

Approach Paper Report on Self-Evaluation Systems (ROSES 2016)

February 24, 2015

Abstract: Self-evaluation systems are mechanisms for learning and mid-course correction that can help the WBG deliver its strategic goals and become a “Solutions Bank” that is focused on addressing complex development problems. This is an opportune time to take stock of the World Bank Group’s self-evaluation given the new strategic directions and the strong focus being placed on results, learning, and innovation. This proposed evaluation will review the World Bank Group’s self-evaluation systems against the objectives of the system and good practice standards and benchmarks and recommend improvements where necessary. Focus will be on (a) systems covering investment, knowledge, and advisory services; (b) integrity of the self-evaluation architecture, including achievement of accountability, performance management, and learning objectives; and (c) behaviors, incentives, and organizational norms and practices that shape how self-evaluation information is produced and used. By appraising the WBG’s self-evaluation systems, IEG is fulfilling a core part of its mandate and work program, intending to enhance operational and organizational effectiveness and building a learning culture in support of the ‘Solutions Bank’.

PART I: Motivation and Approach

Purpose, Objectives, and Audience

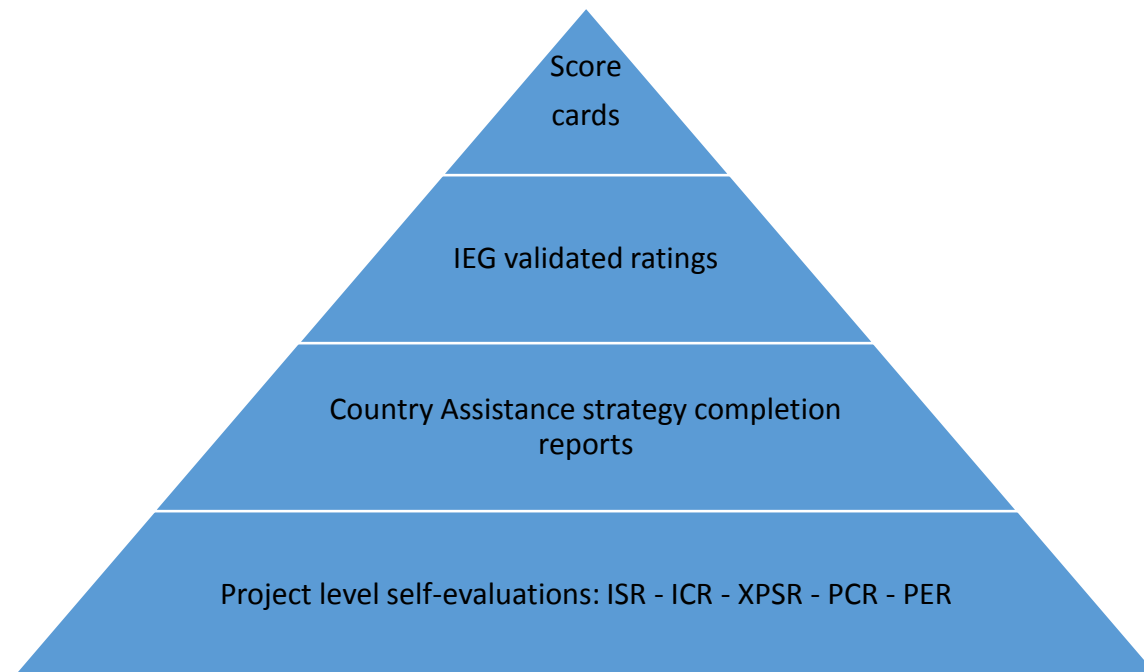
1. The World Bank Group (WBG) has set up a number of systems for measuring and assessing development results at project, country, program, and corporate levels. Self-evaluation by staff is an integral part of these systems; its stated purpose is to foster organizational learning; inform timely management action; and ensure accountability for results and performance.¹ The Independent Evaluation Group (IEG) validates some outputs from the self-evaluation system and these then feed into the corporate scorecards and other performance management systems (figure 1), signaling that the WBG holds itself accountable for achieving results.

2. Self-evaluation can play potentially useful roles at all stages of the project cycle, from approval through supervision, completion, and learning (figure 2). Although portfolio performance is affected by numerous factors, improving the performance of WBG self-evaluation systems is critical and would contribute to turning around the declining trend in portfolio performance and to become a “Solutions Bank” adept at learning and innovation. The self-evaluation architecture needs to be able to support:

¹ See for example OP 13.6 and OPCS (2011). This evaluation defines self-evaluation as the systematic, empirical, and transparent assessment of an ongoing or completed project, program, or policy, its design, implementation, and results done or overseen by someone who is engaged in the actual process (see definitions in box 2 below).

- Accountability – providing publicly available and trusted evidence of project performance;
- Performance management – providing data and information that can assist WBG management take decisions at the portfolio level relevant to improving performance;
- Learning – providing insight, at the project level, about performance, challenges met and addressed, and learning from experience, that can help project teams re-set their projects on an ongoing basis while also offering insights to others undertaking comparable projects.

Figure 1: Some self-evaluation information is aggregated in a cascading manner



3. This evaluation will be the first ever review of the entire WBG self-evaluation architecture.² Its **purpose** is to assess whether WBG self-evaluation systems are adequate to monitor and verify achievement of results, support learning from experience, and promote accountability for results. By doing so, the evaluation will support IEG’s larger objective to enhance operational and organizational effectiveness and building a learning culture, in line with IEG’s mandate.³ While IEG’s annual Results and Performance Report (RAP) covers trends and drivers of results, this evaluation will cover the systems (tools, methods, indicators, processes,

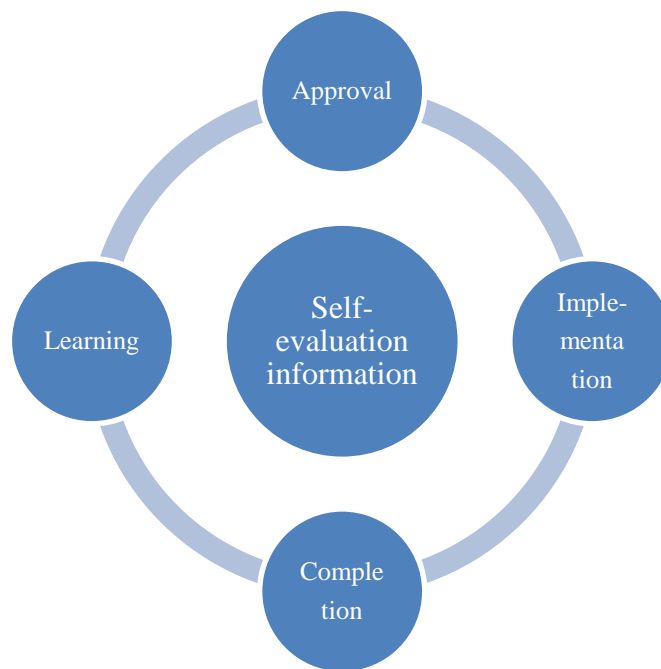
² In the past, IEG’s Biennial Report on Operations Evaluation (BROE) covered IFC (up to 2008) and IFC and MIGA in 2013, while the Annual Report on Operations Evaluation (AROE) covered IDA and IBRD and was published annually from 1998 to 2006.

³ IEG’s mandate lists “Appraising the World Bank Group’s operations self-evaluation and development risk management systems and attesting to their adequacy to the Boards” among the responsibilities of the Director General, Evaluation. See: https://ieg.worldbankgroup.org/Data/dge_mandate_tor.pdf.

data quality, and feedback loops) used by the entire World Bank Group to assess its results and performance.

4. Self-evaluation systems interface other major systems. They have to be understood in the context of the project cycle. Input data come from project M&E. IEG’s independent evaluation and ‘validation’ influence incentives. Self-evaluation information is aggregated in “dashboards” and scorecards, and used in other ways for learning and accountability (figure 3). The WBG’s self-evaluation architecture is thus made up of the totality of systems and scorecards that aggregate their information. The interfaces between systems, gaps in coverage, and the wider context are important aspects of the larger architecture that influence performance.

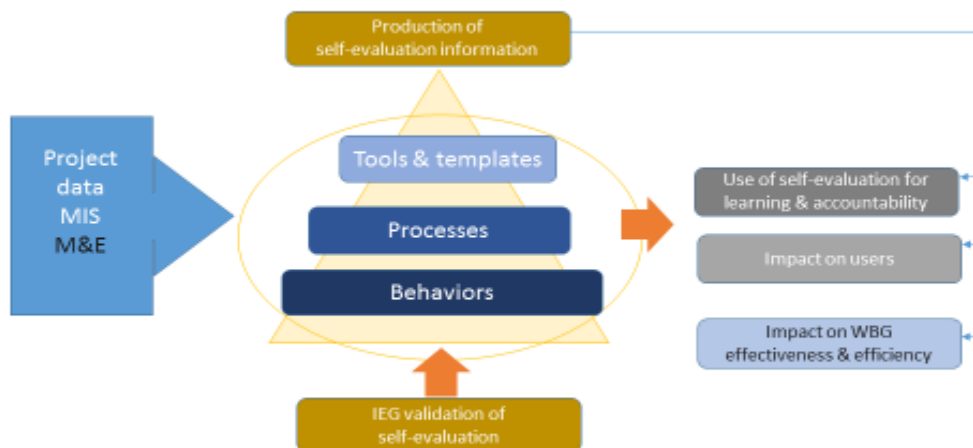
Figure 2: Self-evaluation is part of all stages of the project cycle



5. As explained below, the evaluation will look at broad questions of systems architecture and performance—and the incentives and organizational practices that shape how and why systems work. It will examine both the production and use of self-evaluation information. It will focus on self-evaluation of Bank and IFC investment projects, knowledge, and advisory services, as well as MIGA guarantees. This will involve, among other things: mapping out the systems in their entirety, evaluating the current state of practice, investigating the incentives and other drivers of performance, assessing the demands for self-evaluation information, including for learning purposes, examining the interfaces between self-evaluation and IEG’s independent ‘validation’ and evaluation functions, and developing recommendations for reforms—recognizing that while the business needs of the Bank, IFC, and the Multilateral Investment Guarantee Agency (MIGA) differ, there may also be scope for enhanced comparability of approaches and indicators.

6. The primary audience for this report is the Board, managers, and staff in central units responsible for design and operation of self-evaluation systems (such as Operations Policy and Country Services (OPCS), the Global Practices, and IFC’s Global Economics and Strategy), regional staff, including staff in regional Development Effectiveness and Country Management Units, staff in the results measurement and evaluation stream, IEG staff responsible for validating self-evaluation products, and self-evaluators. There is a potentially wide-ranging secondary audience comprised of other users of self-evaluation, in particular in the bilateral and multilateral donor community.

Figure 3: The chain to achieve better results through self-evaluation



Self-evaluation in the WBG

7. The Bank has expanded the number and coverage of its self-evaluations over the last 15 years. Results-based country assistance strategies (CASs) were piloted between 2003 and 2005 and mainstreamed since 2005, and the Bank stepped up M&E in lending significantly in the early-to-mid 2000s. The International Development Association (IDA) adopted a formal Results Measurement System in 2002, making the Bank the first multilateral development institution to use quantitative indicators to track results. Trust funds and partnership programs’ M&E were strengthened and, for trust funds, integrated in Bank systems in the late 2000s. The Development Impact Evaluation (DIME) program, set up in 2005, has helped scale up the use of impact evaluation in Bank projects. Knowledge activities have lagged behind, with only a skeleton template to self-evaluate results.

8. IFC uses results measurement to generate learning that feeds into IFC’s strategies and ongoing operations, promote transparency of IFC performance to external stakeholders, and provide a basis for rewarding good staff performance. IFC introduced its independent evaluation system in 1996, but prior to that, completed some ad hoc self-evaluations. Self-evaluation of advisory projects began in their current form in 2006. MIGA started self-evaluation in 2008. IFC results measurement systems have evolved in recent years, while also facing increasing pressure from IFC Management to focus on profitability. MIGA is updating its self-evaluation system and aligning it more closely with that of IFC.
9. IFC adopted a Corporate Scorecard in 2005 and the World Bank did the same in 2011 in an effort to consolidate performance indicators and ensure their routine use in decision-making. Similarly, MIGA introduced key performance indicators in 2009. The WBG is currently developing a scorecard that will capture information from the Bank, IFC, and MIGA.
10. The current coverage of self-evaluation in the WBG is described in Box 1, figure 4 and Attachment 3 and covers a mix of mandatory and voluntary self-evaluation. Some areas are not covered, for example innovative development finance and reimbursable services.
11. Information on the costs of self-evaluation is partial and incomplete. Bank Implementation Completion and Results Reports (ICRs) cost around \$35,000 each, corresponding to around 9 percent of total supervision cost over the life of projects. With around 300 projects closing annually, this translates into a cost of \$10.5 million per year. BROE 2013 estimates that IFC spends about \$14 million per year for core M&E activities, or 2.5 percent of IFC’s total administrative budget, and that MIGA spends around 1 percent on self-evaluation.
12. The new World Bank Group Strategy promises to enhance further the orientation toward achieving results, with implications for how self-evaluation should evolve in the future. The Strategy discusses how a “science of delivery” will boost results through a more “rigorous, scientific approach to development” and promises to scale up WBG efforts to capture and share knowledge of what works. To this end, the WBG has set up a results measurement community of practice and is introducing a new requirement for Performance and Lending Reviews for country programs and are making other changes in their evaluation systems and scorecards. At the same time, cost reductions are putting pressure on budgets for conducting self-evaluation and reinforcing the push for simplification and motivating strategic choices as to its coverage and mandatory nature. The WBG is also developing an internal results framework with the scorecard at its apex with elements cascading down into business unit and staff objectives (figure 1); while this new framework has not yet had implications for any of the self-evaluation systems, it suggests a potential for cascading systems and indicators. IFC, in consultation with IEG, has recently simplified its self-evaluation systems for investment projects and has increased reporting of its self-evaluation results in Annual Reports and Corporate Scorecards. MIGA has promoted use of self-evaluation for learning and is strengthening its internal guidelines.

Box 1. What is covered by WBG self-evaluation?

WBG self-evaluation covers operational activities; back office, financial, and managerial functions are covered in apex scorecards but to a more limited extent in primary self-evaluation.

Primary, mandatory self-evaluation systems include:

- ICR (Implementation Completion and Results Report) and ISR (Implementation Supervision Reports) for Bank lending
- CASCR (Country Assistance Strategy Completion Report) 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.

In addition, some WBG activities undergo **voluntary self-evaluation**:

- Impact evaluations for IFC investment and World Bank lending projects
- Evaluative studies, for example, IFC's program performance evaluations.

Aggregated self-evaluation results feed into apex reports that provide more aggregate, corporate accountability:

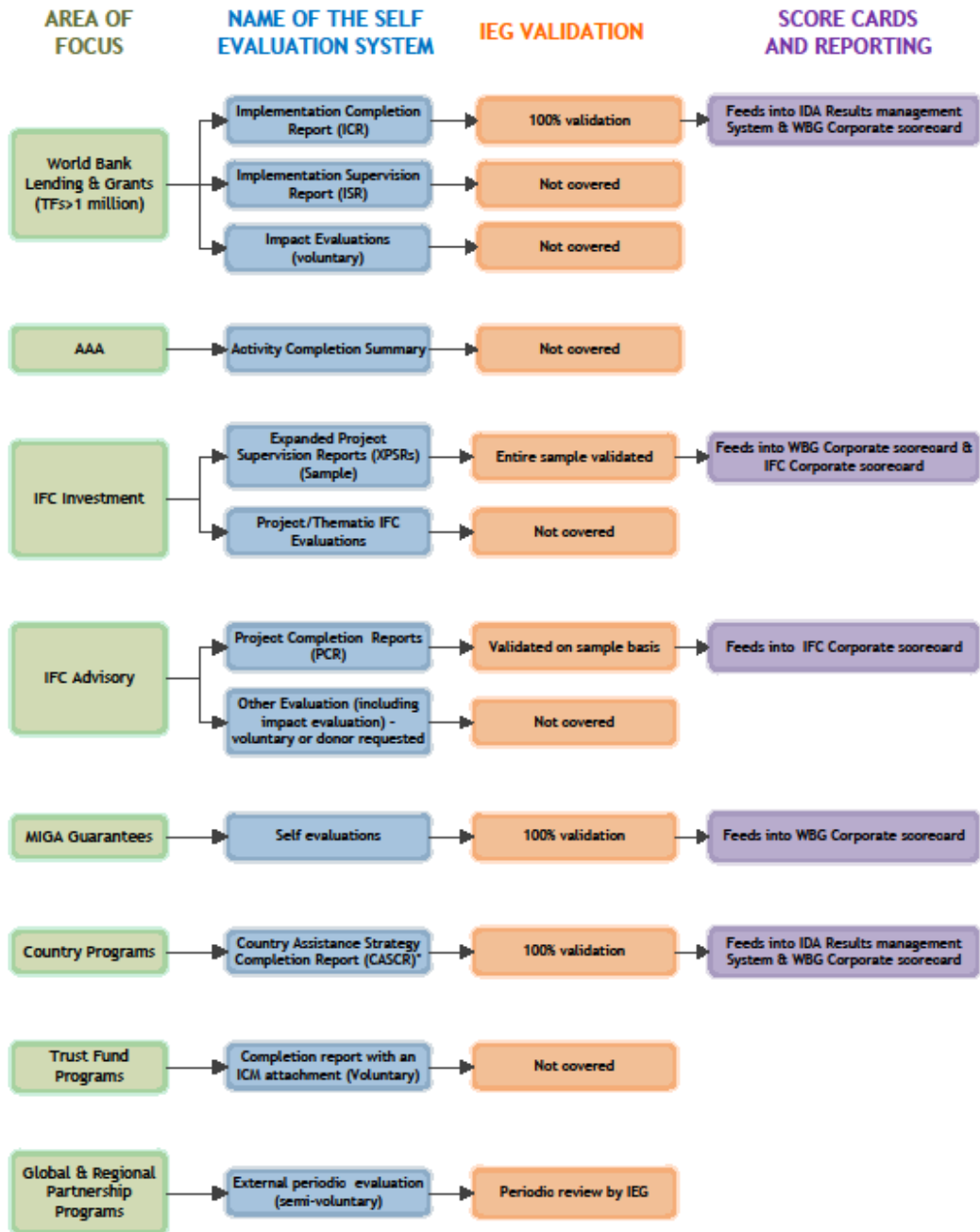
- IDA, IFC, and MIGA scorecards
- The WBG scorecard
- The new website by the President's Delivery Unit.

Some activities are **not yet covered** by self-evaluation, for example:

- Board operations
- Control functions
- Treasury operations and innovative development finance products such as green bonds
- Reimbursable services
- Various assessment tools such CFAA.

Figure 4 and Attachment 3 contain a more detailed inventory.

Figure 4. Major WBG self-evaluation systems



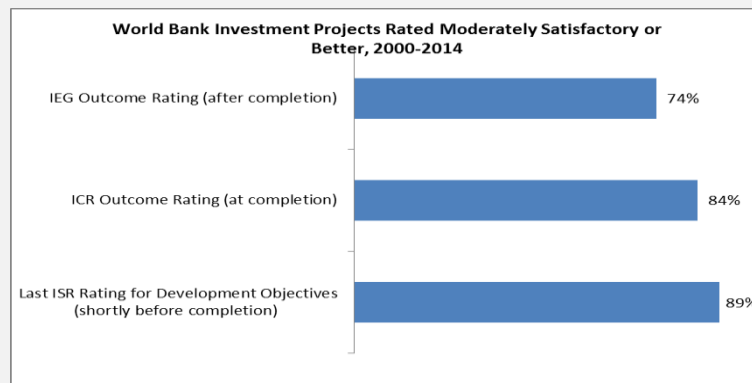
*IEG covers some of these more extensively in its other products beyond the routine validation captured here

Source: IEG staff.

Evaluation record on the state of WBG self-evaluation

13. The stated objectives of self-evaluation systems to foster accountability, performance management, and learning remain as important as ever and could possibly be enhanced. Aggregated project performance ratings are routinely used for reporting, including in corporate scorecards, and for analyzing trends and patterns in operational performance—a key use of self-evaluation information. In certain other areas, the use of self-evaluation information may be below potential. There may also be issues with the quality of the underlying information. Weak project M&E results in weak evidence on results in many self-evaluations. (Lacking or low quality of evidence in self-evaluation reports is one of the factors causing IEG to downgrade ratings, see figure 5). Indeed, across most IEG evaluations, weak M&E continues to be a consistent finding.⁴ And although WBG self-evaluation systems have expanded over the years, some gaps remain, including for Bank reimbursable, advisory services, and analytics.⁵ There is interest in the Bank and among external stakeholders in assessing results of knowledge services, which is critical given the move to a more knowledge-intensive business and the growing spending on knowledge products. While various management studies provide fragmented evidence on WBG’s self-evaluation systems, none provides in-depth and systematic analysis on the key aspects addressed by this evaluation, and none provides a detailed set of potential reforms to the systems. This evaluation attempts to fill these crucial gaps.

Figure 5. There is discrepancy between self-evaluation and independent evaluation ratings



Source: IEG staff based on Business Warehouse.

⁴ For example, IEG, Results and Performance 2013.

⁵ IFC has self-evaluation of Advisory Services but self-evaluation of Bank knowledge products, capacity building, and most Bank-executed trust funds is rather limited in scope and not validated by IEG. A client survey found good overall satisfaction with the quality of Bank knowledge work, but that does not substitute for evaluation. There are several obstacles to achieving evaluability in this area, including the difficulty of establishing logic chains from knowledge or TA to policy changes in client countries and the small size of many activities that make elaborate evaluation procedures uneconomical.

14. IEG’s learning evaluation concludes that, in the Bank, documentary sources of learning are less used than face-to-face sources. World Bank staff to a large extent learn about operational issues via conversations with peers, and do not make full use of written self-evaluation information. ISRs and ICRs are rarely consulted.⁶ The possibility of IEG overturning ratings may inhibit reflection and open dialogue on what works and why. Other evidence suggests that many ICRs are written by consultants, rarely shared among staff, written with a view to justify ratings and avoid IEG downgrades, and generally not used for learning.⁷ The Bank created “Intensive Learning ICRs” but these have been found to be broadly similar to regular ICRs in the type and depth of lessons and overall quality, and very few are conducted.

15. The Bank Group has been stepping up its use of impact evaluations. World Bank impact evaluations have advanced knowledge of the impacts of a large variety of interventions and been especially helpful in small-scale testing and determining whether results can be generalized. However, IEG concluded in 2012 that the feedback loop between impact evaluations and operations was not well developed.⁸

PART II: Evaluation Framework

DEFINITIONS

16. Self-evaluation is the systematic, empirical, and transparent assessment of an ongoing or completed project, program, or policy, its design, implementation, and results done or overseen by someone who is engaged in the actual process. For this evaluation, a “self-evaluation system” comprises tools, templates, work processes, incentives, and behaviors (figure 3). ICR, XPSR, PER, CASCR, and so on are examples of self-evaluation systems. These systems are more than the templates in which information is provided—the attendant processes and behaviors are integral to systems’ functioning. See box 2 for definitions.

Box 2. Definitions

- Self-Evaluation is the systematic, empirical, and transparent assessment of an ongoing or completed project, program, or policy, its design, implementation, and results done or overseen by someone who is engaged in the actual process. In the World Bank Group, self-evaluation is typically conducted by a contractor under the team delivering the program/project, or by the team itself.
- Self-evaluation system comprises tools, templates, work processes, incentives, and behaviors used to conduct self-evaluation.
- Results-Based Monitoring is a continuous process of collecting and analyzing information on key indicators to measure progress toward goals.

⁶ IEG, 2014a learning evaluation. Learning in IFC and MIGA has not been studied to the same degree.

⁷ Elliott 2013.

⁸ IEG evaluation of World Bank Group Impact Evaluations, 2012.

- Monitoring and Evaluation (M&E) is a combination of the continuous (monitoring) and the periodic (evaluation).
- Performance Management is the practice of public service managers using performance data, including data from M&E systems, to help them make decisions to continually improve services to their clients.⁹ In the World Bank, quality assurance of the lending portfolio is a key target of performance management, and ratings from self-evaluations and from IEG provide key inputs to quality assurance processes at various levels.¹⁰
- Organizational Learning is a continuous process of generating, accumulating, and actively using knowledge assets that support and enhance the organization’s ability to achieve its goals. Organizational learning rests on use of existing knowledge (exploitation) and creation of new knowledge (exploration). Organizations need to succeed at both processes in order to be successful.¹¹ It is a stated purpose of self-evaluation to contribute to organizational learning.
- Accountability is the obligation to report on agreed results and on adherence to established standards and processes. IEG could not find an official WBG definition of accountability; common usage in development emphasizes reporting both on results and on adherence to prescribed processes (see annex 4).

Source: IEG.

COVERAGE AND SCOPE

17. This evaluation will selectively investigate key aspects associated with the performance of World Bank, IFC, and MIGA self-evaluation systems. It cannot assess all elements of all systems in depth. Some of the activities that fall outside its scope include: peer review functions, Board operations, research, clients’ monitoring systems, safeguards, other compliance functions, GEF and other formally constituted partnership programs’ internal systems, and IEG’s self-evaluation. It will not cover operational monitoring more broadly (but will in a selective manner assess how monitoring information feeds into self-evaluation which could be critical for the quality of self-evaluation). Focus will be on (a) systems covering investment, advisory, and knowledge; (b) integrity of the self-evaluation architecture, including achievement of learning and accountability objectives; and (c) behaviors and incentives shaping how information is produced and used.

18. The evaluation will examine the interfaces between self-evaluation and IEG’s independent ‘validation’ and evaluation functions but will not provide a formal evaluation of IEG. Self-evaluation systems link up with IEG in multiple ways (figure 3). Monitoring and self-evaluation information is the foundation for independent evaluation. IEG ‘validation’ of self-evaluation—and the signals surrounding it—influences incentives for self-evaluation. More broadly, the nature of the relationship between operations and IEG may influence the utility of systems for learning purposes. IEG’s ‘validation’ functions will be considered alongside other

⁹ Adopted from Hatry 2014.

¹⁰ OPCS 2012.

¹¹ IEG Learning Evaluation.

drivers and incentives of systems performance but will not be formally evaluated. There is an ongoing external assessment of IEG commissioned by CODE that will address some of these issues and provide potentially useful findings. Potential conflicts of interest will be handled in the usual manner by declaring them and by recusing any team member who has or is working extensively on designing self-evaluation.

19. IEG is likely to continue assessing the WBG’s monitoring and evaluation beyond this evaluation. ROSES 2016 may be the first in a programmatic series. Decisions regarding topics to be covered in future evaluations will be made once this evaluation is complete; these could potentially include development policy operations, operational monitoring, detailed review of IFC systems (updating the BROE 2013), and self-evaluation of trust funds, partnership programs, and reimbursable services.¹² IEG will discuss self-evaluation of country programs in its process evaluation of SCD and CPF (due FY16).

DETAILED EVALUATION QUESTIONS

20. The overarching question for this evaluation is

“are the WBG self-evaluation systems adequate to verify achievement of results, inform decision-making, support learning from experience, and promote accountability for results?”¹³

Four specific questions elaborate:

- Do the WBG self-evaluation systems serve accountability purposes (cover the right things, provide timely, relevant, and accurate information)?
- Do the WBG self-evaluation systems inform operational and corporate decision-making?
- Do the WBG self-evaluation systems serve individual and organizational learning purposes?
- Is the WBG’s self-evaluation architecture effective, adaptable, and geared toward strategic needs? How conducive are the behaviors, incentives, and organizational norms

¹² To respond to the evaluative questions, this report will build upon rather than replicate work that was done in IEG’s Biennial Report on Operations Evaluation (BROE) 2013.

¹³ Refer to OP13.60 which frames the purpose of M&E in this way: “The Bank’s objective is to assist its borrowing member countries, individually and collectively, to reduce poverty and achieve sustainable 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.”

and practices shaping how information is produced and used and how learning takes place to achievement of objectives?

Accountability and Performance Management assessment

21. This evaluation will assess to what extent WBG self-evaluation supports accountability and performance management and how these might be strengthened. As explained in Attachment 4, no explicit World Bank definition of accountability could be found; common use of the term ‘accountability’ inside and outside the Bank emphasizes the obligation to report on--and to answer for--results and processes. During implementation, timely and candid reporting can help guide course-corrections and management action to enhance performance. After closing, reporting is used for aggregation of results, to inform stakeholders, and to learn lessons for future activities. The evaluation will examine:

- whether indicators are relevant and informative and other aspects of quality-at-entry and WBG work quality;
- whether reported self-evaluation information is candid, timely, and of quality;
- the extent and quality of reporting on gender and citizen engagement as key cross-cutting requirements;
- the desirability and feasibility of enhancing self-evaluation for Bank knowledge services;
- whether and how information is used to inform operational and corporate decision-making;
- the range of potential consequences that teams and managers face as a result of their reporting and how those potential consequences shape incentives and behaviors for staff and managers;
- the balance of reporting between: results; adherence to safeguards, fiduciary and other processes; and learning;
- incentives, norms, and organizational practices.

Learning assessment

22. The evaluation will assess if, to what extent, and how well WBG self-evaluation supports organizational learning and what reforms might strengthen this dimension by assessing how the WBG:

- Uses existing knowledge from self-evaluation processes for improving current operations—both for implementation and reassessing objectives. For teams, this would mean using the information for improving the current project/program; for managers, it would mean guiding decisions regarding projects/programs as well as providing direction regarding learning for the organization and ensuring the knowledge generated is made available for broader organizational use.
- Generates new knowledge for longer-term operational improvements, building on insights, and feeding lessons from the past into the design and implementation of new operations. For staff, it would imply using the knowledge as a base for future thinking

regarding the design and implementation of new projects/programs. For managers it would mean ensuring knowledge is used to enhance organizational effectiveness.

23. Organizational learning depends on the collective actions of all staff within the framework of organizational practices and norms that shape individuals' learning behaviors in support of the organizational goals. IEG's Learning Evaluation volume 1 found that the factors that support individual learning include: time for reflection; good knowledge management, including IT; access to interpersonal connections and networks that transfer tacit, contextualized and nuanced knowledge; doing team-based work; and budget in support of the above. Additional criteria that might affect learning from self-evaluation specifically which this evaluation will review include:

Substantive Issues

- perceived or actual credibility and usefulness of self-evaluation information for current or future operations (linked to the depth of information and suitability of the tools and approaches used);
- timeliness of the self-evaluation information for the job the staff member is doing or likely to be doing in the future;
- adequacy of IT, knowledge management, and dissemination.

Organizational practices and norms (figure 6)

- the degree to which self-evaluation information is used in meetings and for preparing projects;
- whether candor in self-evaluations is encouraged and rewarded or penalized;
- whether processes are geared toward compliance with tasks or with ensuring learning takes place;
- engagement in the self-evaluation process;
- whether individuals are held accountable for learning.

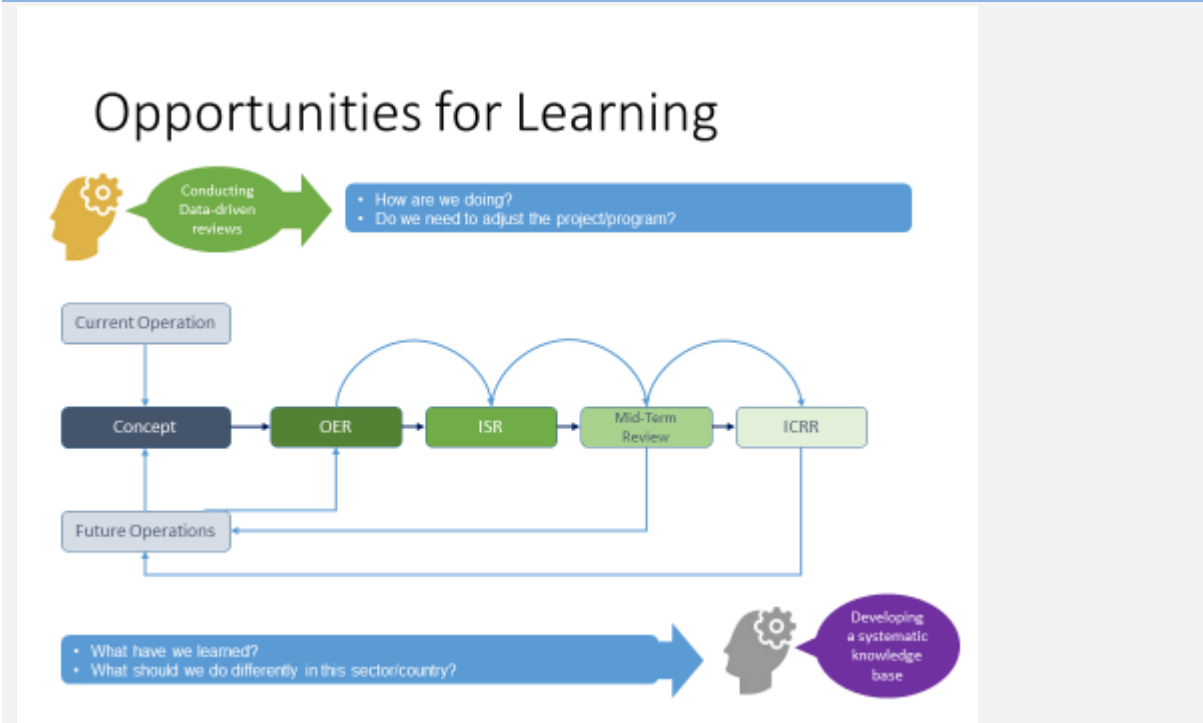
Assessment of self-evaluation architecture

24. The evaluation will assess the adequacy of the WBG's self-evaluation architecture. Is it effective, adaptable, and geared toward strategic needs? Can it respond to new demands and opportunities, and adapt as business needs evolve? Among other things, the evaluation will examine:

- gaps and interfaces between systems, including the way in which information is aggregated in scorecards;
- whether systems were able to respond to changes in demands in the past and are adaptable to meet future changes in the demands being placed on them;
- Incentives, norms, and organizational practices shaping how systems perform and are used;
- whether there is scope for harmonization of Bank, IFC, and MIGA indicators and approaches;

- whether the evaluations can address investment and advisory services in an integrated way; and
- the reformability of systems.

Figure 6. Self-evaluation can contribute to organizational learning under the right circumstances



25. Assessing the learning, accountability, and adaptability dimension of each of the major self-evaluation systems (ISR, ICR, XPSR, PCR, and so on) will allow ROSES to produce a “scorecard” of how well systems perform in different dimensions, including a narrative on their strengths, weaknesses, complementarity, and overlaps.

26. The evaluation will compare self-evaluation in the Bank, IFC, and MIGA. Because their systems have evolved in different directions—responding to different business needs—some aspects of systems have relevant counterfactuals in other parts of the WBG. These counterfactuals may offer evaluative evidence on how design choices shape incentives and system performance. For example,

- The WBG has both mandatory and voluntary self-evaluation. Mandatory self-evaluation relies on pre-determined templates and methodologies and contains ratings. Voluntary evaluative studies (including Bank and IFC impact evaluations) use a wider range of methodologies and do not contain ratings;
- IFC uses standardized indicators while the Bank tends to use diverse, project-specific indicators;

- IFC self-evaluation covers Advisory Services while the Bank struggles with how to assess its knowledge services;
- While several templates cover quality-at-entry and WBG work quality, there is no ability to compare and aggregate these ratings across WBG products and institutions;
- IFC has a “long-term performance award” that provides a feedback from evaluation to explicit staff recognition; the Bank does not have specific feedback from project results to staff career progression.

Evaluation Design and Evaluability Assessment

27. Many existing studies shed light on the quality, coverage, and consistency of evidence used in self-evaluation reports and their use in performance management, allowing ROSES to draw on and synthesize a wealth of information. This includes, for example, an ongoing assessment of the quality of evidence in ICRs led by IEGPS; a AAA evaluability assessment done by IEGCC; benchmarking documents such as the COMPAS report (IFAD 2014) and the Evaluation Cooperation Group (ECG) good practice standards (which addressed the evaluation system of the evaluation offices of the multilateral development Banks); various OPCS reports on portfolio performance and quality of the ICR system; the Internal Audit Department’s (IAD) ongoing Advisory Review of the information quality supporting the Bank’s portfolio monitoring; and IAD’s Advisory Review of the previous risk framework (ORAF). The evaluation will also draw on IEG evaluations such as Learning-in-lending, Results and Performance, BROE and AROEs.

28. The evaluation will collect new data, using the following tools:

- Semi-structured interviews (at least 80 interviews) with WBG authors and users of self-evaluation information representing all major user categories (line managers, Board members, etc.) and some secondary users (IFC and MIGA clients, NGOs, governments). Author interviewees will mostly be sampled in a systematic random fashion (drawing from lists of recent authors), while most users (where groups are smaller) will be purposefully sampled. Interviews will also cover country offices and regional hubs;
- Stakeholder participatory activities (see below);
- Assessment of gender coverage and indicators; review of the extent of citizen engagement; download statistics; expenditure data; review of self-evaluation benchmarks and practices in comparator organizations (such as the Asian Development Bank and the European Commission); study of roles, responsibilities, and the changing context; review of self-evaluation coverage by administrative budget categories; and case studies of good practices from inside the WBG.

29. The evaluation will analyze new and existing information using a multiplicity of methods, including: Statistical analysis of ratings and project performance, including regressions on factors driving outcomes; qualitative text-based analysis (most likely focused on text contained in ISR, ICR, and XPSR) as well as IEG data bases; assessing and benchmarking system features against evaluative criteria; Qualitative Comparative Analysis (QCA), a case-

based qualitative technique to link features of self-evaluation systems to their performance; assessment of future needs and opportunities; and case studies of instances where business units used self-evaluation in novel or innovative ways to enhance results.

30. Stakeholder participatory activities will be used extensively and prominently in this evaluation as part of its data collection and outreach. Staff and managers from all parts of the WBG will be invited to take part in participatory sessions exploring (a) the set of incentives surrounding self-evaluation; (b) drivers and obstacles for changing systems; and (c) eliciting suggestions for reforms. These activities will thus be integral to data collection efforts. They will also help ensure stakeholder engagement, learning, and acceptance of self-evaluation reform. By sharing information and consulting frequently with the evaluatees, the evaluation aims to both tap into the extensive tacit knowledge of stakeholders and to make sure that recommendations are implemented through the desire to improve and not through carrot and stick approaches. A range of activities are planned:

- Three prototyping one-day workshops inspired by User-Centric Design focused on rapid generation of options for improving self-evaluation systems and information on the feasibility and constraints to adoption of the more promising options. Participants would be selectively chosen staff and managers;
- Occasional meetings with an informal ‘sounding board’ group to be convened by OPCS to facilitate two-way information flow and relevance of approaches and recommendations. The expected membership would include OPCS, GP/CCSAs, IFC, and MIGA;
- Game-enabled two-hour sessions using a board game-type exercise to allow participants to conduct and experience a stylized version of self-evaluation and facilitate dialogue on potentially sensitive topics. The game will take participants through the process of designing, implementing, and evaluating a stylized project. The post-game debriefing will emphasize issues around skills, capacity, incentives, and how to improve evaluation. Participants will be members of the Results Measurement Community of Practice, self-evaluation authors, and IEG staff and consultants;
- Workshops and seminars targeting the WBG community interested in performance management and M&E;
- Participatory “campaign” style approaches to feed into learning and to share the results of the report;
- A dedicated SPARK site will support these activities.

31. Some information is unlikely to be available and will limit what this evaluation can do. Fully assessing cost-effectiveness, for example, will likely be infeasible. Preliminary research by the team indicates that data on what it costs to run self-evaluation systems is partial and incomplete. The team also does not expect to be able to estimate the benefits of self-evaluation. Confidentiality makes it unlikely that the ROSES team can obtain human resource data on how use and production of self-evaluation features in staff performance reviews.

Quality Assurance Process

32. The evaluation will be overseen by Nick York, Director, IEGCC and Geeta Batra, Manager, IEGCC. Peer reviewers for the Approach paper are Aart Kraay, Sr Advisor, DEC; Patricia Rogers, Professor of Public Sector Evaluation, RMIT University, Australia; and Preeti Ahuja, Manager, Development Effectiveness, MENA. An additional peer reviewer from a Global Practice will be added for the draft final report. Two eminent evaluators—Ted Kliet and Nils Fostvedt—act as advisors to the team.

Expected Outputs and Dissemination

33. The main output will be a report of approximately 100 pages including an overview and supported by annexes as needed. The report will be accompanied by at least one animated video presenting the analysis in a synthesized and easily digestible format. Dissemination will be designed as continuation of the participatory components mentioned above and is expected to emphasize targeted outreach to segmented constituencies inside and outside the Bank Group.

Resources and timeline

34. The proposed budget is \$900,000 (of which \$570,000 in FY15) including dissemination.

35. *Timeline.* Approach paper of the evaluation will be completed by FY15Q3. Evaluation work and drafting of evaluation report to be completed by FY16Q1. Final report will be presented to CODE in FY16Q3. Internal and external outreach to be completed by FY16Q4.

Attachment 1

Bibliography

- Elliott, Victoria. 2013. Implementation Completion Reporting: Review of Experience and Directions for Change.
- Geli, Patricia, Aart Kraay, and Hoveida Nobakht. 2014. Predicting World Bank Project Outcome Ratings. Policy Research Working Paper 7001
- Hatry, Harry P. 2014. Transforming Performance Measurement for the 21st Century. The Urban Institute.
- IEG. 2009. Enhancing Monitoring and Evaluation for Better Results: Biennial Report on Operations Evaluation in IFC 2008.
- IEG. 2012. World Bank Group Impact Evaluations: Relevance and Effectiveness.
- IEG. 2013. Assessing the Monitoring and Evaluation Systems of IFC and MIGA: Biennial Report on Operations Evaluation 2013.
- IEG. 2014a. Learning and Results in World Bank Operations: How the Bank Learns.
- IEG. 2014b. Results and Performance of the World Bank Group 2013.
- IFAD. 2014. Multilateral Development Banks' Common Performance Assessment System COMPAS 2012.
- OED/IEG. Annual Report on Operations Evaluation (multiple years, 1998-2006).
- OPCS (Operations Policy and Country Services). 2011. Implementation Completion and Results Report: Guidelines. August 2006 updated 2011.
- OPCS (Operations Policy and Country Services). 2012. Delivering Results by Enhancing Our Focus on Quality.
- World Bank Group Corporate Scorecard, April 2014.

Attachment 2

Specific evaluation questions, sub-questions, and data sources:

Overarching question: “are the WBG self-evaluation systems adequate to monitor and verify achievement of results, support learning from experience, and promote accountability for results?”

Question	Data sources
<p>Q1: Do self-evaluation systems serve accountability purposes (cover the right things, provide timely, relevant, and accurate information) and inform operational and corporate decision-making?</p> <p>-Do self-evaluation systems cover the right things, and do they get the balance right between reporting on results and on processes?</p>	<p>Interviews, document reviews, gender coverage and indicator assessment, review of citizen engagement</p>
<p>-Do self-evaluation systems provide timely, candid and accurate information?</p>	<p>Quantitative and qualitative analysis of self-evaluations and IEG validations, including IEG ratings of the quality of ICRs; interviews</p>
<p>-What incentives and other factors influence the quality of information?</p>	<p>Interviews and participatory sessions</p>
<p>-Do self-evaluation systems inform operational, corporate, and strategic decision-making?</p>	<p>Interviews and participatory sessions</p>
<p>Q2: Do self-evaluation systems serve individual and organizational learning purposes?</p>	
<p>-How, and for what purposes, do staff and managers use self-evaluation information?</p>	<p>Interviews, download statistics, IEG learning evaluations, and participatory sessions</p>
<p>-What incentives and organizational processes support learning from self-evaluation?</p>	<p>Interviews; case studies</p>
<p>-What are the barriers to learning from self-evaluation?</p>	<p>Interviews; case studies; re-analysis of data from learning evaluation</p>
<p>Q3: Is the WBG’s self-evaluation architecture effective, adaptable, and geared toward strategic needs?</p>	<p>Assessment of comparators and benchmarks, synthesis of external reviews, review of roles and responsibilities</p>

<p>-Have systems changed in the past as needs evolved and can they do so in the future?</p>	<p>Interviews, document and literature review, and participatory sessions</p>
<p>-Do they meet critical requirements associated with the “Solutions Bank”, the twin goals, environmental sustainability, and SDGs?</p>	<p>Literature review, interviews, participatory sessions</p>
<p>-Is there scope for harmonization of Bank, IFC, and MIGA indicators and approaches?</p>	<p>Review of comparator organizations; interviews;</p>
<p>-What stands in the way of moving toward more ideal systems and using data and information technology to their fullest potential?</p>	<p>Participatory sessions</p>

Attachment 3

World Bank Group Self-Evaluation Tools and Systems

Activity level	Area of focus	Name of the self-evaluation tool (acronym)	Who is responsible for its preparation?	Frequency and coverage	IEG Role & validation coverage	Purpose of self-evaluation	Process details
Project	Lending	Implementation Completion Report (ICR)	TTL/Implementing Team	After project completion for all projects	<ul style="list-style-type: none"> •IEG validates 100% of the ICRs produced. •IEG selects about 20–30 percent of evaluated and reviewed projects for field reassessments (Project Performance Assessment Reports (PPARs) for the World Bank) 	Accountability and Learning	The Regions are responsible for completion reporting; for the quality of ICRs submitted to the Board and, for the processes needed to ensure quality and timeliness (including providing resources). The Country Director allocates funds for ICRs annually in Business Plans/Work Program Agreements. The ICR production process is complete when a final report is approved by the Country Director, submitted to the Board, and disclosed to the public. (OPCS ICR Guidelines; last updated 2011)
		Impact Evaluations	TTL	Ad hoc	IEG does not review impact evaluations	Accountability and Learning	-
	Advisory services and analytics (ASA)	Completion Summary and Results Completion Summary	TTL/Implementing Team	Within six months after delivery to client	Knowledge activities are not systematically covered, but are sometimes covered in sector and thematic, evaluations, country program evaluations, and knowledge evaluations.	Accountability and Learning	
	IFC Investment	Expanded Project Supervision Reports (XPSRs)	Investment staff prepare Expanded Project Supervision Reports (XPSRs) for a random, representative sample of projects	IEG selects a random sample, which covers projects approved five calendar years prior to the current year and have generated at least 18 months of operating revenues (covered by at least one set of company annual audited accounts)	IEG undertakes an independent review of the project's performance and the XPSR's assigned ratings (and adjusts them if needed) to ensure that the prescribed evaluation guidelines and criteria are applied consistently	Accountability and Learning	
		IFC thematic evaluations			No validation	Learning	

Activity level	Area of focus	Name of the self-evaluation tool (acronym)	Who is responsible for its preparation?	Frequency and coverage	IEG Role & validation coverage	Purpose of self-evaluation	Process details
		including Impact Evaluations					
	IFC Advisory	Project Completion Reports Thematic and product assessment (PCR)	Project teams produce the PCRs as the final monitoring report	At completion, for all Advisory Services projects, unless they were dropped or terminated, and are due within three months of project closure	IEG assesses project success and the quality of documentation and summarizes its views and ratings in an Evaluative Note. IEG currently reviews a random sample (51 percent three-year rolling average) of projects closed in the previous fiscal year. IEG's assessment is a desk review of project documents and other sources including any external evaluations. IEG also makes selective field validations.	Accountability and Learning	After PCRs are completed, IFC's Development Impact team (CDI) reviews them to determine whether the project team's self-ratings are supported by evidence and conform to IFC's M&E framework. The department assigns its own project ratings, which become the official IFC rating for all reporting, including IFC's Annual Report and Corporate Scorecard. IEG reviews sample of PCRs and independently assign ratings. CDI and IEG ratings were constantly lower than the original PCR self-ratings. IEG does not validate CDI's ratings, and CDI's ratings are not formally integrated in the PCR self-evaluation system documents.
		Thematic evaluations including Impact Evaluations	Often led by Advisory Services staff but contracted out to external parties.	Ad hoc — based on availability of funding, project team, or donor interest, without a strategic selection framework	IEG does not review impact evaluations	Learning	
	Donor funded AS programs (IFC)	Mid-term and completion evaluation	AS program team with IFC CDI-AS, often commissioned to outside firms	Ad hoc	IEG does not review these evaluations	Accountability and Learning	
	MIGA	Self-evaluations	Self-evaluations are conducted by operational staff in order to emphasize learning	Annually for all eligible, operationally mature projects	IEG independently validates MIGA self-evaluations, based on guidelines developed together with MIGA. In addition, in a transition phase, IEG continues to independently evaluate a sample of guarantee projects.	Accountability and Learning	



Activity level	Area of focus	Name of the self-evaluation tool (acronym)	Who is responsible for its preparation?	Frequency and coverage	IEG Role & validation coverage	Purpose of self-evaluation	Process details
Program	Trust Fund activities (IDA/IBRD)	Implementation Completion Report (ICR)	TTL	After activity completion if trust fund ≥ \$1 million	100% IEG validation of ICRs regardless of the source of finance	Accountability and Learning	IEG also reviews trust-funded activities as part of its CAS Completion Report (CASCR) Reviews, Country Assistance Evaluations (CAEs), and sectoral/thematic reviews (Trust Fund handbook). IDA grants are processed in the same way as IDA credits and are disbursed, monitored, and evaluated in accordance with regular procedures for IDA credits.
	Trust Fund grants (RETFs)	Implementation Completion Report (ICR)	TTL	After activity completion if ≥ \$1 million	100% IEG validation of ICRs regardless of the source of finance	Accountability and Learning	
	Country Program	Country Assistance Strategy Completion Report (CASCR)*	Prepared by the country management unit	Within six months after the end of the previous CAS/CPF period	IEG conducts a desk review of all the CASCRs to validate the self-evaluation. CASCRs are also reviewed in the context of Country Program Evaluations.	Accountability and Learning	
	Sector Program	None	Sector Board	Ad hoc	Sector and thematic reviews	Learning	
	Trust Fund Programs	Completion report through Grant Reporting and Monitoring (GRM)	TTL	Ad hoc, often annual		Accountability and Learning	ICM is required for all funds for which an Initiating Brief for Trust Fund (IBTF) or a Trust Fund Proposal (TFP) was prepared at the initial stage — with total contributions of \$1 million or more. (Trust Funds Website) For grants or activities within a programmatic trust fund, and for trustee-level funds for which total contributions are below \$1 million, TTLs should use the GRM tool to report their assessment of trust fund achievements. For other programmatic trust funds that have disbursed over \$5 million, the TTL of record and the donor(s) agree on the timing and procedures for independent evaluations. These cover both the program and activity levels and are conducted in accordance with the OECD/DAC Evaluation Principles and Standards.
Corporate	Global and Regional Partnership Programs	Independent/external periodic evaluation	Governing body of the GRPP/TTL	Once every five years	IEG used to periodically review a sample as part of its Global Programs Reviews	Accountability and Learning	
		IDA Results Measurement System	OPCS	At the end of a replenishment period	No role for IEG		



Activity level	Area of focus	Name of the self-evaluation tool (acronym)	Who is responsible for its preparation?	Frequency and coverage	IEG Role & validation coverage	Purpose of self-evaluation	Process details
		Corporate Scorecard	OPCS	Annual	IEG Reports on the Corporate Scorecard in the RAP	Accountability	
		Corporate Scorecard (IFC)	IFC Strategy Department	Annual	IEG Reports on the Corporate Scorecard in the RAP	Accountability	
		Corporate Scorecard (MIGA)	MIGA economics Department	Annual	IEG Reports on the Corporate Scorecard in the RAP	Accountability	

Attachment 4: Accountability

Definitions of ‘accountability’ vary but often focus on the obligation to report on agreed results and on adherence to established standards and processes, and sometimes also to answer for any deviations. OECD defines accountability as “Obligation to demonstrate that work has been conducted in compliance with agreed rules and standards or to report fairly and accurately on performance results vis-à-vis mandated roles and/or plans....Accountability in development may refer to the obligations of partners to act according to clearly defined responsibilities, roles and performance expectations”. The International Federation of Red Cross & Red Crescent Societies (IFRC) also defines accountability with respect to both results and processes, including having systems for learning: “The obligation to demonstrate to stakeholders to what extent results have been achieved according to established plans. This definition guides our accountability principles....: explicit standard setting; open monitoring and reporting; transparent information sharing; meaningful beneficiary participation; effective and efficient use of resources; systems for learning and responding to concerns and complaints”.

IEG could not find an official WBG definition. OPCS’s guidelines for ICRs describe the purposes of the system as to “provide a complete and systematic account of the performance and results of each operation.....capture and disseminate experience...provide accountability and transparency ... with respect to the activities of the Bank, borrower, and involved stakeholders”. The ICR template covers, among other topics, results, efficiency of resource use, fiduciary and safeguard compliance, Bank and borrower performance: again, both results and adherence to prescribed processes are emphasized. The implicit definition in the OPCS guidelines (and areas covered by templates) would seem broadly consistent with OECD and IFRC definitions.