5. Conclusions and Recommendations

This evaluation set out to assess whether the operational self-evaluation systems of the World Bank Group are suited to their stated purposes. The evaluation found several positive aspects: 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 as good as or better than those in comparable organizations. The systems produce corporate results measures that are easy to report externally and to compare across time, contexts, and sectors. Guidelines and review processes exist, and there is ongoing, process-driven use of the information generated for performance management and accountability. The systems mesh well with the Bank Group's independent evaluation systems for which they provide information. Compliance with requirements is mostly strong. Stakeholders have unparalleled access to the ratings, self-evaluations, and validation documents.[1] Staff and managers engage seriously and responsibly, and considerable resources go into feeding and using systems (a low-end estimate puts the cost of producing self-evaluation at $15 million, 0.6 percent of the Bank Group's annual administrative budget).

Yet the emphasis in the 2013 World Bank Group Strategy on developing a "Solutions Bank" and learning to enhance performance is not well-served by existing self-evaluation systems. Information generated through the current systems is not systematically mined for learning except by the Independent Evaluation Group (IEG) and use of the systems for project and portfolio performance management can be improved. The focus on corporate results reporting for accountability has sidelined use of the systems for these other purposes.

Some of the shortcomings identified by this evaluation are inherent in the design of the systems, others relate to how they are used. The systems are mostly project-focused, objective-based, and geared toward accountability ("did activities achieve their stated objectives?"), and thus have built-in limitations for driving performance ("what needs to change so that we can deliver better for clients?") or generating learning ("what worked well and what could we have done better?"). Also, using results-based management systems blindly can lead to excessive focus on simple outputs and underinvestment in complex, long-term strengthening of client systems. Finally, ratings are a useful part of the systems but tensions associated with IEG's rating validation process are unnecessarily prominent and distracting.

In economics, it is well-established that multiple goals cannot be achieved with a single instrument. The same applies to self-evaluation. In the current organizational environment, it is unrealistic to expect that self-evaluation systems can simultaneously and fully deliver on performance management, robust measurement of results for accountability, and learning. There are trade-offs among these objectives that have been insufficiently recognized and, in practice, the main thrust has been on results measurement for external reporting.

Evolution of the Self-Evaluation Systems

The Bank Group has not had a coherent approach to how, how often, and in what direction systems ought to evolve. Several documents establish the current expectations for the systems as encompassing support for performance management, accountability and rigorous measurement of results, and learning, but no single document sets out guiding principles or priorities. The 2013 Strategy adds an ambition of linking evaluation to the institution's twin goals, which are yet to be achieved. The International Finance Corporation (IFC) has expressed a desire to reform its monitoring and evaluation (M&E) system to better meet its learning and business needs but reconciling this with the reporting and accountability functions provided by the existing systems proved contentious. The Bank has simplified the Implementation Status and Results Report (ISR), whereas the most recent major change to the Implementation Completion Report (ICR) was in 2006.

There has been talk about integrating the diverse results measurement systems in place across the Bank Group institutions and product lines. Doing so would be misguided. Already, corporate results reporting overshadows other purposes so that information from systems is less useful and less used for performance management and learning. Different product lines differ in their information needs and, to be relevant and useful, systems should respond to these needs in the first place. Also, the International Development Association (IDA) needs an IDA-specific results framework for demonstrating its results.

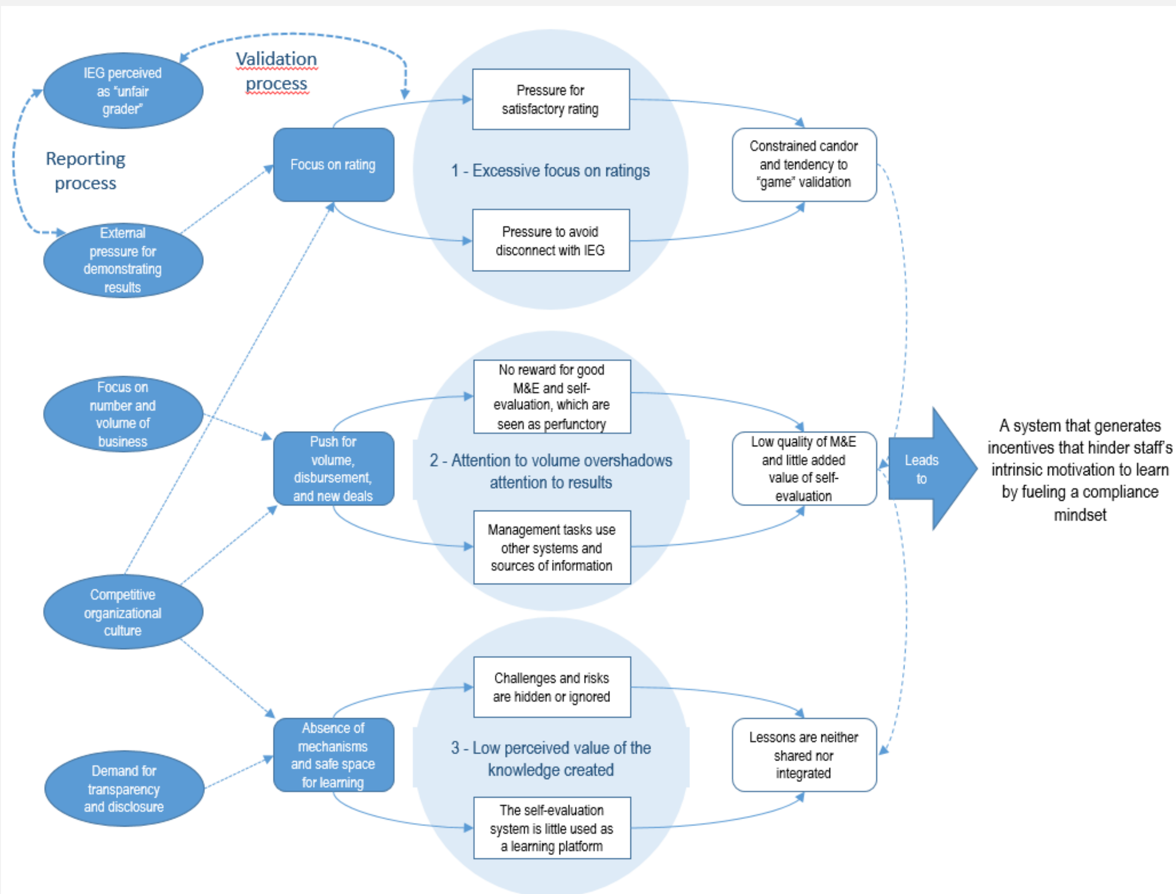
Mapping Behaviors and Incentives

Key groups of people engage with the systems in ways that are fueled by a compliance mindset more than a learning mindset. Levels of frustration and mistrust are high, and many perceive systems to add little value. The systems map (figure 5.1) illustrates in three loops the ways in which behaviors and incentives for staff and managers constrain the usefulness of the systems:[2]

- There is excessive focus on ratings in how the systems are used, exacerbated by a competitive organizational culture. This can make staff focus on avoiding negative ratings and disconnect and can limit candor and lead to attempts at "gaming the process," making results reporting less than fully accurate (first loop).

- Attention to volume overshadows attention to results. The push for new deals, lending, and disbursements displaces incentives to invest in M&E and, without good data, systems create little value and are only partially used for project performance management. Thus, many managerial tasks rely on other data and occur outside the systems (second loop).
- The perceived value of the knowledge created is low, too many risks and failures are hidden, safe spaces to learn from failure are missing, lessons and knowledge are not mined, and systems therefore create little organizational learning (third loop).

Figure 5.1. Behaviors, Incentives, and Motivations



Interactions with systems need to more often trigger reflection, course correction, and learning and less often trigger frustration and mechanical reporting. The user experience for staff must improve (box 5.1). Interview respondents from across the

Bank Group characterized the self-evaluation processes as an elaborate architecture to "feed the bureaucratic beast" with data that add little value. Staff did not understand how management and the Board use information produced by the systems. Consistent with the external panel review of IEG (see box 3.4), people who were interviewed or participated in focus groups were eager for reform that, in their view, should not result in additional work pressure and complexity and needed to address incentives. As one manager said: "Self-evaluations are only as good as the intention, candor, and use to which they are put. Systems may change at the margin, but unless signals and other factors change, not much will improve."

Staff is not the only group for whom systems fail to produce much value. The main focus is on the Board, donors, senior management, and arguably IEG, to some extent at the expense of other stakeholders, particularly governments, implementing agencies, firms, and even beneficiaries and citizens. In some cases these stakeholders do not find the value they are looking for and instead find the systems to be burdensome, bureaucratic, and irrelevant. It may not be possible for the Bank Group to realize all of the potential value for all potential stakeholders, but systems need to produce value to the primary beneficiaries of the "Solutions Bank," and to the team and line management where the need for learning arguably is strongest.

Many staff are intrinsically motivated to help clients deliver results, and value working toward improvement and learning, but managerial signals and organizational habits distract. IEG's report *Learning* and *Results in World Bank Operations: Toward a New Learning Strategy* suggests that the Bank needs a fresh approach to learning and knowledge sharing, one that that affords sufficient weight to behavioral drivers, to rigorous measurement of results so that meaningful learning can take place, and to achieving results so that learning for learning's sake is avoided. Both this evaluation and IEG's two evaluations on Bank learning call for wide-ranging changes to deep-rooted organizational habits. How should such reforms be designed? This evaluation was not able to identify any comparable organization with clearly better systems that could be imitated. The four user-centric design workshops conducted as part of this evaluation indicate that reforms will be hard to design because many different systems are intertwined, stakeholders have conflicting needs, and, for people deeply familiar with existing systems, it is hard to visualize what highly functioning systems look like.

Box 5-1. Applying User-Centric Analysis to Understanding Self-Evaluation

User centric analysis offers several important and additional insights into the practice of self-evaluation more generally, and into the challenges specific to the World Bank Group's self-evaluation systems. User-centric analysis considers "usability" as a sub-set of the user experience. Usability describes the extent to which a system, product, or service can be deployed by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction.

Unique to user-centric analysis is the dimension of "satisfaction." The "nudge team" of the United Kingdom has proposed four dimensions for user satisfaction—"easy," "attractive," "social," and "timely" (EAST).

Do the World Bank Group's self-evaluation systems offer "user satisfaction"? There are two kinds of users—those users that feed the system and those users that look to finding the ratings or the records to offer a realistic description of the past. Neither user experience appears to be anywhere close to being easy, attractive, social, or timely:

Usability Dimensions	*User Experience: Feeding the systems*	*User Experience: Taking ratings, records, and lessons from the systems*
Effectiveness	Users do not trust the system overall	Data and lessons are not consistently of high quality and systems do not serve well the "Solutions Bank"
Efficiency	Users find the data input experience costly in terms of time. Templates do not support efficient recording of lessons	Efficient for using the ratings for corporate performance reporting. Inefficient for using records for learning purposes
Satisfaction (easy, attractive, social, timely)	Users find feeding the system a lonely and unsatisfying experience with little if any personal rewards	Users describe the process on a range between "time consuming" to a "waste of time"

In interviews with users, dissatisfaction was the dominant theme and few if any cited positive attributes to their actual experience with systems. There was a lack of trust and little sense that systems provide a service to the user. Positive aspects named, if any, pertained more to the overall function of having accountability, which is needed, and not to the actual experience.

Source: IEG.

Unleashing the Potential of Self-Evaluation

Staff and management perceive IEG's validation function as yet another obstacle to overcome and many staff erroneously believe IEG to be the "owner" of systems that, in fact, are owned by management. Yet because IEG has worked collaboratively with management over the years in designing, maintaining, evolving, and refining systems, the current state of affairs is a shared responsibility between management, IEG, and to some extent the Board on whose behalf IEG conducts validations.

Realizing the potential of self-evaluation to support the Bank Group's strategy and the twin goals will require greater clarity and better balance between accountability, performance management, and learning objectives. The accountability function of mandatory self-evaluation is essential and should not be sacrificed, so when reforming systems, options to enhance learning should be explored while maintaining the accountability function. There is a need to work toward a more learning-oriented culture where users trust systems and have dramatically more positive experiences interacting with them.

Recommendations

This evaluation identifies three broad causes of misaligned incentives for writing and using self-evaluations (illustrated in the three loops in figure 5.1): (1) excessive focus on ratings, (2) attention to volume that overshadows attention to results, and (3) 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, and 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 has a fragmented approach to lesson learning with no clear framework for capturing, storing and acting on lessons and no high-level champion for this has emerged.

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).