

2. Managing Performance with Self-Evaluation

Highlights

- ❖ Management has access to, and makes use of, data that can track performance, identify problem areas, and foster corrective action but some prominently tracked indicators are not on a suitable timescale.
- ❖ When the Bank's Implementation Status and Results Report (ISR) system works as intended, when flags are raised at the right time, and when teams and managers act on these flags, problem projects can be turned around.
- ❖ The Bank has room for earlier and more periodic mid-term project reviews and for more adaptable project design and simpler restructuring procedures.

Quality assurance of the operational portfolio is a major purpose of the Bank Group's performance management. A self-evaluation system that supports performance management should measure performance well, generate the right responses, and be supported by the right incentives and an environment that enables change where change is needed (box 2.1). The World Bank, International Finance Corporation (IFC), and Multilateral Investment Guarantee Agency (MIGA) use very different tools and approaches to manage performance and this chapter is mainly concerned with World Bank lending while making comparisons to IFC (IFC's systems are more fully discussed in chapter 3). The chapter assesses the extent to which self-evaluation systems are used to identify challenges and spark necessary course corrections and identify the factors that influence their use and effectiveness.

Monitoring Performance

KEY INSTRUMENTS AND PROCESSES

In the Bank, ISRs are filed by team leaders every six months for all active projects. Together with Aide Memoires and back-to-office reports, ISRs help manage active projects. ISRs contain a brief narrative, report on outcome indicators, and assign ratings, including on progress toward the achievement of outcomes, implementation progress, risks, safeguards, and monitoring and evaluation (M&E). Like ICRs, ISRs receive attention from management on both sides of the matrix. The indicators and ratings feed into corporate databases. Together with other information, ISR ratings form the basis for "flags" of issues and projects and help identify "problem projects"

CHAPTER 2 MANAGING PERFORMANCE WITH SELF-EVALUATION

in need of management attention. The ISR template has been reformed and simplified twice in the past five years by Operations Policy and Country Services (OPCS) in consultation with the Board and operational staff, resulting in a concise and focused reporting tool. ISRs are made public, with the exception of a specific confidential section.

Box 2-1. Framework for Assessing Self-Evaluation for Performance Management

The literature points to three aspects of a successful performance management system:

- *Measure performance well.* The system tracks performance regularly, identifies challenges to achieving targets, keeps implementation processes in check, and warns teams and managers if projects or programs are not on track to achieving their objectives. The following criteria of quality performance information should be met: relevance, timeliness, credibility, and comprehensiveness.
- *Generate the right responses to the observed performance.* Managers and staff need to learn from the data and take appropriate action; they build on data to make small or large adjustments to the implementation plan, if warranted. The system allows people to propose changes and try out alternative scenarios to put the plan back on course. Processes such as after-action reviews, quarterly business reviews, and data-driven meetings can help.
- *Be supported by the right incentives and an environment that enables change where change is needed.* The active use of self-evaluation for performance management depends on organizational factors, such as attitudes to risk, incentives, leadership signals, and trust in the system. Goal displacement—individuals changing behavior in areas where they are being measured so as to improve a particular performance measure—may occur:¹ “unfortunately, and to the detriment of the program, focusing on improving the wrong behavior can happen at the expense of the more desirable program outcomes.” A performance culture and the right signals from leadership teams can mitigate goal displacement.

Sources: Behn 2002, 2014; Bohte and Meier 2000; DeLancer Julnes 2006; Havens 1983; Moynihan 2008; Mark and others 2000; Newcomer 2007; Radin 2006.

Implementation Completion Reports (ICRs) are filed after project completion and aim to “provide a complete and systematic account of the performance and results of each operation,” according to guidelines. The Independent Evaluation Group (IEG) validates ICRs but not ISRs.

There is a parallel approach for country programs. Country Learning Reviews (CLRs) assess the performance of country programs at the end of the country strategy period, with a progress report in the middle of the cycle. IEG validates the CLRs but not the progress reports.

Bank management has created and is actively relying on a comprehensive, cascading monitoring system. Regular (currently monthly) meetings between senior Bank management and operational units review results and portfolio performance, including problem projects, supported by data systems and processes, some of which draw on information from ISRs and ICRs. These data systems aim to track performance, alert management to problem areas, and enable corrective action. The management dashboard is a useful tool for accessing operational data cascading down from the corporate scorecards (and in some instances also the website of the President's Delivery Unit) with good ability to drill down on specific indicators, Regions, and Global Practices. There is also the *Quarterly Portfolio and Pipeline Quality Report*, and operational updates are presented at the ABCDQ meetings which are chaired by a Managing Director.

IFC has separate systems for monitoring and for self-evaluation.

- IFC monitors its portfolio based on the triple bottom line: Financial (credit risk, profitability) indicators are tracked through separate systems. Development results are monitored annually (for investment) and semi-annually (for advisory) through the Development Outcome Tracking System (DOTS), which uses a number of standard and non-standard indicators (not all mandatory) that are filled in by IFC staff. Environmental and social issues are managed by monitoring compliance with performance standards through the environmental and social risk system, updated annually.
- For self-evaluation of investment, IFC relies on Expanded Project Supervision Reports (XPSRs), which are different from the monitoring systems and are used for a representative sample of mature projects. IEG samples and validates all XPSRs. These systems are discussed in chapter 3.
- IFC advisory services are assessed more like Bank projects with results frameworks focused on development outcomes, semi-annual supervision reports, self-evaluation of all projects at completion (the Project Completion Reports [PCRs]), and a strong role for M&E officers. IEG selects a sample of PCRs for validation. Hence, much of the discussion about self-evaluation of Bank investments applies equally to IFC advisory services.

MONITORING QUALITY AND RESULTS OF THE BANK'S PORTFOLIO

Management has scaled up the use of internal and external client satisfaction surveys for portfolio monitoring purposes (the 2-Minute Feedback Survey and the World Bank Satisfaction Survey). These short surveys cover all lending and all ASA with a country client and are fielded to clients and Bank staff in relevant roles at project milestones. Results are available in real time and are used by Senior Management and

CHAPTER 2 MANAGING PERFORMANCE WITH SELF-EVALUATION

as performance indicators for Regions and Global Practices. Other key indicators of operational quality and results tend to have issues with timeliness or reliability:

- **Projects and commitments at risk and proactivity** (monitoring actions to deal with flagged projects in the preceding 12 months). These are useful and timely indicators for performance management, but they rely on ISRs for the correct identification of problem projects.
- **Projects with baseline data available in the first ISR.** Recently added, this is a useful but also partial indicator of M&E quality. Having baseline data is good, but the indicators in the results frameworks also need strengthening.
- **Satisfactory outcomes, Bank performance at entry, and Bank performance during supervision.** These indicators draw on IEG's ICR reviews of projects exiting the portfolio and hence cannot assist with management of the active portfolio, although they help identify issues in need of attention.
- **Net disconnect** (the difference between the percentage of projects rated as unsatisfactory on outcomes by IEG and the percentage rated in the final ISR as unsatisfactory in achieving their development objectives. Because it relies on ICR reviews, this indicator has a lag time. It is also somewhat imprecise: the ISR measures the likelihood of achieving project outcomes, whereas IEG rates a combination of relevance, efficacy, and efficiency, a subtly different concept.
- **Candor gap** compares recent exits to the current portfolio and is hence more timely, but the term is problematic because it implies that the disconnect is caused by teams being less than fully open, honest, or sincere in their ISR ratings when, in fact, divergent ratings could be caused by a number of factors, including excessive optimism.²

Some indicators on cross-cutting priorities are captured at project or program design or closing and are not tracked during implementation and hence cannot assist with ongoing performance management:

- A “gender flag” is used by both the Bank and IFC to identify gender-informed projects, but the flag assesses project design at entry and does not track or help manage gender-related action during the implementation and completion phases. (Efforts are underway to improve gender tracking during implementation.)³ Since the gender flag does not ensure that attention is paid to gender after the design phase, it may reflect a relatively superficial integration of gender into project design, such as consulting with women during preparation, or disaggregating the number of expected beneficiaries by gender. There is as yet no clear guidance on what “gender-informed” means, and many projects that should have been flagged were not.⁴ The risk is that

easily quantifiable metrics can overshadow more complex challenges of achieving long-term, transformative impact.⁵ The eight interviewed gender coordinators in the World Bank and IFC all said that current systems do not adequately support their work and can lead to pro-forma, “box ticking” approaches to gender.

- Likewise, indicators of gender-informed country strategies and projects and commitments with climate co-benefits are captured only at design and do not support ongoing management of these issues.
- The safeguards section in the ISR is updated, but nothing ensures that this is done by the safeguard specialist on record. The Bank is reportedly setting up a new Environmental Performance Tracking System that is separate from the ISR and from the Integrated Safeguard Data Sheet, though a unified system would be preferable.

When the System Generates the Right Responses, Project Performance Can Improve

THE ROLE OF M&E

Good M&E can significantly boost the performance of operations. The reverse is also true: shortcomings in project monitoring systems hinder performance management.

Regression analyses of Bank projects based on ICR and ICR review data developed for this evaluation show that Bank projects with good-quality M&E tend to have substantially and statistically significant higher outcome ratings than similar projects, controlling for other factors. Establishing causality between M&E quality and outcomes is complicated by the fact that, since 2006, IEG downgrades projects with weak evidence of outcomes, and these are projects that also have weak M&E ratings. The analysis accounts for potential endogeneity in two different ways and also controls for project size, identity of the team leader, expected duration, sector, and borrowers’ performance.⁶ First it uses propensity score matching to compare IEG’s outcome ratings for projects with good M&E to otherwise similar projects with weak M&E. The estimated effect of an increase in M&E quality from “modest” to “substantial” is comparable in magnitude to a one-step jump in ratings on the six-point scale. Second, the analysis uses the outcome rating measured by the ICR. The effect of M&E on outcomes remains statistically significant, but of lower magnitude in this specification.

A detailed analysis of IEG’s ICR reviews for a stratified random sample of 144 investment projects that closed between FY12 and FY14 finds that commonly occurring implementation issues are more prevalent among unsuccessful projects

CHAPTER 2
MANAGING PERFORMANCE WITH SELF-EVALUATION

(those rated marginally unsatisfactory [MU] and below) than among successful projects (those rated marginally satisfactory [MS] and above). The results are shown in table 2.1.⁷

Table 2.1. Most Common Implementation Issues in a Sample of World Bank Investment Projects (FY12-14 exits)

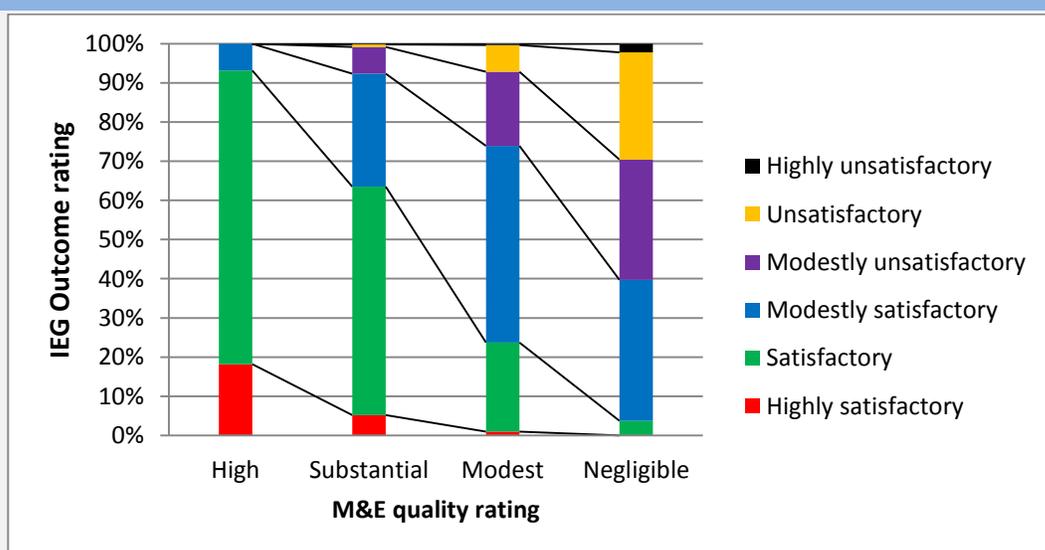
	MU and below	MS and above
Inadequate attention to M&E	45%	30%
Weak project management	28%	20%
ISRs rating too optimistic	30%	8%
Sample Size	83	61

Note: MU=marginally unsatisfactory; MS= marginally satisfactory.

A recent review by the Internal Audit Department (IAD 2015) finds that 85 percent of problem projects that have a satisfactory M&E rating end up being rated satisfactory by IEG compared to 45 percent of problem projects with unsatisfactory M&E ratings.

Given the association between M&E and outcome rating, depicted also in figure 2.1, it would be beneficial to identify M&E shortcomings early and address them, but teams do not often use the ISR to do so. In the ISR, teams can flag M&E issues but they only do so in 18 percent of active projects whereas IEG rates 74 percent of projects negligible or modest on M&E quality and use.⁸ Box 2.2 offers one example of a project with strong M&E design and use.

Figure 2.1. Association between M&E Quality and IEG Outcome Rating for Bank Projects



Box 2-2. Example of high-quality M&E design, implementation, and use: The Kazakhstan Moinak Electricity Transmission Project

The project aimed to increase and improve the supply of electricity to business enterprises and households in southern Kazakhstan in an economically and environmentally sustainable manner. IEG rated the project high on quality of M&E for the following reasons:

- The project used a simple, measurable, and outcome-oriented M&E framework. The outcome indicators reflected well the project objectives without being overly complicated. The intermediate outcome indicators helped monitoring implementation progress by focusing on timely completion of tender documents, timely contract awards, and exact items of equipment delivered, installed, and commissioned in accordance with the plan.
- The project had baseline data in place at the time of project design for the four quantitative measures that it tracked and which were mostly available through the existing management information system (reduction of power deficit, load shedding, wholesale price of electricity, and CO2 emissions).
- The project used progress indicators to keep track of progress and identify implementation challenges. Implementation support missions used this data to identify procurement and other slippages and reach agreement on efforts needed to speed up the process.
- The implementing agency made active use of the information and found some of the indicators so useful that it adopted them for use in future work and integrated them into its regular monitoring activities.

Source: IEG ICR Review, P114766 Kazakhstan Moinak Electricity Transmission Project

THE ROLE OF FLAGS AND PROBLEM PROJECT STATUS

The IAD study finds that when the Bank’s ISR system works as intended, flags are raised at the right time, and teams and managers act on these flags, problem projects can be turned around and ultimately obtain a satisfactory rating. This finding is in line with an earlier study by Cevdet Denizer and others (2013) who found that projects flagged as a problem in the first half of their implementation period and turned around and no longer a problem during the second half of their cycle, had an 83 percent chance of yielding satisfactory results – compared to 75 percent for projects that were never flagged as problem projects.

The ISR system could be improved as an early warning mechanism if team leaders were quicker to raise risk flags and assign cautious ratings once issues surface. Approximately 20 percent of the active Bank investment portfolio is designated as problem projects and these are often identified during the first half of the project life. In interviews, managers and directors described a heightened attention to problem projects driven by periodic senior management reviews. Yet around 23 percent of projects that end up with unsatisfactory IEG outcome ratings were never identified

CHAPTER 2

MANAGING PERFORMANCE WITH SELF-EVALUATION

as problem projects.⁹ Further, only 1 percent of projects that closed in the FY09–14 period and were flagged as problem projects have been flagged as “potential problem project”.¹⁰ As one interviewee put it “what this means, if we do an analogy with a traffic light, is that we have green and then we go directly to red, there is no yellow in the system.”

The poor ability of the ISR to predict project success was also analyzed by Patricia Geli and others (2014), who concluded that “opportunities to take mid-course corrective actions on projects in difficulty are missed due to overly optimistic ISR-DO [development outcome] ratings.” They found that the ISR-DO is a poorer predictor of unsuccessful projects than a simple model made up of project characteristics that are observable early in project life.¹¹ Their model can anticipate between 40 and 46 percent of projects with IEG unsatisfactory outcomes, whereas the ISR-DO ratings in the first quarter of the life of a project anticipates 3 percent of those, and those in the second quarter do so correctly only 17 percent of the time.

Bank management is fully aware that more accurate ISR ratings would improve the early identification of projects in need of attention. To improve the early flagging of issues, two things need to change:

- Projects need to have reliable monitoring data. Obtaining data on project indicators is a major challenge, according to staff.
- Team leaders need incentives to report and rate accurately and flag up issues. According to interviews, some team leaders are hesitant to raise flags because it might generate pointed questions and lead to additional work without additional support to resolve issues.

REMEDIAL ACTION AND RESTRUCTURING OF BANK PROJECTS

While remedial actions are not necessarily decided upon during a formal mid-term review (MTR), the MTR is nevertheless a key decision moment in the Bank project lifecycle. The IAD study shows that the timing of the remedial action is critical to whether a problem project can be turned around; this suggests that MTRs are more useful when they take place early in the project cycle. The World Development Report 2015 found evidence of sunk cost bias among Bank staff (sunk cost bias is the human tendency to continue pursuing activities that have already received substantial investment, even if these activities are no longer likely to be successful). This reinforces the importance of conducting early MTRs or similar in-depth reviews aimed at identifying critical issues.

There is room to conduct the MTRs earlier. As of June 2015, 95 Bank projects had gone more than three years since effectiveness without an MTR, despite guidance to

the contrary.¹² A review of MTR occurrences conducted for this evaluation showed that, among all investment projects that closed in the past three years, about 8 percent (42 projects) had an MTR well before the midpoint. The majority (65 percent) conducted the MTR right around the midpoint, while 27 percent held the MTR in the third quarter of the project life, measured from the date of effectiveness. Consistent with this, a pattern emerging from game-enabled simulations was that, for the fictional projects that had problems in their design or early implementation steps, the MTR appeared too late in the process: there would have been opportunities for course correction earlier in the project's lifetime, but by the time of the MTR, when people realized the problem existed, it was too late and the opportunity was gone. Team leaders found themselves stuck with bad trajectories that could have been corrected if learning had occurred earlier. For this reason, a few Global Practices in certain Regions already attempt to restructure projects before the MTR.

Regardless of when the MTR is done, changing course to improve results is difficult when it involves formal restructuring, especially "level 1" restructuring of project objectives which need Board approval. Not only are internal Bank processes lengthy, many client countries also take a long time to approve restructuring which in some countries may involve ratification by Parliament or approval by the Presidency. Analysis of IEG's ICR reviews for a stratified random sample of 144 investment projects that closed between FY12 and FY14 found that 22 percent of unsuccessful projects (those rated MU and below) missed the opportunity for or delayed restructuring (at any level). The same was true for 13 percent of projects rated MS and above (see box 2.3). IAD also found that the responsiveness to flags and alerts raised by the system can be improved. The indicator of proactivity has declined from 81 percent in FY08 to 66 percent at present, in part because some pro-active measures, such as restructuring, are difficult, time consuming, and depend on borrowers' capacity and commitment to take actions.¹³

Box 2.3. Examples of Delayed Restructuring

The Bank-supported "Market-Led Smallholder Development in the Zambezi Valley" project (in Mozambique) incurred a delay of over two years between the MTR and the final approval of project restructuring, the key action recommended by the MTR. The restructuring was formally requested by the government over a year after the MTR, and preparation of the restructuring could only start once the request was made. There were protracted discussions about what changes to make, and most of the eventually agreed changes required formal approval to be in place. In another project (in Zambia), the ICR review noted that performance problems were identified quickly, but the restructuring took almost two years to complete and Bank management gave little guidance on how to address the issues flagged by the ISRs.

Source: IEG ICR Review, P098040 Mozambique: Market-Led Smallholder Development in the Zambezi Valley

The situation is very different in IFC investment projects. There, changes to projects occur more frequently due to changing market conditions and are considered business decisions that need not get Board approval and therefore can be processed quickly. The lesson is that simpler procedures promotes adaptable project management.

Incentives Affecting Performance Monitoring and Management

Low quality and use of M&E is a cross-cutting finding of this report and can be traced to a lack of rewards and incentives for results-based management. Interviews and focus group discussions made it clear that the self-evaluation systems are not consistently seen as a source of relevant, timely, comprehensive, and credible information that help team leaders, investment officers, and underwriters manage projects. It thus becomes a perfunctory exercise.

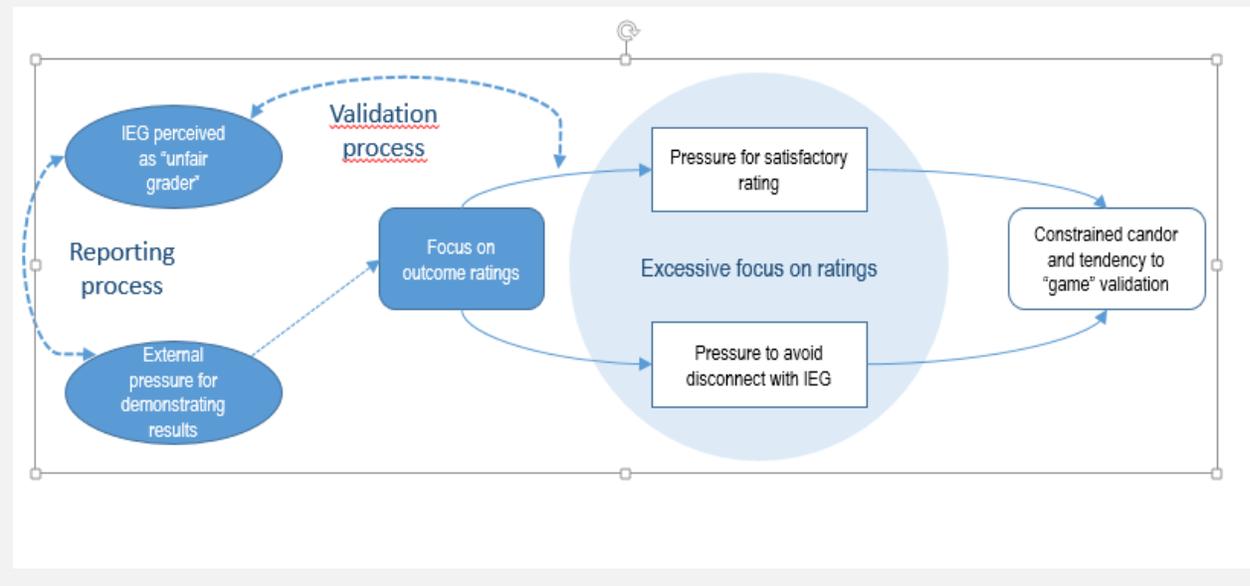
Partly, the signals come from outside the systems with pressure for lending volume and a perception that individual success depends more on obtaining new deals and ensuring timely disbursement than on quality implementation and, ultimately, results. This view was particularly frequent among IFC interviewees. Out of 17 interviews with IFC staff and managers where this topic was discussed, seven mentioned the drive for volume and closing new deals as the primary motivator and 12 thought that there was no incentive to take self-evaluation seriously. Reaching targets and complying with reporting requirements was often perceived as getting in the way of pursuing results. The Bank's heavy reliance on consultants to write ICRs also sends a signal about the lack of importance.

Yet part of the signals also come from the acute focus on outcome ratings. Staff and management are concerned with obtaining good ratings and avoiding disconnect with IEG. Thus, the validation process has significant influences over behaviors and incentives, and affects the content, candor and usefulness of the self-evaluations (figure 2.2).

Across the World Bank Group, there is room for managerial signals to more consistently emphasize excellence in implementation support geared toward development results. IEG's 2014 Results and Performance Report analyzed the scope to improve the quality of implementation support for both the Bank and IFC. The ability to solve implementation problems is a key factor, and determined in part by the frequency and quality of client contact.¹⁴ However, out of 41 interviewees who

specifically discussed rewards and incentives, 31 stated that staff do not get rewarded for fixing problem projects or for doing an honest and quality evaluation.

Figure 2.2. The Incentive Signals Underlying Performance Management



In both the Bank and IFC, prestige was perceived as coming from peer recognition of successes, particularly through getting new projects approved. Fear of damage to one’s reputation and concerns about reputational risks attached to poor results was a recurrent theme in both the Bank and IFC and linked to limits in candor: acknowledging that a project is not performing well was described as “exposing one’s dirty laundry” and best avoided. Safe space for trial-and-error was missing.¹⁵ Systems were often used defensively (for example, to manage indicators of disconnect), more than as a source for data on how to boost results. This creates goal displacement, where the internal needs of team leaders and teams are not well served by the system.

A number of staff and managers mentioned getting important information from alternative sources that they deem more useful and credible than the self-evaluation system, such as conversations with colleagues, clients, or implementing agencies; letters from civil society organizations; and operational systems that focus on procurement and financial transactions (this information may or may not be reflected in ISRs and back-to-office reports). In instances where self-evaluation information was deemed useful, it was because it had sparked further discussion within a team (for example, prompted by a country or practice director who picked up on an issue flagged in an ISR). The Public-Private Partnership team started conducting transaction review meetings half way through project implementation, tapping into lessons from PCRs and other platforms to address emerging challenges.

CHAPTER 2 MANAGING PERFORMANCE WITH SELF-EVALUATION

All 34 interviewees, including 12 managers, with whom the use of self-evaluation for strategic decision-making was discussed, reported that information from the systems was not used to make strategic change at the level of the portfolio (as opposed to addressing implementation issues in specific projects). Even if not entirely accurate, this perception is one reason that staff and line managers are demoralized about the value of systems.

There is Opportunity to Do Better

The success of self-evaluation for performance also lies in being able to change course as often as necessary, informed by a continuous flow of information about how a project is performing. The data revolution has transformed a number of industries and may have the potential to transform development and boost performance, including via rapid data flows.¹⁶ The practice of adaptive management – small but frequent course corrections – is better suited to capitalize on the data revolution than the prevalent model in the Bank Group, which concentrates the bulk of the effort in the design phase. As one interviewee put it, “implementation and evaluation remain afterthoughts to design.” Some development agencies have begun to rely more on adaptive management (box 2.4).

Box 2.4. The United Kingdom’s Department for International Development’s (DFID) Experience with Adaptive Management

Since 2011, DFID has done a comprehensive reform of its project design and results reporting systems. However, DFID’s self- and external assessments suggest that strengthening project design and M&E does not automatically translate into the effective transfer of knowledge, project management, and delivery of results. Rather, tighter rules increased the pressure to comply and drifted the staff’s attention and time away from effective delivery and self-reflection. In response, DFID shifted toward “adaptive management” to bring greater flexibility, timeliness, and simplicity to the project management cycle and to allow more innovation and adaptive learning. DFID is also preparing a Learning Strategy that is expected to address many of the organizational barriers to learning. DFID’s example shows that deliberate systemwide efforts are needed to promote an organizational culture of learning that encompasses incentives, systems, and processes to facilitate learning along with loosened compliance pressures in areas where that is possible.

Source: DFID (2013); Independent Commission for Aid Impact, (2014); (2015). See also Appendix B.

Embedding impact evaluations in projects is not only useful for measuring results and allowing for attribution, it also has potential to add value by enhancing the quality of logic chains, results frameworks, and data collection, with positive

spillover for other M&E activities that are not necessarily related to the impact evaluation. Additionally, while the first generation of impact evaluations focused on rigid evaluation of implementation of the project as initially designed, there are current efforts to test variations around intervention design and incorporate quick feedback loops that allow for adaptive management.

Summing Up

A self-evaluation system that supports performance management is a system that tracks performance using relevant, credible, and timely information and allows the managerial team to use that data to reflect on progress and challenges. It is a system that is supported by incentives to acknowledge issues and make course corrections. If the self-evaluation systems of the Bank Group more consistently embodied these critical elements, they would more effectively facilitate early warning and course correction.

There is active management of a number of prominently tracked aggregated performance indicators. Indicators aggregated from the Bank's ISRs and IFC's DOTS are timely but insufficiently precise because of weaknesses in the underlying M&E systems, lack of quality control of data inputs, and teams' tendency toward excessive optimism. Other indicators, including gender flags, most citizen engagement indicators, and outcome ratings, are often not on a timescale where they can support ongoing management of the performance of projects and portfolios. MTRs sometimes take place late, as does remedial action to address identified problems. Restructuring of Bank projects is complicated because of lengthy Bank and client procedures. Incentives and managerial signals need to more often reward teams for good M&E and identification and fixing of problems and for reduced pressure around quantitatively tracked indicators.