercise for learning. If the self-evaluation systems had been set up to primarily serve learning, they would have been more forward-looking (how can we do better?), more selective (which projects offer the greatest learning opportunities?), more programmatic (are there synergies across activities and countries?), attuned to unintended positive and negative consequences, and more often done in real-time.

Support and guidance on writing and learning lessons is missing. Lessons are recorded but rarely used and too often of low quality: many of them are too generic, not sufficiently based in evidence, fail to recommend what specifically should be done differently in the future, or fail to address critical internal organizational issues. In the Bank Group's face-to-face culture, dialogue formats would likely help staff explore key findings and lessons and spur more learning.

Parts of the system not focused on corporate reporting, such as impact evaluations and other voluntary self-evaluations, are more valued by respondents. Impact evaluations are optional, seen as technically credible, invest in monitoring, are undertaken when there is a specific interest in learning, and regarded as a valuable tool to increase development effectiveness. Thus, when conditions are right, the World Bank Group has strong demand for evaluative learning and a robust ability to supply it.

### Unleashing the Potential of Self-Evaluation

The main reasons for the observed shortcomings lie in incentives and behaviors rather than templates and processes. Incentives created inside and outside systems, including through ratings and validation processes, are not conducive to conducting high-quality self-evaluation and most staff do not view the self-evaluation systems as a source of timely, credible, and comprehensive information. Staff engage with the systems with a compliance mindset where candor and thoughtful analysis of drivers of results and failures suffer.

The external panel review of IEG, which also reflected on larger systems beyond IEG's control, concluded "the current overall system and processes are broken…. Learning is not prioritized, accountability is mechanical and does not support necessary learning or continuous improvement…. Improving the self-evaluation system is key for the success of [Bank Group's] new strategy and for strengthening the basis for IEG's validation and review..." IEG has worked collaboratively with management in designing and operating the systems and must therefore share in the responsibility for the state of affairs.

This evaluation identifies three broad causes of misaligned incentives for writing and using self-evaluations: excessive focus on ratings, attention to volume that overshadows attention to results, and low perceived value of the knowledge created. The evaluation proposes five recommendations to address these issues.

### First Loop: Excessive Focus on Ratings

The planned reform of the ICR process, template, and guidelines is an opportunity to correct the incentives and signals surrounding self-evaluation, building on the heightened attention that management has started to pay to results frameworks. Staff perceive that the prevailing interpretation of the IEG/OPSC harmonized objectives-based approach to rating and validating ICRs limits the appetite for innovation and causes inflexibility for project management. Adaptability can be promoted through increased flexibility in project design that minimizes the need to amend legal agreements as well as through simplified Bank and client restructuring procedures. There is a need to promote more constructive interactions between IEG and operational departments over project validations without losing sight of IEG's accountability function. Something that would help with this would be a mechanism to flag up when unsuccessful outcomes are caused by major shocks outside the control of the Bank such as, for example, disasters, conflict, or economic crises. The harmonized ICR rating and validation guidelines give insufficient attention to beneficiaries' views and to unintended positive and negative consequences.

**Recommendation 1: Reform the ICR system and its validation to make it more compatible with innovation and course**

**corrections**. As the report explains, project teams should be able to change course faster and more often. The ICR system should better account for unintended positive and negative outcomes, beneficiaries' perspectives, and unforeseeable shocks in how results are measured and projects are rated (applies to the World Bank and to IEG's role in validation).

Measuring and rating project outcomes at closing against objectives stated at design years earlier has become a source of tension and perceived rigidity, given that the quality assurance of results frameworks at the time of project design is insufficient and that the options of restructuring and adaptive project management have not taken root.

**Recommendation 2: Help staff understand that project objectives pertaining to innovating, piloting, and testing are feasible and that projects with such objectives are rated appropriately, provided the project development objective and indicators are set in the right way** (applies to World Bank and IFC with implications for IEG).

### Second Loop: Attention to Volume Sometimes Overshadows Results

Demand from the Bank Group Board and management for knowledge and evidence to enhance development effectiveness has not been matched by a corporate learning culture. Managerial signals emphasize business volume more than they do results, performance, and good self-evaluation; tensions over ratings and disconnects distract from learning; and there is room to more consistently infuse existing learning, strategic, and planning processes with evaluative evidence. The Board has a role also to reinforce these signals.

**Recommendation 3: Strengthen rewards and leadership signals at all levels of the organization to reinforce the importance of self-evaluation.** For example, this can be done by promoting use of the knowledge generated from self-evaluations by teams, practices, and senior management, and by balancing the current excessive focus on outcome ratings and disconnects with more deliberative use of monitoring and self-evaluation information by teams and managers (applies to World Bank and IFC).

Identification of problems and solutions could be strengthened by having more reliable monitoring data and using that data more consistently in safe space deliberative meetings aimed at identifying and discussing problems. The M&E systems that generate the underlying evidence for results have long-standing shortcomings, despite various initiatives to strengthen M&E and results orientation. Strengthening M&E is especially important for projects with new or innovative designs and will also require building client M&E capacity in collaboration with partners.

**Recommendation 4: Formulate a more systematic approach to improving M&E quality.** As the report explains, this would entail building staff and clients' M&E capacity, demonstrating to clients the value of M&E, and provisioning of specialized M&E skills at key moments of the project cycle for targeted projects (applies to the World Bank and IFC).

### Third Loop: The Perceived Value of Knowledge from Self-evaluation is Low

Corporate requirements specify the scope, timing, and content of self-evaluations in a way that supports reporting more than it does learning. For example, most self-evaluations continue to be project-specific, with similar approach and depth, regardless of the learning potential. Mandatory and voluntary self-evaluations are not used strategically to meet knowledge gaps and approaches to using them for lesson learning are fragmented, further fueling staff perceptions of low importance. There is scope to strengthen Bank-wide oversight and the regional and thematic selectivity of impact evaluations, the uptake of findings from impact evaluations, and the use of information systems for capturing, classification, and availability of Bank Group mandatory and voluntary self-evaluations. IFC lacks a framework for capturing and acting on evaluative lessons.

**Recommendation 5: Expand voluntary evaluations that respond to learning needs of management and teams.** These include impact and process evaluations, retrospectives, and beneficiary surveys and need not be project-specific but can cover multiple interventions in a given sector, country, or region, depending on learning needs. Building on recent progress, further enhance the manner in which impact evaluations respond to learning needs through greater regional and thematic selectivity and enhance the uptake of findings from impact evaluations. Ensure that information technology systems capture and make accessible knowledge from self-evaluations (applies to the World Bank and IFC).