

Chapter 3

Evaluation Highlights

- The Bank's overall approach to M&E has many strengths.
- Progress has been made in updating policies and frameworks, but there is considerable room to improve how M&E is put into practice.
- At the project level, the overall quality of M&E is low.
- At the country level, results frameworks are increasingly produced but are often poorly formulated.
- The independence of external evaluations of global programs is improving, but their M&E systems are often weak.
- Impact evaluations are a useful addition, but topics need to be chosen strategically.



Child studying in public school in the Amazon region of Brazil; photo by Julio Pantoja, courtesy of the World Bank Photo Library.

Underpinning Impact— M&E and Results Management

Monitoring and Evaluation Systems at the Project Level

The Bank's overall approach to M&E has many strengths (see appendix B for an overview). Indeed, its policies for monitoring and evaluating projects have been revised, to place greater emphasis on project outcomes. Specifically, the Bank:

- Required (a) investment projects to include results frameworks, outlining (final) project and intermediate outcomes as well as the process for carrying out M&E;¹ and (b) development policy operations to specify expected outcomes and measurable indicators for M&E in 2004 (World Bank 2004a).
- Replaced the Project Supervision Report with the Implementation Status and Results report (ISR) in 2005. The ISR now gives more prominence to project outcome and intermediate outcome indicators by including them in the main report, whereas in the past they were in an optional annex (infrequently updated).
- Harmonized the Bank and IEG's evaluation criteria for Implementation Completion and Results reports (ICRs) and IEG's ICR reviews in 2006. The procedures for programmatic development policy loan ICRs were also simplified to improve the effectiveness of the ICRs.

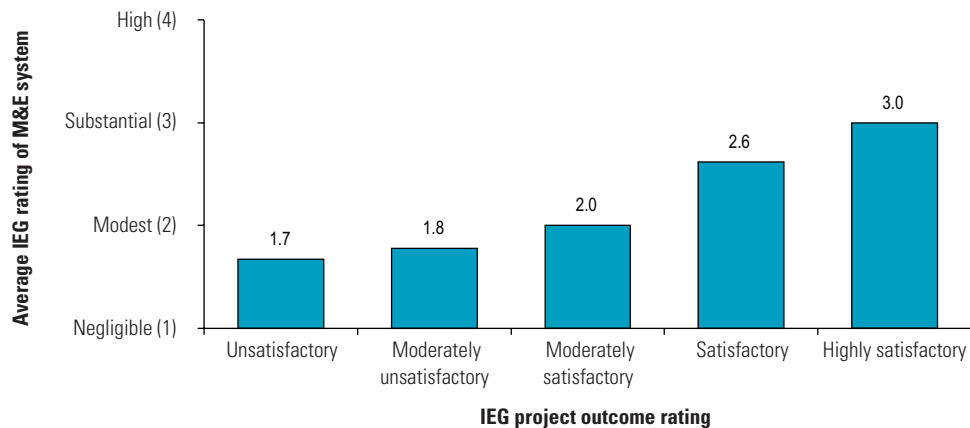
As part of the revisions, ICRs are now required to include an assessment of project M&E quality along three dimensions—design, implementation, and utilization. The review of M&E *design* examines the extent to which adequate indicators

were identified to monitor progress toward project development objectives. The assessment of M&E *implementation* is the extent to which appropriate data were actually collected, and *utilization* reviews the extent to which appropriate data were evaluated and used to inform decision making and resource allocation. IEG, as part of its ICR reviews, has started rating overall M&E quality using a four-point scale: high, substantial, modest, and negligible.² This quantitative assessment does not reflect a methodology that has been agreed to by Bank management; ICRs assess, but are not required to rate, M&E quality.

A review of the quality of project M&E showed a positive association between good project M&E and better project outcomes. It found that projects with highly satisfactory project outcome ratings had, on average, higher M&E quality rating (3.0 or equivalent to a rating of “substantial”) compared with projects with unsatisfactory outcome ratings that had lower M&E quality (1.7 or equivalent to a rating of “negligible”) as shown in (figure 3.1).

There is a positive association between good project M&E and better project outcomes, but the overall quality of project M&E is low.

Figure 3.1: Projects with Higher Outcome Ratings Have Better M&E Ratings



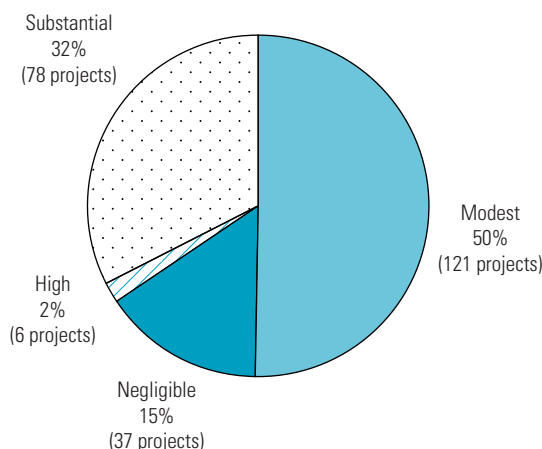
Source: IEG.

But the overall quality of project M&E has been low. Of the 242 ICR reviews for which M&E quality ratings are available,³ IEG rated that quality as modest or negligible in two-thirds of cases, as shown in figure 3.2. Poor design was a key factor in low M&E quality ratings. Some projects had weak results frameworks where the project outcomes were poorly defined and/or the link among the project outcomes, that is, the project development objectives, intermediate outcomes, and the project outputs was not clear. Another factor was poorly designed M&E

systems, where the monitoring indicators lacked baselines and targets and not enough attention was given to implementation.

Projects with high M&E quality ratings included M&E systems that were well designed, implemented, and used. The Second Rural Roads Project in Peru included performance monitoring indicators for outputs, intermediate outcomes, and outcomes that were conceptually clear, realistic, and measurable and were linked closely to the project’s four main objectives. Impact evaluations were designed into the project, and the evaluation of the previous project provided the baseline for this project. The M&E framework also constituted an active learning process for the project’s executing agency. The project was rated satisfactory for outcome. The Partnership for Polio Eradication Project in Pakistan was rated highly satisfactory for project outcome and high for M&E quality. The project was able to use existing systems, combined with substantial external technical and financial support, to develop a system and create demand for data, which are essential for disease eradication.

Figure 3.2: M&E Is Rated Modest or Lower in Two-Thirds of ICR Reviews



Source: IEG.

The impact of the changes in Bank M&E procedures may not yet be fully reflected in this analysis because most projects approved by the Board since fiscal 2005 are still under implemen-

tation. Of the 242 projects with M&E quality ratings, 23 have been approved since fiscal 2005, of which 21 are development policy operations. Close to half (48 percent) of these projects were rated substantial or high for M&E quality. However, it is not clear whether this is due to better M&E quality ratings for development policy operations or because of improvements in M&E policies. A clearer picture will emerge in subsequent years as more projects from fiscal 2005 and afterward exit the portfolio and have ICRs and ICR reviews prepared.

It may not necessarily be policies and procedures, but rather a lack of incentives and priority that is constraining the design and use of M&E systems. As far back as 2000, a Bank working group on improving M&E found that M&E tended to be perceived as the least important of the dimensions of operational quality (World Bank 2000). Focus group participants for IEG consultations stated that they sometimes set ambitious development objectives and then faced difficulty linking them to projects, that there was excessive emphasis on quantitative data, and that monitor-

ing got less priority during implementation. Difficulties in M&E have also been highlighted in several IEG sector and thematic evaluations, as noted in box 3.1.

Many of these IEG findings are echoed in reports from the Bank's Quality Assurance Group. The QAG's most recent *Quality of Supervision (QSA7)* assessment noted satisfactory focus on management of inputs and outputs, but less so on outcomes (World Bank 2007a). The QAG's latest quality at entry assessment (World Bank 2008b) found improvements in results framework design (where more than 95 percent of the project were rated marginally satisfactory or better), but also observed that project development objective clarity and realism, consistency of project design with outcomes, and impact and outcome measurement could be improved. The Bank is proposing to strengthen monitoring under IDA15 by tracking the quality of the project development objectives at design and the adequacy of baselines during implementation (World Bank 2007b).

Poor M&E impairs the ability to manage ongoing projects and subsequently to evaluate them robustly.

Box 3.1: M&E Findings and Recommendations in Recent IEG Evaluations

IEG's evaluation of the World Bank's project-based and World Bank Institute training (IEG 2008d) found that Bank projects and the World Bank Institute both report on outputs (the number of people trained, the number of training days, participant satisfaction with training), but seldom on outcomes (changes in workplace behavior, institutional capacity development). It recommended that the Bank: (a) develop guidance and quality criteria for the design and implementation of training, to enable quality assurance and monitoring and evaluation of all its training support; and (b) improve the quality and impact of training by making available technical expertise on the design, implementation, and monitoring and evaluation of training to its staff and borrowers.

IEG's evaluation of World Bank assistance to agriculture in Sub-Saharan Africa (IEG 2007d) found that data systems and support for M&E have been insufficient to adequately inform the Bank's efforts to develop agriculture in Africa. M&E at the project level is often of limited value for answering fundamental outcome, impact, and efficiency questions, such as who benefited, which

crops received support and how, what has been the comparative cost effectiveness, and to what can one attribute gains. It recommended that the Bank improve data systems to better track activities it supports and strengthen M&E to report on project activities in various agro-ecological zones for different crops and farmer categories, including women.

IEG's recent evaluation of the Bank's Group's work in supporting environmental sustainability (IEG 2008a) recommended that the Bank Group, as a whole, improve its ability to monitor and evaluate the impact of its environment-related interventions. This included using current International Finance Corporation and Multilateral Investment Guarantee Agency evaluation systems as a starting point for better monitoring for the whole Bank Group. It could further mobilize efforts for: (a) better baseline environmental assessment studies; (b) more holistic evaluation methods; (c) further development of environmental performance indicators; and (d) improved monitoring and reporting on project implementation and results.

Country-Level M&E: Early Evidence from Results-Based CASs

The Bank adopted the results-based Country Assistance Strategy (CAS) approach in 2005,⁴ following a review that found that the evaluation framework remained the weakest area of the CAS, with the key issue being the “missing middle” between the CAS objectives and Bank operations (World Bank 2003a). A results-based CAS would be “a country assistance strategy that contains a strong orientation toward achieving realistic outcomes and results-oriented monitoring and evaluation system” (World Bank 2005a).

The Bank strengthened the results focus of its CASs.

The two main innovations of the results-based CAS approach were the CASCR as a self-evaluation tool and the results framework as a monitoring, management, and evaluation tool for the Bank country program. The results framework is expected to link higher-level country outcomes with Bank CAS outcomes, intermediate outcomes/milestones, and operations, and would include measurable indicators of progress that can be tracked through CAS implementation, encouraging active management and allowing both self- and independent evaluation (World Bank 2005a).

Experience with this new approach suggests that, so far, the CAS results framework typically has been used as a reference for evaluation, but not for monitoring or program management. This emerged from a review of results-based or results-oriented CASs and interviews with the Task Team Leaders of the reports to examine the extent the results framework was used for monitoring, management, and evaluation and to enhance the Bank’s development effectiveness. CAS results frameworks served as an inventory of commitments in the CAS Progress Reports and were used to report on the status of CAS outcomes and/or

intermediate outcomes without further analysis. The results framework was also used as a reference in evaluating the country program in the CASCRs, which included an annex with the current status of the CAS outcomes. One exception is Armenia, which in-

corporated the CAS outcomes into their country portfolio performance review and is jointly managing the Bank program with the borrower to focus on CAS outcomes (box 3.2). Poorly designed results frameworks limited their usefulness for country program monitoring and management. It is difficult to interpret the information on a project’s status if the expected outcome is vaguely defined (“banking system financial capacity strengthened”) and/or lacks baselines and targets to measure progress. Another issue is the volume of information to be provided. Assembling the data has proven challenging; the Task Team Leaders for the Progress and Completion Reports mentioned that they had to devote a significant amount of time and effort to collecting the data and updating the results framework. In some cases, the amount of information provided was overwhelming and yet failed to answer the basic question: “Is the Bank making sufficient progress toward achieving its CAS objectives.” The Ghana Progress and Completion Reports addressed this issue by reporting actuals only for the 35 CAS monitoring indicators, which (except for those under the Governance Pillar) had baselines and targets.

IEG has also raised concerns over the poor design of CAS results frameworks. Many CAS

Box 3.2: Armenia Joint Country Portfolio Performance Review

As preparation for the 2007 CAS Progress Report, the Bank conducted a joint portfolio review with the government of Armenia, which covered not only the overall status of the portfolio, implementation issues, follow-up on previous commitments, and individual problem projects, but also the progress toward CAS objectives and the contribution of Bank instruments. Information was prepared on the status of 29 CAS outcomes and discussed with the government. Recommendations included specific steps, typically broader and policy-oriented, which are needed to improve project performance and to ensure achievement of CAS outcomes.

CAS results frameworks are used as a basis for evaluation but less so for country program monitoring and management.

results frameworks had poorly articulated results chains—including in several IDA countries where the distinction between CAS outcomes and intermediate outcomes/milestones was unclear and the links between Bank products and CAS outcomes were poor. Many results frameworks had too many outcome measures—more than 40 CAS outcomes indicators and over 60 intermediate indicators—and/or lacked indicators with baselines and targets, making them less effective as a management and evaluation tool.

The lack of outcome data makes it more difficult to assess the achievement of Bank CAS objectives in the CASCR reviews. While most of the CASs reviewed to date were prepared before the results-based CASs were introduced, some did include expected outcomes and/or results frameworks that were retrofitted in the CASCRs. IEG reviewers noted that the CASCRs discussed what the Bank did (process, inputs and outputs), but did not necessarily provide convincing evidence of the success of the Bank country program in achieving its objectives.

Whether staff have the right incentives to promote effective results-based frameworks for CASs remains open to question. In interviews for this report, some Task Team Leaders talked of the difficulty of getting support from their sector colleagues when updating the information in the CAS results frameworks. Some sector staff did not see value in the monitoring indicators because they were not involved in the formulation of the original CAS results framework. Given the demands on their time, monitoring of the CAS results framework and analyzing Bank country performance gets low priority, reflecting some skepticism about the value added by the exercise. This echoes findings from previous IEG evaluations (IEG 2004a, 2006d) which concluded that “getting results” was not yet part of the reward system for individual staff and that the Bank’s incentives rewarded the work that the staff do at the early stages of the project cycle—that is, preparing new operations—more highly than work at the later stages of supervision, such as evaluation and learning lessons.

One of the lessons learned from the Progress Report and CASCRs was the need to better align Bank activities with CAS outcomes. The Armenia CAS Progress Report noted that it is better to build the results framework around Poverty Reduction Strategy Paper outcomes and indicators. Baselines are important and other indicators should be sought if baseline data are not available, and data sources should be verified to make sure that information can be made available in a timely manner. The Ghana CASCR observed that one of the weaknesses was that some indicators were only indirectly related to Bank interventions. The Mozambique CASCR also commented on the importance of the Poverty Reduction Strategy Paper, but noted that the CAS should set more realistic targets, distinguishing them from Poverty Reduction Strategy Paper targets and choosing such targets only if they are related to the Poverty Reduction Support Credit (PRSC) or where the Bank is the main source of support to achieve the Poverty Reduction Strategy Paper target.

Another lesson was the importance of strengthening M&E capacity at the sector or project level. The Mozambique CASCR noted that projects also need to strengthen their monitoring of results, and focus on helping sectors to develop monitoring and evaluation systems, which would help project data collection and reporting. The Bosnia-Herzegovina CASCR made a similar point, noting that the focus on results should go beyond Bank-financed operations and that the Bank teams should help their counterparts to develop appropriate results frameworks for all their operations.

Linking the CAS and project results frameworks and strengthening the country’s capacity to conduct M&E would not only reduce the cost of data collection, but also increase its relevance and use.

While the discussion focused on countries with CASs, M&E is also important in Low-Income

Poor design limited the usefulness of the results frameworks for Bank’s country programs.

Incentives favoring project preparation and delivery over M&E continue to be a concern.

It is important to link CAS and project results frameworks, and strengthen country capacity to conduct M&E.

M&E is just as critical in Low-Income Countries Under Stress. Countries Under Stress (LICUS), which are covered by Interim Strategy Notes or Transitional Support Strategies. An IEG evaluation (2006e) concluded that M&E is just as critical, maybe more so, in LICUS (box 3.3).

Going forward, country teams should examine results chains and explore synergies among various interventions in country programs (projects, analytic and advisory activities, trust-funded activities, and so on) to identify shared higher-level outcomes. This could lead to a reduction in the number of country outcomes to be measured, making the frameworks more strategically focused and usable. Teams could also

The Bank could take steps to strengthen the results frameworks and better manage for results. identify areas where the project results framework could be linked into the CAS results framework. Establishing synergies between interventions and identifying possible gaps in the Bank program could go toward reducing the disconnect between country and project portfolio outcomes.

The basics need to be ensured. CAS results frameworks need to have monitoring indicators with baselines and targets to measure progress.

Systems need to be in place to ensure that the information is collected. This may require trading off data availability with desirability.

Finally, integrating the CAS results framework into Bank operational procedures would establish the use of the performance data and could provide an incentive to update and utilize the CAS results framework. There are some pilots under way. At the country level, the Moldova Country Management Unit is piloting a scorecard to link CAS outcomes, operational performance, and Bank budget allocation (box 3.4). At the regional level, the East Asia and Pacific Region piloted and is expanding an approach to link the achievement of CAS objectives with budget allocation (box 3.4). Linking Bank resource allocation with progress toward CAS outcomes would also provide an incentive to monitor and analyze Bank country performance indicators. The Bank could learn from and build on these pilots.

Managing Global Programs and Partnerships: An Emerging Agenda

The Bank has put in place new business processes for global programs and partnerships (GPPs), which recognize these as a separate product line of the Bank, and which aim to integrate GPP business processes and information with the Bank's regular operational business systems (IEG 2004b). Throughout their life cycle, GPPs now follow procedures similar to a simplified lending process—using familiar concepts and systems and allowing new GPPs to be tracked from their start through to evaluation and results assessment.

Enhanced processes have also been put in place to encourage greater selectivity although their effect remains to be assessed fully. Proposals for new GPPs and requests for new Development Grant Facility (DGF) funding are now presented in a standard Partnership Review Note and considered, based on input from relevant parties in a review meeting chaired by a director. The GPP Group and the Legal Department are providing early advice to task teams with regard to governance, which focuses on the quality of the governance structure (how the partners will form the partnership and interact with each other),

Box 3.3: Results Measurement and Monitoring in LICUS

Monitoring and evaluation are at least as important in LICUS as they are in any other country. Monitoring and evaluation are crucial in LICUS for a number of reasons. First, the Bank, like other donors, is still learning what approaches work in LICUS contexts. Closely monitoring experiences in order to draw lessons is critical, and learning and sharing needs to become a more prominent feature of LICUS work. Second, given that progress is often slow in these countries, it is important to reassess continually whether the program is on course to achieve the desired outcomes. Third, a constantly changing and volatile LICUS environment where progress is often nonlinear means that program adaptation is essential—closely tracking performance will help determine when and what kind of adaptation is necessary. Effective learning-by-doing to improve the Bank's future effectiveness in LICUS can only happen with strong monitoring and evaluation.

Source: IEG 2006e.

Box 3.4: Use of CAS Results Frameworks in Country Program Management

Moldova Results Scorecard

The Moldova Country Management Unit is developing a Moldova Results Scorecard which will integrate, in one place, the different elements of country program management: progress toward country strategic outcomes, lending and delivery costs of analytic and advisory activities, and quality (QAG and IEG assessments). The scorecard will be used to link country program management and resource allocation and improve program and project results.

Strategically Managing Country Programs in East Asia

The East Asia and Pacific Region piloted the use of

CAS results frameworks to increase management accountability for results. The new approach is grounded in CASs and, at the same time, focused across core Regional priorities and actionable milestones. It is intended to be consistent across Regional and country levels and enables Regional aggregation to allow discussion with senior management. Country teams took their respective CAS results frameworks, agreed on the critical CAS outcomes where they would focus their efforts, and narrowed the monitoring indicators to a manageable number in a one-page table. This information was used in the annual and midyear program/strategy discussions.

legitimacy, the voice of developing countries, and the Bank's roles (particularly whether the Bank's accountability and responsibility are aligned with the Bank's formal authority and actual control). A new results framework and performance indicators were added to the Partnership Review Note and the DGF Progress Report on a voluntary basis in fiscal 2007, and made compulsory in fiscal 2008.

Quality-at-entry reviews of DGF-supported programs by QAG also provide feedback that contributes to enhanced selectivity. QAG noted several improvements in fiscal 2007, as compared with fiscal 2006, including stronger quality of program design and improved readiness for implementation, as well as a number of areas needing improvement, such as unrealistic objectives, weak results frameworks, and poor documentation.

Taken together, these represent a substantial improvement over the previous partnership approval and tracking system for GPPs, which was not integrated into the Bank's regular operational business systems. However, to date the improvements have mainly occurred with respect to DGF-supported programs, since these programs have to comply with the new business processes in order to receive grants. Except for DGF grants received, the financial information in the Bank's databases (on the sources and uses of funds for each program) is not complete or reliable. None of

the Bank's existing information systems include financial resources that are not channeled through the Bank, and the Bank's Task Team Leaders are not required to provide or update this information on an annual basis.

The requirement for all programs receiving DGF funding of \$300,000 or more, over the life of the program, to undertake an independent evaluation every three to five years has led to an increasing number of program-level evaluations. However, the Bank has not yet extended this requirement to the many other programs not receiving DGF support, as previously recommended by IEG (2002), though it is currently revising the relevant Operational Policy to achieve this. Implementing such a requirement should be fairly straightforward for programs supported by Bank-administered trust funds. For programs that are not supported either by the DGF or by Bank-administered trust funds, the Bank should still use its influence, as a member of the governing body, to encourage periodic independent evaluations.

The independence of external evaluations is improving. Out of the first seven evaluations that IEG reviewed, five were executed independent of the management of the program, and without

Bank management has put in place new business processes that aim to integrate GPPs with the Bank's regular business.

Feedback from QAG quality-at-entry reviews of DGF-supported programs is contributing to enhanced selectivity.

The independence of external evaluations is improving. apparent conflicts of interest. The governing bodies generally commissioned the evaluation, approved the terms of reference, managed the selection process, and reviewed the draft report independent of program management. In four out of these five cases, the selection of the external evaluators was competitive.

The quality of some evaluations was compromised by a weak M&E system for the program. However, the quality of three of the seven was compromised by a weak M&E system for the program. Either the objectives and strategies of the program were not well defined (too diffuse, process-oriented, difficult to measure, or open to different interpretations by different stakeholders), the M&E system was not well designed (focusing only on inputs and outputs, and not outcomes), or the data on the progress of activities and on the achievement of outcomes were not systematically collected. And in three of the seven programs an inadequate budget did not permit any fieldwork to verify facts on the ground or elicit the perceptions of implementers and beneficiaries in developing countries. As a result, neither the external evaluations nor IEG found much systematic evidence relating to the achievement of the seven programs' objectives at the outcome level, and it is difficult to say whether the \$100 million spent on the programs annually ultimately had a substantial effect on the ground.

External evaluations are influencing the strategic directions and designs of GPPs. Notwithstanding these shortcomings, the external evaluations have had significant impacts on the programs reviewed. The evaluation led to a revision in the strategic direction or design of the program in six out of seven cases. The evaluation led to a change in the governance of the program in four cases, and led directly to continued or increased funding in four cases.

Improving Our Understanding of Causality: The Use of Impact Evaluations

In response to the pressures for accountability, the development community in recent years has accorded increasing attention to impact evaluation as another tool for generating knowledge

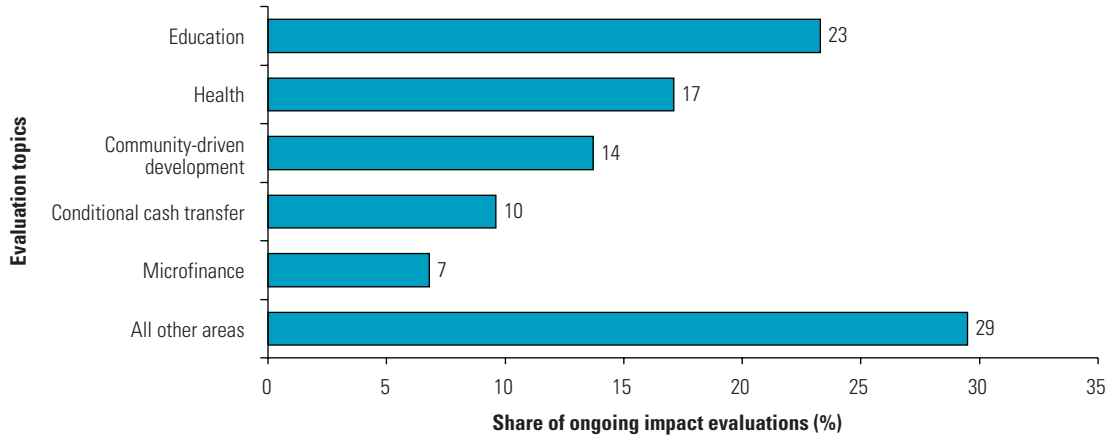
about the effectiveness of development interventions. The challenge now is how best to use this approach for increasing knowledge regarding development issues, advancing the broader results agenda, and stimulating client engagement in, and capacity for, monitoring and evaluation.

The results agenda can potentially benefit from impact evaluations, although they are clearly not a panacea. Impact evaluations explicitly identify a counterfactual through experimental or quasi-experimental designs to measure the true impact of development interventions. Accurate measurement of results can help for efficient allocation of resources and effective project/policy selection and design. The potential benefits of an impact evaluation extend beyond any particular project or program because the evaluation can provide information and lessons for the broader development community. It is possible, therefore, that impact evaluations may be underfunded unless there is explicit support from management and clients. On the other hand, it must be recognized that impact evaluations are highly specific to the context and often quite costly to launch, and their benefits are uncertain until they are completed.

The number of impact evaluations at the Bank has grown rapidly recently, with a rise from 60 under way in 2006, to 158 in mid-2008.⁵ Yet, the evaluations are concentrated in a few sectors and topics. Almost 70 percent of the ongoing evaluations are clustered in five areas, as shown in figure 3.3, and three of these—education, health, and conditional cash transfers (many of which also focus on education outcomes)—also comprise the top categories of previously completed evaluations. There is also a concentration by Region, with nearly two-thirds of ongoing evaluations located in Sub-Saharan Africa and South Asia (see figure 3.4).

Why evaluations are concentrated in certain areas and whether the observed distribution is efficient are questions for Bank management. Some substantive areas are more amenable to impact evaluations, and Bank clients differ in their receptivity to evaluations. But staff incentives may

Figure 3.3: About 70 Percent of Ongoing Evaluations Are Clustered in Five Areas



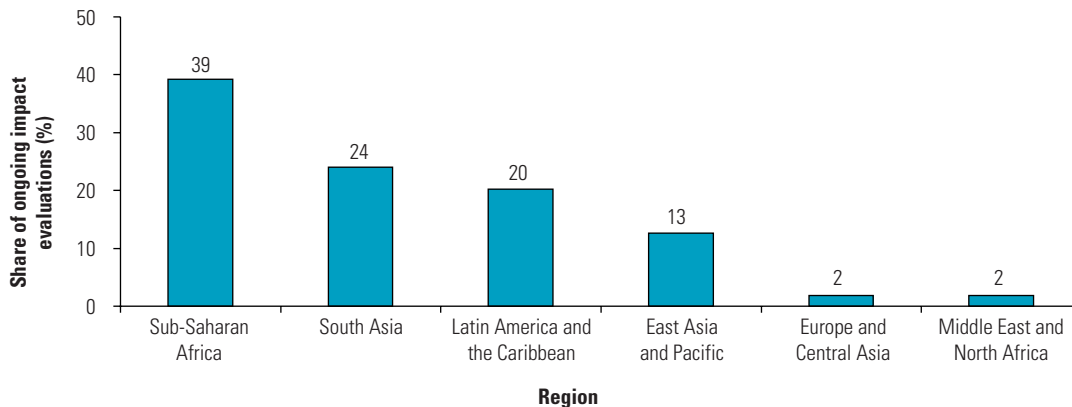
Source: Development Impact Evaluation database.

also play an important role, and care must be taken not to overexamine some topics while others are underresearched. In an ideal setting, the decision to fund impact evaluations in a given area would take into account one or more of the following five criteria: the value of answering the question in terms of benefits and costs of a specific project, the value of answering the question for other current or future projects, the cost of the evaluation, the innovative nature of the project, and the likely feasibility of designing a convincing impact evaluation.

The Bank has the capacity and resources to be a global leader in using impact evaluations to drive effective development assistance. Required steps include identifying important gaps in our knowledge; closing these gaps with effective and replicated evaluations; leveraging knowledge from impact evaluations through effective discussion, dissemination, and application across the institution; and shifting expenditures from unsuccessful to successful interventions. Clients should be closely engaged and ideally, in many

Impact evaluations are not a panacea, but they are an important part of the results agenda.

Figure 3.4: Nearly Two-Thirds of Ongoing Evaluations Are Located in Two Regions



Source: Development Impact Evaluation database.

Impact evaluation topics should be chosen strategically and the results should influence future funding decisions.

cases, leading the process of implementing evaluations and learning from their findings. A major challenge facing the Bank is to assess and compare the relative effectiveness of alternative interventions rather than largely assessing the effectiveness of different ways of implementing a given intervention. Existing initiatives, such as the Development Impact Evaluation Initiative, have made some progress toward these goals.

IEG itself is aiming to increase its own use of impact evaluations and capacity to review impact evaluations within the Bank. Through meta-evaluations and mixed methods approaches, impact evaluations provide important opportunities to enhance IEG's evaluation role.

Monitoring Institutional Effectiveness

The development community is focusing on results on the ground, with the Millennium Development Goals serving as a common framework. In 2002, the heads of the multilateral development banks issued a joint statement on their commitment to scale up their work on measuring, monitoring, and managing for results, following the Monterrey Conference on Financing for Development.

The development community is focused on results on the ground and is interested in the impact of Bank operations.

The IDA15 replenishment negotiations distilled this focus on results in a very real way. During the negotiations, the Bank faced competition for donor funding from different organizations and programs, many of them focused on specific issues. Donor governments had to answer to their citizens on the effective use of their resources and therefore requested that IDA demonstrate that it was achieving results on the ground and what those results were. IDA's competition often focuses on single issues—such as children, health, and the environment—and are better able to present

aggregated output data (such as numbers of beneficiaries assisted, goods provided, and so on) and to describe their results. The Bank, on the other hand, carries out not only projects but also additional activities, such as policy

and institutional development or donor coordination; impacts are harder to measure and attribute to the Bank. In response, the report “Additions to IDA Resources: Fifteenth Replenishment” (World Bank 2008c) served to illustrate accomplishments in individual projects, sectors, and countries, thus complementing the information provided in the IDA Results Measurement System (RMS).

The IDA RMS assesses Bank-wide performance, but only for IDA countries. The RMS has two tiers. *Tier 1* focuses on monitoring country outcomes and includes 14 country outcome indicators that are consistent with the Millennium Development Goals, are priorities in poverty reduction strategies, and reflect IDA's activities. These are high-level country outcomes that are influenced by the Bank, but for which the Bank is not solely responsible. *Tier 2* monitors IDA's effectiveness. It consists of indicators: (a) at the country level, where the cumulative introduction of results-based CASs in IDA countries is monitored; (b) at the project level, where four project quality indicators are monitored; and (c) outputs from completed projects in four sectors (health, education, water supply, and rural transport).

IDA prepared a progress report on the RMS in 2007 and made suggestions for improved monitoring under IDA15. It reported, among other things, that IDA had built 7,500 km of road, trained 81,400 health professionals and 282,500 teachers, and made 87,000 new water connections based on IDA project ICRs in fiscal 2006–07. The report concluded that the introduction of the RMS “has generated a sharper focus on results at the country level and a stronger internal results culture.” It also acknowledged that internal process weaknesses remain in the quality of Tier 2 outcome indicators and the availability of baselines for all outcome indicators at entry in IDA projects. IDA recognized that because “many ICRs still lack figures on outputs, or report them in a variety of forms (different units, no standard categories), aggregation remains very challenging.” For IDA15, the report proposed that IDA track the quality of the project development objectives at design, the adequacy of baselines during implementation, and the

The IDA Results Monitoring System and AfricaRMS provide a broader view of results, but not for all countries.

reporting quality of outputs and outcomes in ICRs, based on the original results framework as part of Tier 2. The report also suggested including ratings on CAS implementation from IEG's CASCR reviews to measure the quality of Bank country programs.

The Africa Region developed the Africa Results Monitoring System (AfricaRMS) as a comprehensive online system in the Bank for monitoring on-the-ground results of Bank activities. It consists of a set of harmonized project and country indicators to facilitate counting and aggregating results at the country, sector, Region, and Bank-wide levels. AfricaRMS partnered with the education, malaria, HIV, water, private sector development, governance, agriculture, energy, and roads teams to build monitoring frameworks that are fully aligned with existing global initiatives. It also includes write-ups to tell the story behind the numbers, describing how good results are changing people's lives, and descriptions of impact evaluations carried out in Africa. The AfricaRMS is aligned with the Africa Action Program.

The goal of the AfricaRMS is to use select information to show the results of Bank operations and development progress for both

the Region and individual countries. It not only helps teams focus on achieving results but also provides internal and external clients access to live information about stories from projects and beneficiaries, impact evaluation results, and data for priority sectors.

The system is still being refined and faces several challenges. The Region is working to improve the availability of country data and harmonized indicator measurement across Bank operations. It is also making efforts to better integrate financial data, and output and outcome data from operations at the sector and country level. Measurement systems across countries and among development partners will also need to be harmonized.

IEG's mandate is to evaluate the outcome of Bank-supported projects and programs, as measured against their objectives, and its work—including ARDE 2008—obviously represents a part of the Bank Group's efforts to monitor institutional effectiveness. An assessment of IEG's effectiveness, including the perspectives of its clients and the extent to which its recent recommendations have been implemented, can be found in appendix C.