4. Learning from Self-Evaluation

Highlights

- Having a self-evaluation system in which the entire organization writes substantive end-of-project reports is a noteworthy accomplishment.
- Knowledge from the mandatory self-evaluation systems is rarely valued or used, and there is little effort to extract and synthesize evidence and lessons or to inform operations.
- The focus of the systems on accountability drives the shape, scope, timing, and content of reporting and limits their usefulness for learning.
- Tensions and concerns over ratings and disconnects distract from learning.
- There is more learning from impact evaluations, which are optional, seen as technically credible, and done in response to specific learning interests.

The Place of Self-Evaluation in Organizational Learning

Having a self-evaluation system in which the entire organization writes substantive end-of-project reports is a noteworthy accomplishment, one that few other organizations can claim. In principle, this could contribute significantly to individual and organizational learning, as articulated in Operational Policy 13.60 and by the International Finance Corporation (IFC). Evaluation has long been viewed as an instrument for accountability, but evaluators argue that the ultimate value of evaluation is in stimulating organizational learning with a view to improving performance by management and staff who are responsible for the design and implementation of policies, programs, and projects.[1] Yet learning from evaluation (of any kind) does not occur automatically. IEG's evaluations of how the Bank learns from lending found that the Bank lacks a robust learning culture.

A vast literature on evaluation use emphasizes that, to enable their use, evaluations must be timely, relevant, based on sound data, perceived as technically credible, delivered in an understandable format, based on collaboration and follow-up between evaluators and those being evaluated, and contain clear messages and new lessons. Use also depends on the receptivity and political environment in the organization receiving evaluation findings (box 4.1).[2]

Scholars and evaluators observe tensions between different objectives of evaluation. According to John Mayne (2015:47), "evaluation is often seen by those being evaluated a bit like an audit, something to be avoided or at least controlled as much as possible." Self-evaluation that is subject to independent validation can have the same audit connotation. Other scholars note that evaluation may serve a symbolic function that confers legitimacy but is delinked from organizational decision-making and learning in a context where the primary purpose of evaluation gradually becomes to satisfy funders more than to assess effectiveness.[3] Disclosing evaluation information to external audiences raises the stakes further, and can lead to risk aversion, deter learning from failure, and hinder innovation.[4] A review of OECD-DAC members' systems for measuring results finds that much results information is used for accountability and concise external reporting at the cost of shedding light on how long-term results have been achieved, which would support learning.[5]

This chapter addresses the degree to which self-evaluations serve individual and organizational learning, taking into account the observations in literature regarding factors that enable evaluation use and organizational learning and tensions between accountability and learning.[6]

Box 4-1. Organizational Learning

Organizational learning has numerous definitions and conceptualizations, but the basic notion is that the organization engages in a comprehensive effort to create knowledge and facilitate active learning among its staff in support of its goals. Building on IEG's recent evaluations of Learning in Bank Operations (2014; 2015) and external research, this chapter posits that organizational learning takes place when an organization institutes an enabling environment—policies, processes, structures, and incentives—for its staff to:

- Generate, share, and apply knowledge that is timely and based on credible data and analysis.
- Participate in active learning from and with others.

This should be done so as to further the goals of the organization.

Sources: Argyris and Schon 1978; Davenport and Prusack 2000; Mallon, Clarey, and Vickers 2012; Frost 2014; IEG 2014 and 2015; and Senge 1990.

Organizational Learning from Self-Evaluations: The State of Affairs

The Bank Group has instituted policies and processes for generating and sharing knowledge from mandatory self-evaluations. Set processes define their timing and formats. Templates guide the information generated and contain "lessons" sections meant to capture knowledge of wider relevance. Over the years, a vast number of self-evaluation reports has accumulated.[7] Impact evaluations, although not mandatory, have been driven institutionally in the Bank through the Development Impact Evaluation Initiative (DIME), the Strategic Impact Evaluation Fund (SIEF), the Health Results Innovation Trust Fund, and the Africa Gender Lab.[8]

There is demand in the Bank Group Board and management for knowledge and evidence to enhance development effectiveness and Bank Group management has taken important steps to promote results orientation and strengthen self-evaluation use. [9] For example, the Operations Policy and Country Services (OPCS) Vice Presidential Unit has proposed that the agenda for project concept note and decision meetings include a discussion of the evaluative evidence that has informed the design and the plan for collecting baseline data.[10] Eighty-one global lead positions have been created to provide technical leadership and strengthen evidence-based learning and knowledge sharing in core Global Practice areas. Ongoing work aims to refocus the Bank's advisory services and analytics (ASA) to better meet client needs, new knowledge hubs have been set up to share development experiences with partners, and the Science of Delivery initiative aims to create a cumulative knowledge base of delivery know-how.

Box 4-2. What the External Panel Said About Learning Culture and Self-Evaluation

According to the external panel review of IEG commissioned by the Committee on Development Effectiveness (CODE), the Bank Group has insufficient attention to learning, course corrections, and (self) evaluation use. The panel reviewed IEG "within the larger, interdependent system in which it operates, including core institutional processes around learning and accountability" and found that "the current overall system and processes are broken. They do not support a mindset of learning, course correction, continuous improvement and accountability. Nor do they create the cycles of learning and accountability necessary to make progress toward key development goals. Learning is not prioritized, accountability is mechanical and does not support necessary learning or continuous improvement, and while there is some single-loop learning (are we doing it right?), there is less discussion of the critically important double-loop questions about whether or not the Bank is doing the right things to reach their goals….. Improving the self-evaluation system is key for the success of [the Bank Group's] new strategy and for strengthening the basis for IEG's validation and review—and thereby its contribution to the Corporate Scorecard."

Source: External Panel Review of IEG, 2015.

The demand for knowledge and evidence to enhance development effectiveness has not been matched with an active learning culture (see box 4.2) and the mandatory self-evaluation systems have not yielded a strong repository of knowledge that is mined, shared, and used regularly by staff, although there are exceptions. Interviewees across the Bank Group almost unanimously described the process of conducting and writing a self-evaluation as a useful learning exercise for them individually, but with few benefits accruing beyond themselves. Fifty percent of those interviewed noted that they had learned *something* through the self-evaluation system. Authors of self-evaluation reports noted that they:

- learned about sectors and countries in which the self-evaluation took place
- benefited from reflection and the chance to think retrospectively
- better understood client relationships.

According to a survey conducted for IEG's Learning and Results evaluation, 23 percent of Bank task team leaders use ICRs to a "substantial" or "very large" extent for learning for new operations during project preparation (contrasted with 50 percent for documents produced by clients and 51 percent for analytical, advisory, and economic work).[11] And 31 percent indicated that they use ICRs for learning from previous operations during implementation (in contrast with 54 percent for documents produced by clients and 42 percent for knowledge products financed out of project loans and credit proceeds).

Self-evaluations are not used regularly for extracting and synthesizing evidence and lessons that would be used to inform new or ongoing operations, and if a particular self-evaluation report were to raise policy or strategic issues, no mechanism exists to elevate it for management's attention. Said one staff: "There is no learning loop, or systematic approach to feed the lessons of projects into any larger agenda." A study of lesson transmission in IFC from one project to another estimates this at only 7 percent. This evaluation identified relatively few instances where business units mine or accumulate lessons or insights from mandatory self-evaluations (box 4.3). Interviews with staff reveal that not much value is placed on systematic learning from self-evaluations even as some project design document templates contain a mandatory section on how past lessons have informed the proposed design. This imposes a norm of using self-evaluation information. Nonetheless, staff cautioned that filling out such a section can be a gesture of compliance, not necessarily one of absorbing lessons learned. To promote a culture of applying evaluative lessons, mandatory sections will not suffice.

A study of IFC's effectiveness at lesson learning (through self-evaluation or in other ways) conducted for this evaluation concludes that IFC has a fragmented approach to lesson learning with no clear framework for capturing, storing and acting on lessons and that no high-level champion for this has emerged.[12] All 14 staff and managers interviewed for the study thought IFC's lesson-learning system is in need of overhaul. Participants in the electronic survey of all IFC staff were asked to rate the effectiveness of IFC's lesson learning by selecting one of five categories: Completely ineffective, slightly effective, moderately effective, very effective, and totally effective. These were converted into scores from 0 to 4, with 4 being "totally effective." The average effectiveness score is 1.81 out of 4 (figure 4.1).[13] Staff at grades GG (senior) have the least favorable perception of IFC's lesson learning effectiveness, while staff at grades GA-GD (administrative and client support) have the most favorable perception. Numerous projects are believed to have failed as a result of repeating the mistakes of the past.

Box 4-3. Good Practice Approaches to Learning from Self-Evaluation

The Governance Global Practice organized a "boot camp" in 2014, in which in-depth reviews of ICRs were undertaken with the objective of learning lessons to feed into ongoing and future operations. The teams used the ICRs as a springboard to present, analyze, and interpret the lessons, contextualizing them with rich tacit information they had from their experiences. This strategic use of evidence with discussion and debate proved to be valuable and insightful for the participants.

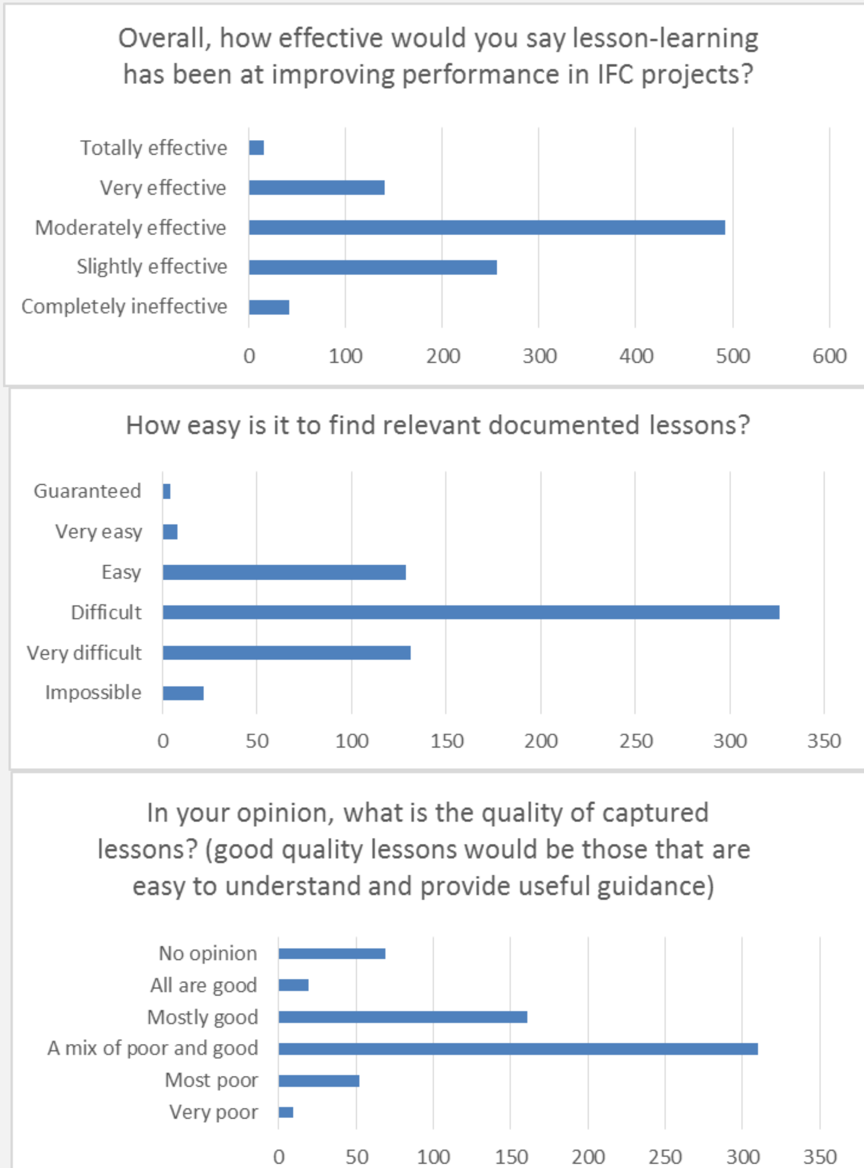
The Public-Private Partnership team set up a process where small meetings are used to capture critical lessons from each transaction. The Africa Region synthesized ICRs over the period 2011 to 2014 to inform actions to improve portfolio performance. IFC's Results Measurement Unit reviews PCRs for lessons and reasons for failure and success. MIGA has seminars to present project self-evaluations to MIGA staff.

Source: IEG interviews.

Staff often prefer tacit knowledge—"having coffee with peers"—to obtain nuanced knowledge and experience, but self-evaluation systems do not exploit dialogue formats as part of the learning process. In IFC, this tacit oral approach is regarded as "IFC style" and it works well for experienced staff in Washington. However, interviews recognized that this approach was not sustainable as IFC grows in size and geographic reach. Dialogue and tacit knowledge alone is insufficient—experts on knowledge management note that once individual lessons, evidence, or information are generated, they should be culled, codified, and turned into actionable guidance for implementation or strategy formulation as weaknesses in documenting key lessons and over-reliance on personal connections can lead to inefficiencies and loss of important knowledge. Combining written and dialogue-based formats could boost learning from self-evaluation. The health, airline, and energy industries are more attuned to the value of good lesson learning, which can be mission-critical or lifesaving. Many hospitals, for example, conduct post-mortems to this end. When researchers traced the source of hospital infections to improper handwashing, this was developed into a checklist that is now widely used and has reduced hospital infections.[14]

The World Bank Group could usefully build more dialogue into self-evaluation processes. Some parts of the Bank Group use deliberative meetings to reflect on experiences—most systematically in the PPP group and in parts of IFC's Advisory Services—and these were seen as useful safe spaces for learning. The validation process could include (non-antagonistic) dialogue between the author, the project

Figure 4.1. Assessment of the Effectiveness of Lesson Learning in IFC by Survey Respondents



team, and the validator aiming to explore the set of relevant lessons. Peer review processes could require dialogue formats instead of report formats for sharing knowledge on past projects, involving peer-to-peer learning.

There are also opportunities for more consistently exploiting self-evaluations to drive on-the-job learning and professional growth of junior staff (box 4.4). None of the staff and managers interviewed discussed strategically choosing an ICR or XPSR author with a view to promote learning, say to feed into a follow-on operation or to address strategic issues. Many ICRs and CLRs are written by consultants rather than staff, according to interviews and reports' acknowledgment pages.[15] The reasons they are outsourced are varied and legitimate—time constraints, desire for impartiality, and skills in writing a report that meets the requirements—but, by using consultants, the Bank forgoes an opportunity for contextual learning by staff and also signals the low priority placed on self-evaluation. In IFC, junior investment officers write XPSRs, thus seizing this opportunity for learning about project design, processing, and execution from investments made by the department (although for accountability purposes the XPSRs are sampled randomly by IEG and not strategically to meet learning needs). The same applies to MIGA, where junior underwriters write PERs.

Box 4-4. Facilitating Active Learning With and From Others through Self-Evaluation

The literature indicates that in addition to knowledge generation, organizational learning is supported by creating an environment in which people are expected to learn constantly (through a range of modalities, such as on-the-job learning, mentoring, and training) and opportunities are available for the application of that knowledge. Organizations that value learning also promote a culture in which ways of thinking and mental models are challenged in an environment of trust (Senge 1990). Research shows that when companies adopt "formalized informal learning," those programs outperform formal training by 3 to 1 (Jackson and Williamson 2011; Mallon, Clarey, and Vickers 2012). In these companies the corporate training team not only trains people, it puts in place programs to help employees learn on the job, an important aspect of transmitting tacit knowledge. Concrete practices and processes are required; simply having an environment supportive of learning is insufficient (Garvin, Edmonson, and Gino 2008). Leaders in the field of evaluation also note the importance of participatory approaches to enhance learning (Mayne 2015).

LESSONS

ICR lessons have a justified reputation of being rather obvious and generic. The evaluation team's review of ICR lessons covered 60 ICRs with an average of 5.8 lessons per ICR. The majority of the lessons pertained to sectoral issues (70 percent); 10 percent to country-level issues, primarily in development policy lending (DPL) operations; and the remaining 20 percent were cross-cutting. Eighty-eight percent of the lessons were worded as "lessons" (as opposed to "findings"), and ought therefore to be generalizable to future operations in other countries. Lessons were often written in very general terms, without specific recommendations on how to do things differently in the future (for example, "complex project design in a low-capacity environment leads to poor implementation and non-attainment of objectives"[16]). Further, 74 percent of the lessons pertained to design issues; 21 percent to implementation; 3 percent to internal institutional issues; and 2 percent to external causes.

The evidence behind lessons was sometimes weak:

- 18 percent of lessons were backed up by solid evidence presented in the ICR that discusses the issue and the consequences of the issue.
- 30 percent of lessons were backed up by some supporting evidence in the ICR.
- 34 percent of lessons lack supporting evidence and analysis.
- 18 percent of lessons appear to come completely out of the blue.

The lessons were not always applicable:

- 28 percent of lessons were very specific on how things should be done differently in the future.
- 47 percent pointed toward a direction, but readers would need more information to know specifically what to do.
- 24 percent were too broad and did not specify what to do in the future.

Thus, several issues hamper the potential for better lesson learning from the ICR. The ICR document is both a reporting and a lesson learning tool and does not allow for a systematic approach to recording lessons. Lesson learning would require reading the entire document. Lessons are not consistently quality-controlled and evidence-based and may not sufficiently cover internal institutional issues. And whereas each ICR provides one data point, lesson-learning should be based on mining a set of experiences to ensure that lessons are turned into knowledge with applicability across contexts.

In interviews, Bank and IFC staff placed low value on information and lessons from self-evaluations and expressed the view that the "right" lessons are not being captured and that lessons captured fail to address the most critical issues, are too generic, or too specific. Across the World Bank Group, 48 percent of those interviewed cited one or more major obstacles to using the lessons from the mandatory self-evaluations. Staff observed that similar types of lessons appear in project after project, year after year, yet they are not acted upon and addressed in future operations.[17] For example, interviewees noted that self-evaluations are normally silent on lessons pertaining to Bank Group internal constraints such as team leader turnover, the factors leading to excessive complexity of projects, and client-related issues. Such critical factors can result in mistakes and problems that are worth learning from but tend to be left out of self-evaluations. In part, this is because Bank (but not IFC and MIGA) self-evaluations are disclosed to the public.

IFC lessons were found to be of variable quality. IFC lessons were assessed for quality using a system that recognizes that lessons have certain components, referred to in the military as Observations, Insights, Lessons representing the train of thought from an observation through to deriving a recommendation for future projects. Average lesson quality in IFC was found to be relatively low, even though there are some good examples within IFC. Lesson quality is highest (though still variable) in the lessons in the XPSRs and LessonFinder, although even these were described by interviewees as poor or variable in quality. Likewise, a majority of IFC survey respondents thought that lessons are a mix of good and bad quality.[18] Where quality was poor, a large proportion of the "lessons" were observations rather than lessons. Forty percent of the lessons in the "Lessons of Experience" and "Learning By Doing" and 50 percent in the Post Vivems, for example, contained no recommendations for the future, or only weak generic statements. Similarly, the majority of the lessons within SmartLessons are observations or mini-case studies under a vague heading such as "raise awareness at multiple levels" or "partner with the press." Root cause analysis in many of the lessons is superficial, and looks primarily at external root causes rather than addressing issues within IFC. The mixed quality of lessons was also recognized by survey participants. If staff find brief or vague statements rather than useful content, they stop seeking.

There is no systematic support offered to Bank and IFC self-evaluation authors or users to facilitate lessons identification and learning. There is little guidance on how to write good lessons and no processes of using dialogue formats to help authors discover key findings and lessons—a missed opportunity because, in the Bank Group's face-to-face culture, dialogue would likely spur better lessons and greater use of them. IFC stores its lessons in different systems; sometimes as individual lesson documents collected within a file folder, sometimes as sections within project reports. Few survey respondents were aware that lessons were also collected in a lessons database, LessonFinder.

Impact Evaluations

Results from Bank impact evaluations are well-regarded but still underused in reporting on project effectiveness or integrating them as lessons. World Bank and other impact evaluation hubs put a lot of emphasis on disseminating information about these evaluations through newsletters, research publications, seminars, and other media. Some World Bank sector strategies have included the findings in areas in which there is large body of evidence from impact evaluations, such as education and social protection, thus reflecting systematic use of knowledge for organizational purposes.[19] There continues to be room to use impact evaluations to a greater extent to inform operational decisions, according to IEG's 2012 evaluation of impact evaluations, interviews done for this evaluation, and IEG's report on Social Safety Nets and Gender. According to the latter, if projects are not conscious of potential gender impacts, they do not collect gender-disaggregated data and do not make the best use of existing impact evaluation evidence. Coupled with the lack of attention to gender in project monitoring, this raises questions about missed opportunities for learning.[20] Interviews with team leaders indicated that they have little time to familiarize themselves with recent findings and rely on their own networks of colleagues when they have questions, making lessons application somewhat idiosyncratic.

Bank-sponsored impact evaluations could improve how they serve operations by more effectively brokering knowledge and by explicitly including reflections on the evaluated project and lessons in future ones. More could be done to mine the evidence, for example by conducting and better using existing systematic reviews and by better bridging the agendas and priorities of researchers and operational staff. While several regional chief economists' offices have an impact evaluation point of contact, they are generally responsible for conducting and supporting impact evaluations in their Region rather than for disseminating evaluation findings. Assigning responsibility for knowledge translation to dedicated "knowledge brokers" could help transfer information from impact evaluations into actionable lessons in the competitive space for staff attention. Some parts of the Bank, the Africa Region for example, have seen good results from engaging in several of these modalities and may be a useful template upon which other Regions and Global Practices can build.

Shape, Scope, Timing, and Content of Reporting

Driven by corporate requirements (Operational Policy 13.60 for the Bank), the vast majority of self-evaluations are project-specific (CLRs are an exception) and summative in nature. There are benefits to this way of doing things from a reporting and accountability perspective, but clear drawbacks from a learning perspective.

First, the aid architecture emphasizes programmatic approaches, yet the gravity of the self-evaluation architecture remains the project (except CLRs). As Bank management has emphasized, this "project mentality" does not square with the "development solution" mentality implied by the Bank Group Strategy. To facilitate learning and guide strategic decisions, it can help to focus evaluations around themes, sectors, or clusters of similar projects (IEG does this in its evaluations and learning products, as do evaluation departments in other organizations). Interviewees and focus group participants noted that self-evaluations rarely address questions of strategic importance for upcoming operations or to the sector, but that there is potential to institute this approach for clusters of projects. Impact evaluation hubs[21] sponsor impact evaluations of individual projects that are clustered around themes and within Regions, for example on gender in Africa or results-based health financing.

Second, and related, the systems pay little attention to synergies (or lack thereof) across activities. For example: Do knowledge, lending, and policy dialogue activities mesh well? For trans-border issues such as water and transport, are there synergies between activities in adjacent countries?

Third, funding is tied to project evaluations. Business units can commission evaluations on any topic they desire, but IEG did not identify any routine evaluation funding sources other than donor funds for impact and program evaluations and the administrative budget procedures that are used to finance the mandatory self-evaluations.[22] Funding for formative, voluntary evaluations is therefore not readily available and it is not known how many are conducted. Key informants from the Bank noted the difficulty in securing funding for evaluations of government interventions in areas where the Bank does not have an active lending program, limiting opportunities to engage.

Fourth, there is room to improve on self-evaluation timing to support timely learning and decision-making:

- For Bank investment projects for which ICR reviews were completed in FY15, the most frequent year in which they were approved was FY06, nine years earlier, and they hence shed little light on how well current approaches to project design tackle development problems (the lag time is a few years less for XPSRs and policy lending). This is because they are done after closing.
- The timing of XPSRs is somewhat flexible, and CLRs are timed to inform the next country program, but ICR timing is not flexible: always done within six months of closing, ICRs come too late to inform follow-on operations which are prepared before project closing. Hence there is no room to consider optimal evaluation timing.
- Decisions about course corrections and scaling up pilot interventions need to benefit from accurate and timely evaluative results.[23] IEG's evaluation of the poverty orientation of country programs therefore recommended "attention at project inception to evaluability" and "explicit evaluation protocols for piloted interventions to capture lessons from experience on poverty reduction, with a view towards opportunities for scaling up successful interventions."[24]
- The writers of some ICRs and PCRs have not had time to complete planned beneficiary surveys or other data collection that would facilitate accurate

measurement. For IFC advisory services, BROE 2013 recommended post-completion monitoring to address the timing issues.[25]

- Likewise, interviewees expressed a desire for more timely impact evaluation findings. Leaders in the impact evaluation community have indicated that they are aware of this concern and are working to integrate impact evaluation methods into project monitoring systems to be able to provide mid-course interim findings to help projects make needed course corrections.

Consistent with this, key informants from the Bank advocated for more flexibility: some projects may need frequent assessments during implementation and some projects may need to be revisited five years after closing depending on their profile and impact. Users who participated in focus groups want flexible systems that are transparent, adaptable, and promote real-time learning and information sharing. They also argued that more could be done to capture knowledge gained during implementation, ideally right after missions for easy recall.[26]

Fifth, a more comprehensive assessment of unintended positive and negative consequences could promote learning. As Vinod Thomas and Xubei Luo (2012:9) argued, "Unintended results can provide a rich source of learning for future activities and checks on current ones." An evolving good practice for impact evaluations is to pair quantitative methods with qualitative methods to understand not only what happened and what the results were, but also how the program was implemented and why the outcomes came out as they did.

Sixth, some of the nuts and bolts such as sector and theme codes and core sector indicators facilitate the aggregation of project information. According to guidelines, "the Bank's theme and sector coding system provides the basis for analyzing and reporting on the content of Bank activities," and "responds to shareholder recommendations for standard reporting."[27] Teams do not have the flexibility to use theme codes that align with common knowledge topics (such as child labor or school feeding).[28] Imposing core sector indicators can promote useful standard reporting but also crowd out the ability to adapt metrics to the project context and to learning needs.

Box 4-5. Learning from Evaluation in Other Agencies

The evaluation community has responded in various ways to enhance uptake and learning, yet learning from self- and independent evaluations remains weaker than desired in several development agencies, according to studies. For example, a study on the uptake of learning in the European Union's Directorate-General for International Cooperation and Development cites issues such as lack of systematic attempts in most reports to compile lessons, rigid methodologies that disincentivize learning, tendency toward bureaucratic compliance, and lack of staff time for learning. Both the Asian Development Bank (ADB) and the African Development Bank (AfDB) have launched knowledge platforms to enhance sharing of findings, lessons, and recommendations from past projects. An evaluation of the International Monetary Fund's (IMF) self-evaluation system finds learning to be weak. The evaluation community has adopted good practice guidelines, and, to improve timeliness, started conducting more formative (or real-time) evaluations.

Sources: European Commission (2014); Nielsen, Turksema, and Knaap (2015); Independent Evaluation Office (IMF)(2015); Thomas and Luo (2012). See also Appendix B.

Summing up, summative (backward-looking) evaluation purposes sideline more formative (learning-oriented) purposes in how systems operate. If the self-evaluation systems had been set up to primarily serve learning, they would have been more forward-looking (how can we do better?), more selective (which projects and programs offer the greatest learning opportunities?), more programmatic (are there synergies across activities and countries?), attuned to unintended consequences, and more often done in real-time. As an operational Practice Manager expressed, "fundamentally, [self-evaluations] should be formative and not summative. They cannot do both for a range of reasons…As an institution we need to pick our objective, we can't have it both ways." The Bank Group is not alone in facing weak learning from self-evaluation (box 4.5).

Incentives to Learn from Self-Evaluations

Almost 70 percent of Bank staff agree or strongly agree that lending pressure crowds out learning.[29] Similarly, in interviews for the current evaluation, staff noted that there is an implicit "pressure to lend" and the self-evaluations are primarily a tool for reporting, although impact evaluations are supporting learning.

The Bank Group's strong culture of success and competition leads staff to be wary about acknowledging issues or problems that may be interpreted as failure in projects (box 4.6). An overwhelming majority—78 percent—of the interviewees specifically mentioned that there are either no incentives or negative incentives for candid self-evaluation. Forty percent noted negative incentives for reporting issues that may be interpreted as failure; some worried about the implications on their professional reputation.

Box 4-6. Learning from Failure

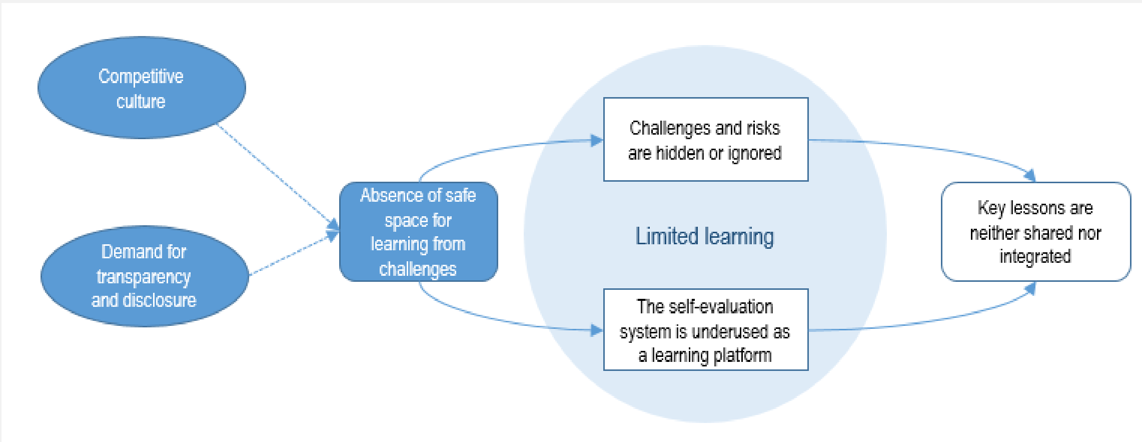
Literature on organizational development states that the critical examination of failure can trigger learning, especially when organizations diagnose not only the proximal causes of failure but also examine the underlying causes—policies, norms, and objectives—and develop mechanisms for improvement, which can also lead to innovations.[a] The World Development Report (2015 chapter 10) also emphasized that it is important to recognize that "'failure' is sometimes unavoidable in development and encouraging individuals to learn, rather than hide, from it." A review of results measurement systems among bilateral donors emphasizes the need for a "strong and mature results culture with incentives to strengthen results measurement and [an] enabling environment to discuss poor and good performance.[b]

Notes: a. Argyris and Shon 1978; Edmonson 2011; Frese and Keith 2015. b. OECD-DAC 2013.

The absence of a safe space for trying things out, identifying and discussing problems and failures, and accumulating knowledge from failure was a recurrent theme in interviews and focus groups. Interviews done for this evaluation also suggest that the Bank has room to better embrace the "failures" identified by Bank-sponsored impact evaluations. Some impact evaluations reporting "null" result—findings of weak or no results—have met with lukewarm or obstructionist responses, though in other instances researchers have been able to use null results to impel closer collaboration and investigation with the client country. Lack of candor is equally applicable to IFC, as evidenced by interviews and BROE (2013), even though ratings are less salient there, with profitability the bigger concern. The staff, therefore, in the words of an interviewee, "focus on what is needed" to be consistent with guidelines and to avoid a downgrade. Lesson learning has no high-level IFC champion, and many of the signals staff perceive (or interpret) from management promote short-term actions, and some interviewed IFC staff expressed cynicism about lesson learning.

The system's focus on accountability and reporting creates negative associations among intended users, leading to under-use. Ratings can, in principle, focus attention and stimulate action. Yet users reported overwhelmingly negative experiences with the ratings and validation processes; these frustrating experiences caused negative perceptions of the systems in general and IEG's role in particular. Staff perceive that ratings and validations focus too rigidly on documentation requirements associated with the initial project objectives and results frameworks, and often feel unfairly assessed by IEG, making them disassociate from the process and the information it generates. Sixty percent of the Bank staff interviewed stated that they are concerned with ratings and potential disconnect with IEG and that this preoccupation leads them to focus less on learning from self-evaluation (figure 4.2).

Figure 4.2. Incentives around Learning



IEG reviewed a random sample of 74 substantive email responses received from Global Practices in response to ICR reviews and found that nearly all (72 of 74) disputed ratings (often arguing that IEG had misinterpreted evidence, results frameworks, or that guidance was unclear). These responses only rarely discussed learning and lessons: eight mentioned learning, but six in the context of defending the ICR and 18 mentioned lessons, but 11 in the context of defending the ICR. The review also judged the tone of ICR review responses to be mostly factual but at times crossing into antagonistic (16 percent of responses, but only in parts) and personal or emotional (14 percent).[30] On a similar note, interviewees from the Bank noted that meetings to review draft ICRs rarely focus on lessons and implications and, instead, tend to focus on proposed ratings and their congruence with the available evidence in anticipation of the reaction from IEG's validator.

In interviews, ratings were the second-most frequently cited obstacle to learning, after the nature of the lessons. The issues noted by staff square well with findings from educational scholars on the impact of grading on students focus, learning, and motivation (box 4.7)—although potentially ratings can also drive attention and action. This said, the ratings validation process is far from the only reason that learning is below potential. Some quotes illustrate how interviewees perceived the impact of ratings on learning:

- "We do not learn from the graveyards around us" because "ratings are a lightning rod."

- "Framing self-evaluation as an accountability tool automatically makes it confrontational."
- "As a manager, every month I take a look at the dashboard and what unfortunately focuses the attention is the disconnect with IEG. If there is no disconnect, then there is a feeling of relief and the team moves on without further reflection."
- "Learning is hindered by the tension created by judging/ratings and the need for accountability/justifying use of resources for projects. The Bank environment is competitive and focused on promotions, so people respond to ratings and this hinders learning."

Box 4-7. Grades and Learning

Educational scholars long have studied the effects of feedback in general and grades in particular on students from elementary school to college. The literature is far too vast to be summarized here, but a few noteworthy themes are worth highlighting.

First, feedback has a powerful influence on student learning and achievement. To be useful, feedback should be frequent, specific, and on a small chunk of course content. It should be timely to help students pay attention to further learning while it still matters.

Second, grading shapes incentives in powerful ways and tends to dominate students' focus and interest. A number of studies have described students receiving their assignment back, glancing at the mark at the bottom, and then throwing it away, including all the feedback. "Students may tackle essays that are intended as learning activities so as to maximize grades they obtain rather than maximizing the learning achieved from engaging with the assignment" (Gibbs and Simpson 2005). Likewise, studies of higher education students have found them to spend considerable effort on discovering what portion of the curriculum that is likely to appear in exams.

Third, a "grading orientation" is different from, and in many ways opposed to, a "learning orientation." Extrinsic motivation (desire to get better grades) can undermine intrinsic motivation (desire to learn for its own sake) even in higher education, for example by inducing a preference for easier tasks, avoidance of unnecessary intellectual risks, and a tendency for skimming books for what is likely to come up in tests. Grade-oriented environments experience increased levels of cheating, and fear of failure even in high-achieving students.

Educational institutions have been slow to take note of these findings. Some have responded by providing more frequent and focused feedback, and some medical schools and many Ph.D. programs have moved to pass/fail systems rather than grading.

Sources: Anderman and Murdock 2007; Crooks 1933; De Zouche 1945; Gibbs and Simpson 2005; Kirschenbaum, Simon, and Napier 1971; Kohn 1999a, 1999b, Pulfrey and others 2011.

Summing Up

Self-evaluation generates some individual learning but the potential of the systems for organizational learning is unfulfilled. Knowledge from systems is rarely valued or used, except by IEG, and there is little effort to extract and synthesize evidence and lessons or to inform operations. Lessons have a justified reputation for being of low value.

The systems' focus on accountability drives the shape, scope, timing, and content of reporting and limit the usefulness of the exercise for learning. Reporting against objectives for all individual projects at closing makes sense from an accountability perspective, but does not foster learning and has become a source of tension and perceived rigidity. Staff often feel unfairly assessed, making them disassociate from the process and the information it generates.

These shortcomings have to be understood within the context of a corporate culture that often rewards delivery over learning. Parts of the system not focused on accountability such as impact evaluations and other voluntary self-evaluations produce far more learning, indicating that when conditions are right, the World Bank Group has a strong demand for evaluative learning and a robust ability to supply it.