



Assessing Our Work on Impact Evaluation

Markus Goldstein, Emmanuel Skoufias, and Ariel Fiszbein

Over the last several years, the World Bank has increasingly engaged in impact evaluations as means of building evidence for results. During this process, the Bank has also produced an extensive variety of knowledge products. However, there are several institutional and resource issues that constrain the effectiveness of our impact evaluation work. This brief outlines recent gains in the Bank's work on impact evaluation, highlights several issues, and proposes some options to continue improving and expanding the Bank's efforts in this area.

Introduction

Demand for evidence of the results of development assistance is increasing. Donors and country partners alike are keen to see demonstrable results from the programs and interventions that are supported through Bank assistance. In addition, there is increasing scrutiny outside of the donor community. For example, academics are arguing for more rigorous evaluations¹ and the Center for Global Development is spearheading an independent evaluation initiative. Finally, as the Bank scales up its interventions, the ability to demonstrate the impact of aid on development outcomes is critical for attracting additional support, particularly from bilateral donors. In response to these and other pressures, the Bank has increased its emphasis on the use of monitoring and evaluation tools.

Among monitoring and evaluation techniques, impact evaluation is an important tool for showing the effect of interventions. The rigorous nature of impact evaluation provides a clearly attributable measure of the effect of interventions and hence is a powerful tool to show their value. In addition, the rigor of this approach strengthens the argument for the portability of lessons, allowing the program design and evaluation results in one country to influence others.

Given the power of this tool, the Bank is supporting an increasing number of impact evaluations (figure 1). Moreover, the knowledge generated through impact evaluation is a public good, which is an important rationale for Bank sup-

port of impact evaluation of interventions that are not directly funded by the Bank.

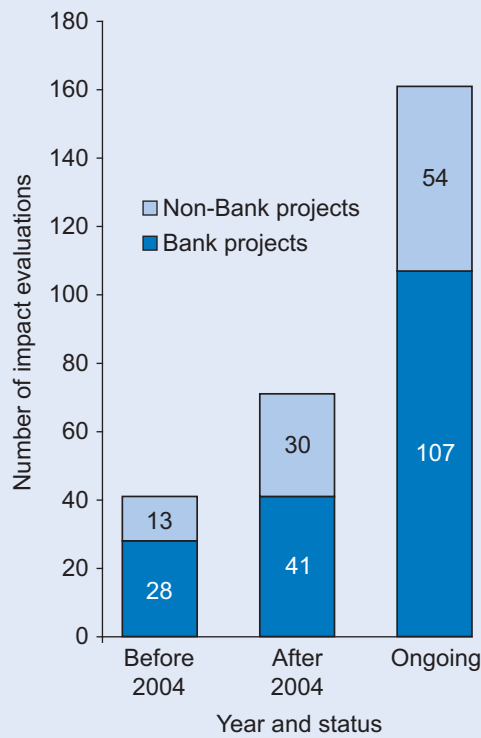
Of the ongoing and completed impact evaluations shown in figure 1, over one third are cases where the Bank is engaged in the impact evaluation, but is not providing financing for the intervention being evaluated.² Thus there

Box 1: Impact evaluation: definition and methods

Impact evaluations compare the outcomes of a program against a counterfactual that shows what would have happened to beneficiaries without the program. The counterfactual corresponds to a group that is as similar as possible (in observable and unobservable dimensions) to those benefiting from the program. The counterfactual group is established by either assigning the program randomly (experimental designs) or through a range of statistical techniques (quasi-experimental techniques). Unlike other forms of evaluation, impact evaluations allow attributing observed changes in outcomes to the program being evaluated.

While in many cases randomization does reduce the number of assumptions necessary to draw causal inferences, it is not feasible or desirable in all situations. The general best-practice approach at the Bank has been to use whichever experimental or quasi-experimental technique works best and to be explicit about the limitations of the results obtained.

Figure 1: World Bank Impact Evaluations, by Year and Status



Source: World Bank Impact Evaluation Database at: <http://www.worldbank.org/impactevaluation>.

are two overlapping roles for impact evaluation within the Bank: first, as a tool to evaluate our own work, and second, as a tool and service we can provide for others. As bilateral donors and national governments increase their demand for rigorous evidence, we can expect this latter role to assume increased importance.

A number of key actors have helped to promote this large increase in impact evaluation over the past few years. The Development Impact Evaluation (DIME) initiative³ of the Chief Economist has organized and help coordinate clusters of evaluation on similar topics, while the Poverty Group of the Poverty Reduction and Economic Management (PREM) Network, through the Thematic Group on Poverty Impact Analysis, Monitoring, and Evaluation, has helped develop staff capacity through a wide range of knowledge products. The Bank's Development Research Group (DECRG) provides additional capacity building and is currently undertaking a significant number of impact evaluations. The Results Secretariat of the Operations Policy and Country Services Network (OPCRX) led

the establishment of impact evaluation as a recognized AAA product, thereby providing a formal venue for this work to be recognized. A large share of current evaluations has been initiated in the networks and the Regions. On the network side, Human Development (HD) is leading a large program of impact evaluation and regional workshops. Regions are pursuing a range of strategies, from the centralized Africa Impact Evaluation Initiative to the more decentralized efforts of South Asia (SAR), East Asia and the Pacific (EAP), and Latin America and the Caribbean (LAC).

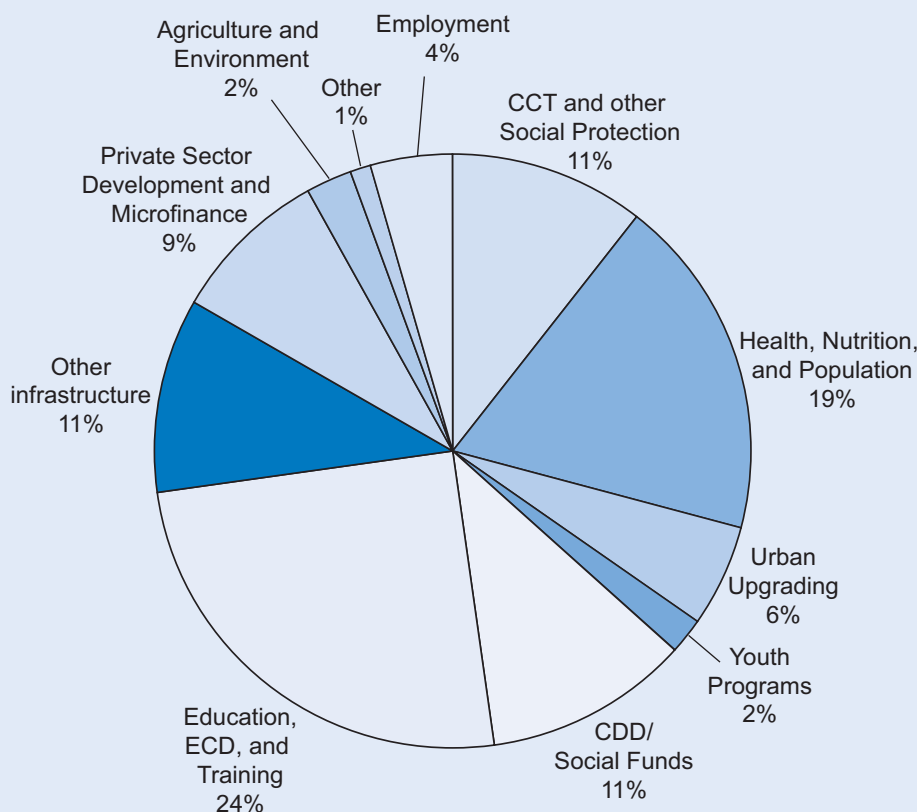
Central Issues

Although the number of impact evaluations is growing overall, some Regions and networks are more active than others. Given that knowledge gaps about what works and what does not exist in all sectors and Regions, there is a fair amount of work to be done to encourage staff to include impact evaluation in their work, especially in the Sustainable Development Network (SDN), Private Sector Development Network (PSD), and the Europe and Central Asia Region (ECA). Most ongoing impact evaluations are in the social sectors (figure 2), which reflects not only the support provided by the HD Network, but also that there is more of an evaluation tradition in these areas and that the projects are more amenable to impact evaluation techniques. Appendix 1 shows the breakdown of impact evaluations by sector against lending by sector for FY05–07.

The Regional picture is also a skewed one. Africa is the leader with 61 ongoing⁴ evaluations, followed by SAR (39), LAC (32), and EAP (23). Middle East and North Africa (MENA) and ECA have 3 evaluations each.⁵ Appendix 2 provides a lending comparison for the regional breakdown of evaluations. A significant reason for Africa's lead is the Africa Impact Evaluation Initiative, which provides dedicated staff for promoting impact evaluation in the region.

While this large number of ongoing impact evaluations is encouraging, many of them are still in early stages (figure 3), which makes it essential that we maintain quality and continuity going forward. Figure 3 shows that, of the 225 evaluations now in progress in the Bank, 138 have no tangible commitment; that is, they have not reached the point of collecting baseline data (appendix 3 provides a breakdown of the status of ongoing evaluations by Region and by sector). In ad-

Figure 2: Impact Evaluations (Ongoing) by Category



Source: World Bank Impact Evaluation Database at: <http://www.worldbank.org/impactevaluation>.

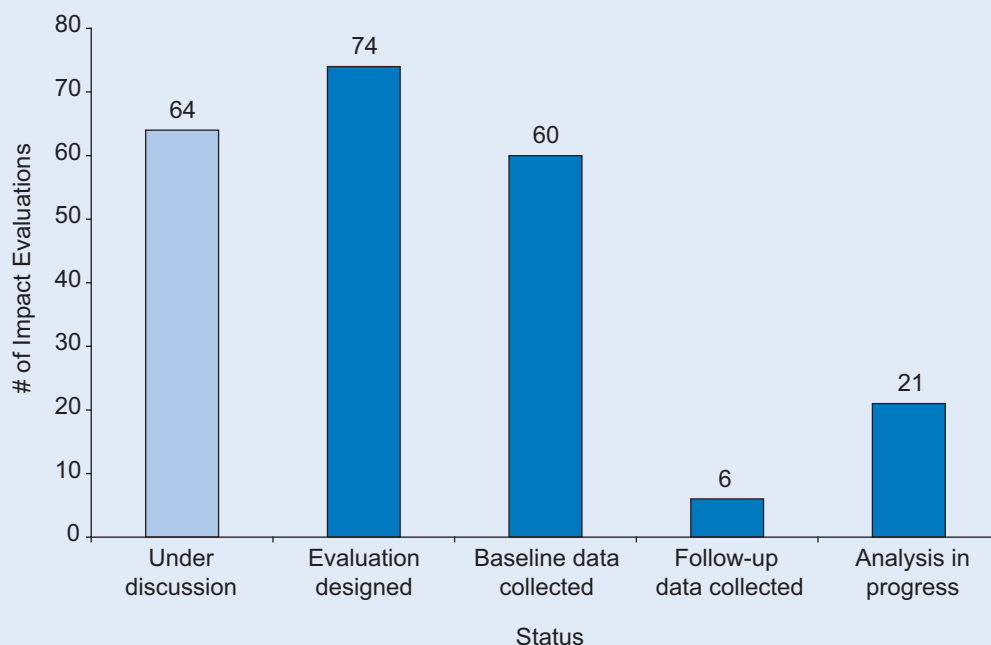
dition, the typical impact evaluation will take 4–5 years to produce findings. These ongoing evaluations potentially are a rich pipeline of evidence, but their uneven status also raises some concerns. Therefore, it is essential that we maintain continuity and quality as evaluations move along. A number of factors make quality control a challenge. First, most impact evaluations will outlast the average tenure of a task team leader (TTL) and there is no guarantee that the next TTL will be committed to this activity. This is compounded by the fact that Regional and sectoral support staff for impact evaluation are thin (if it exists at all) and thus are incapable of guiding all ongoing evaluations through to completion.⁶ Finally, despite the creation of impact evaluation as a product line, there is no central mechanism to review and monitor the quality of these evaluations

There is insufficient strategic direction to decisions about which Bank interventions get evaluated. Current practice relies mostly on the initiative

and interest of individual TTLs. Objective criteria for which interventions should be evaluated could include the following: (i) an intervention that is innovative in delivery or content, (ii) an intervention that will inform the evolution of sector strategy, (iii) an intervention that involves significant risk due to its size, (iv) the application of an intervention that has worked elsewhere to a new context (such as conditional cash transfers in Africa), and (v) the feasibility of developing a counterfactual. The solution here is for sector/regional management teams to review the pipeline of projects that meet these criteria and facilitate the evaluations.

Limited government ownership is a possible constraint to the expansion and effective use of impact evaluation. The big exception to this point comes from Latin America, where countries such as Mexico and Colombia press the Bank for help on impact evaluation. These requests include demands for the evaluation of individual interventions and also Bank support for

Figure 3: Status of World Bank Impact Evaluations



Source: World Bank Impact Evaluation Database at: <http://www.worldbank.org/impactevaluation>.

building national results management capabilities. These demands for national capacity are instructive because they suggest a model for making this work: building a role for impact evaluation into work on national monitoring and evaluation systems. Note also that impact evaluation is an area of business in which we are experiencing increasing demand from our middle-income clients (for example in LAC, but also countries such as Indonesia).

Options to Help Meet Impact Evaluation Challenges

Any strategy to address these issues will likely need to include a number of underlying attributes. First, it will be important to support the current groundswell of impact evaluations initiated by project TTLs. Such support can be delivered through appropriate incentives and resources, accompanied by a strategic direction for new or underserved areas and a commitment to maintaining quality. Second, a strategy for impact evaluations will need to balance twin, and potentially conflicting, goals: (i) country ownership, and (ii) the Bank's need to identify priority areas for impact evaluation based on a global perspective. Finally, the strategy must

ensure that lessons from impact evaluations (successes and failures both) are widely disseminated and used for policy. Below, we describe six components of a potential strategy to support impact evaluations.

1. Identify priorities for evaluation

Identifying priorities requires work on both sides of the Bank's organizing structure. Networks, through a process they determine, should identify priorities for clusters of evaluations based on gaps identified in their sector strategies. On the other side, country teams should identify priorities based on country dialogue or country program issues. Cross-sectoral coordination can be provided by DIME. For example, one of the current DIME clusters is local governance, which brings together impact evaluations from Social Development (SDV), the Least Developed Countries Expert Group (LEG), PREM, and DECRG.

2. Ensure quality

A three pronged approach could support the delivery of high-quality impact evaluations:

- Staff capacity to supervise and execute impact evaluations needs to be strengthened (discussed below). Given that the best

approach to impact evaluation involves a close and ongoing engagement between the project team and the evaluation team from the start of project design, this training is critical to ensuring quality.

- b. Support teams could be put in place. These could be located in the network VPUs or in network-Region teams, as in the example discussed in Box 2 on Africa's education initiative.
- c. The Bank's review process can be used to monitor quality as the evaluation is carried out. The impact evaluation guidelines call for a concept note and a peer review. In addition, this being a form of AAA, regional chief economists should be made responsible for ensuring quality. Currently, much of this feedback process has been going on through informal meetings. A more formal process could help ensure that, with the growing number of impact evaluations, individual evaluations get the feedback they need.

3. Maintain continuity in the face of staff transitions

Bank staff engage in impact evaluation in two main ways: supervising the execution of the evaluation (and using the results) and doing the evaluation (especially the analysis) themselves. In the latter case, the main issue for continuity will be ensuring that the replacement for the evaluator is sufficiently skilled to continue the

work on the evaluation. This suggests a human resource issue: making sure that there is a sufficient skill base across regions and sectors (or in central units). In the case where the staff member is supervising an external evaluator, the issue is again one of capacity, but this is easier to rectify through the type of training we identify below, perhaps combined with some back-stopping support from sectoral support teams.

4. Build capacity

As part of the process of increasing the quantity and scope of evaluations, it would be desirable to provide additional training to Bank staff and client governments, in addition to developing global learning materials such as the *Doing Impact Evaluation* series.⁷ Training for Bank staff could be provided at two levels. The first level would be to provide a more intuitive explanation of impact evaluation concepts and methods coupled with practical considerations of how evaluation could be used for policy (this level could be something akin to the current PREM week course). This level would equip staff to supervise and intelligently consume impact evaluations. The second level would build on the graduate school training of staff to develop their technical skills to actually carry out evaluations (this level could be like an expanded version of the current DECRCG one-week module). In terms of developing the skills of client governments to supervise and consume impact evaluations, we could provide training similar to that used for Bank staff.

Integrating this training with the development of national results management capabilities could help to not only place the tool of impact evaluation in context, but also enhance the demand for a broader monitoring and evaluation system. To build client capacity for executing impact evaluations, we propose that each Bank-sponsored impact evaluation be required to involve 1–2 local researchers and a local research institution. Given the current rate of Bank evaluations, this learning-by-doing approach should yield a worldwide cadre of 200–300 trained researchers in the next few years, providing significant progress towards one of the goals outlined at the Hanoi Roundtable on Results this past year. As these researchers interact with policy makers, they will function as powerful domestic advocates for improved monitoring and evaluation more generally.

Box 2: The Africa Program for Education Impact Evaluation

The Africa Program for Education Impact Evaluation provides one model for sectoral-regional teams on impact evaluation. Under an initiative funded by the Education for All Trust Fund, the program solicited country and TTL interest in education evaluations. Eleven countries expressed interest in participating. Program management will be provided by the regional impact evaluation team, with supervision and support from the Africa education group. In addition, a technical support team will be set up to provide help with issues in impact evaluation design, household and facility surveys, sampling, and costing of interventions. This structure will support government evaluation and evaluation research teams, which work in collaboration with Bank TTLs.

5. Learn from the results

Putting the lessons of individual evaluations together heightens their utility for policy making. Given their alignment with sectoral strategies, network teams would likely have a clear role in analyzing, aggregating, and disseminating the results of impact evaluations. However, the presence of cross-sectoral clusters of activities aimed at a common problem suggests the need for a cross-sectoral vehicle to coordinate this knowledge accumulation and dissemination. This cross-sectoral learning mechanism could also be well placed to tackle the task of comparing results across interventions, as well as evaluating programs that intervene in multiple sectors simultaneously. For example, we have clear evidence that conditional cash transfers increase enrollment and have effects on health and poverty. This has no doubt contributed to the proliferation of these programs. But what we do not know is whether CCTs are the most cost effective way to get these benefits. In addition, a more comprehensive understanding of the impacts of interventions will require a broader perspective and examination of the major binding constraints facing economic actors. For example, interventions such as CCTs to increase the demand for schooling will be less effective where the supply of schooling is constrained. It is important that we understand the constraints on the supply side and the demand side. Finally, markets may fail for reasons entirely unrelated to individual firms and people, so it is important to understand that the impact of interventions is conditional on the structure of markets. For example, it may be difficult to design interventions to match workers with jobs if labor markets are constrained by regulations such as hiring and firing restrictions or binding legislated minimum wages.

This analysis and lesson collecting should be complemented by steps to make sure that the individual results are quickly and easily available. To this end, the Poverty Reduction Group within PREM currently maintains a searchable Web-based database of over 200 completed Bank and nonBank impact evaluations. Demand for this data seems fairly strong, with over 44,000 hits last year. However, with the increasing number of evaluations both inside and outside the Bank in coming years, additional support will be needed to make sure that this database remains current.

6. Commit more resources and incentives for more and better impact evaluation

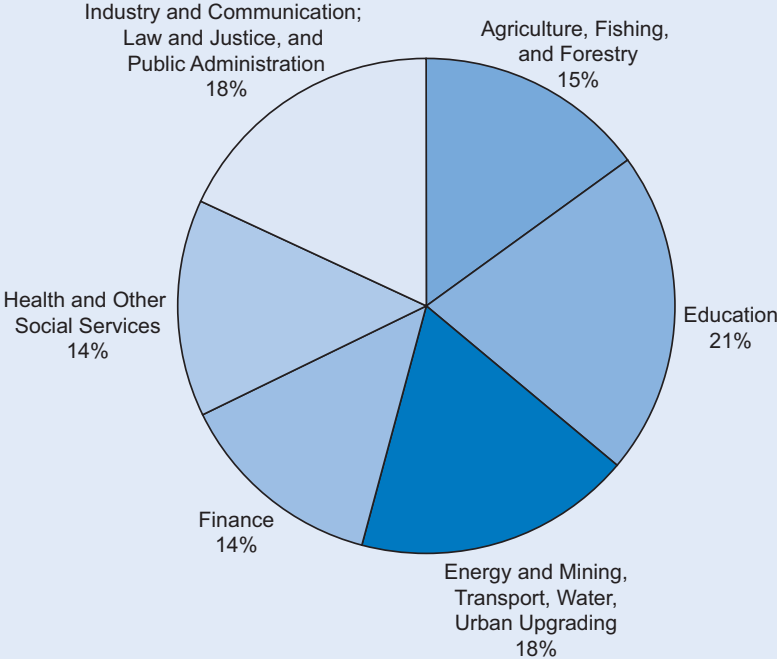
Current incentives and resources for impact evaluation are inadequate. While impact evaluation now exists as an AAA product, not many management units are devoting significant resources to this activity. Given that the evaluation program will be driven in part by knowledge gaps identified in their sector strategies, networks (including their regional affiliates) should make some of their existing resources available for impact evaluation. In addition, at the staff level, there is currently a lack of coherent incentives to initiate evaluations. The result is that most of the TTLs who engage in evaluation either do so on their own initiative or at the request of their immediate manager or the client government. These incentives should be altered to provide the proper signals to staff.

Both the capacity development and staffing of sector support teams will require the commitment of resources in terms of funding and staff time. Project and government funds are often used to fund evaluation activities once a project is underway, but the management of different evaluation programs, pre-program evaluation data collection, and training will all require other sources of funds. While the recent Spanish Trust Fund (SIEF) received by the HD Network provides one example of ways to mobilize new funds for impact evaluation support, there are some restrictions. These include limits on sectoral coverage and lack of support for capacity development originating from Bank headquarters. The fact that over 160 applications for a total of US\$40 million were received in response to the latest SIEF call for proposals also provides evidence of a clear resource constraint (SIEF will provide US\$6 million in this round).

The lessons provided by impact evaluations are a global public good. For example, as we have seen with conditional cash transfer programs, countries can learn quite a lot from each others' rigorously analyzed experiences. This public good aspect provides a powerful argument for some sort of international subsidy.

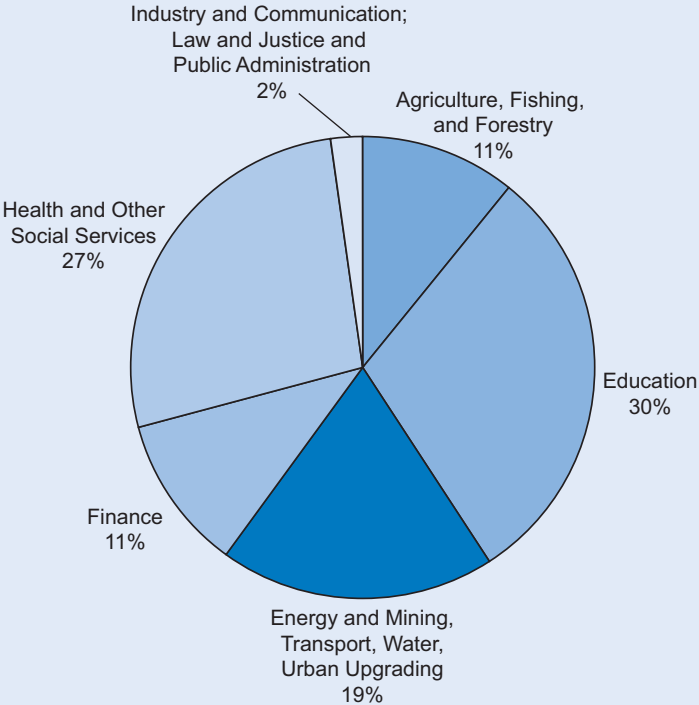
Appendix 1: Lending and Ongoing Impact Evaluations, by Sector

Figure A1.1: World Bank Lending by Bank Sector



Source: World Bank Impact Evaluation Database at: <http://www.worldbank.org/impactevaluation>.

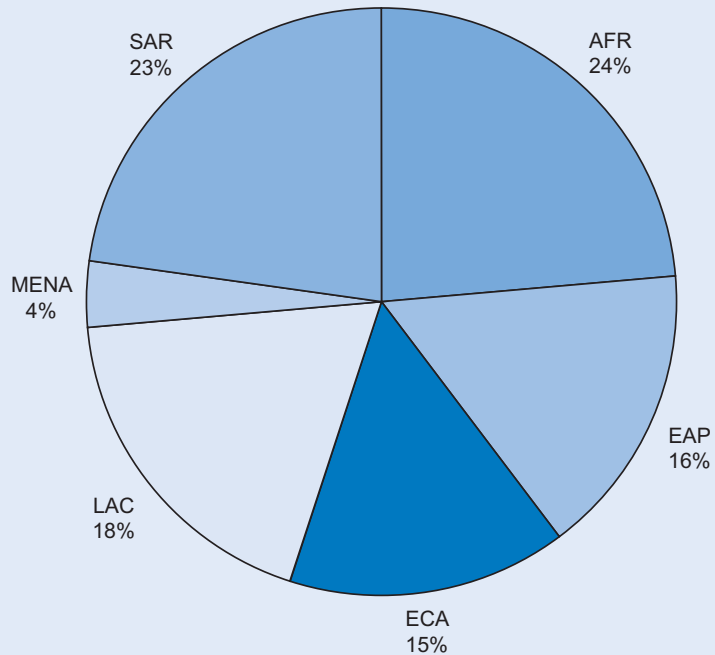
Figure A1.2: Impact Evaluations by Bank Sector



Source: World Bank Impact Evaluation Database at: <http://www.worldbank.org/impactevaluation>.

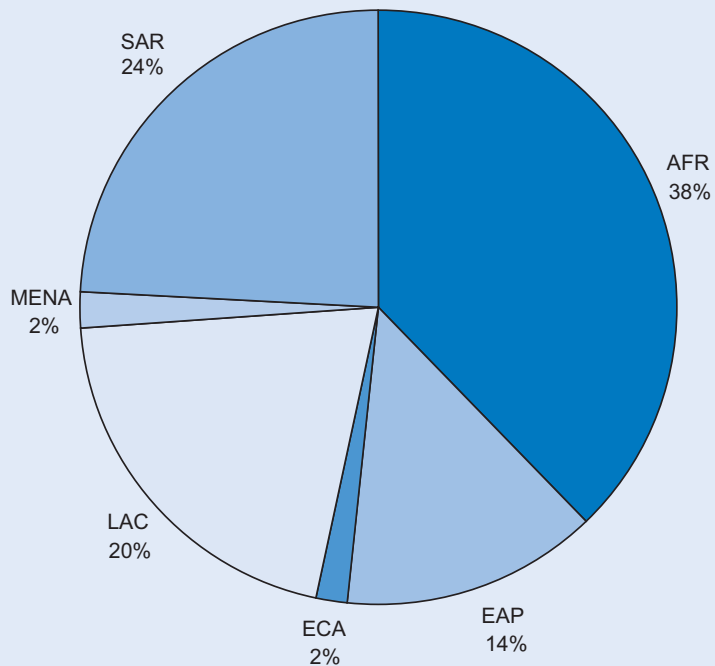
Appendix 2: Lending and Impact Evaluations, by Region

Figure A2.1: World Bank Lending by Region



Source: World Bank Impact Evaluation Database at: <http://www.worldbank.org/impacetevaluation>.

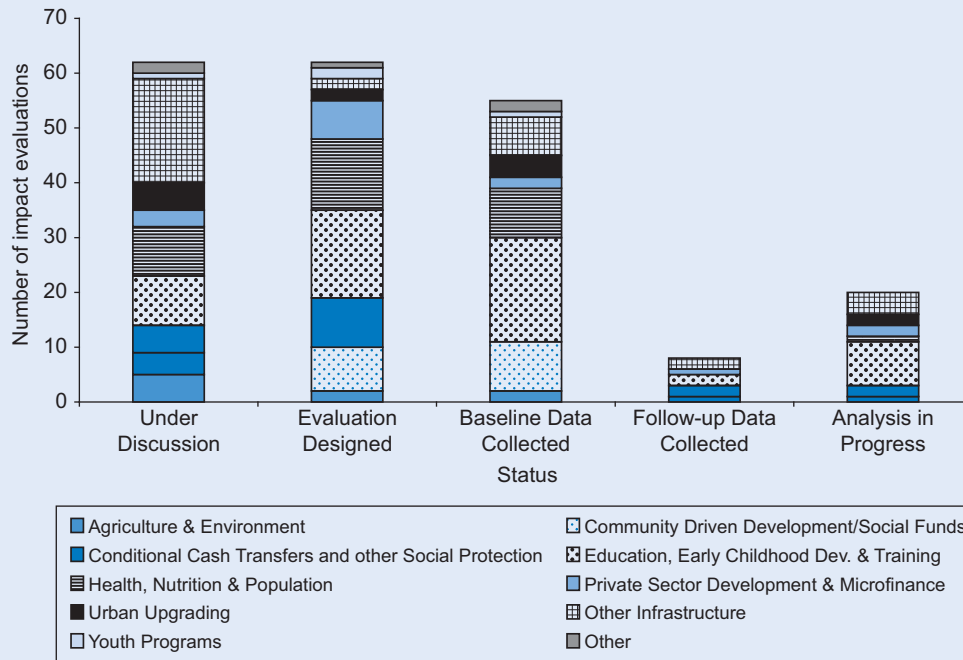
Figure A2.2: Impact Evaluations (Ongoing) by Region



Source: World Bank Impact Evaluation Database at: <http://www.worldbank.org/impacetevaluation>.

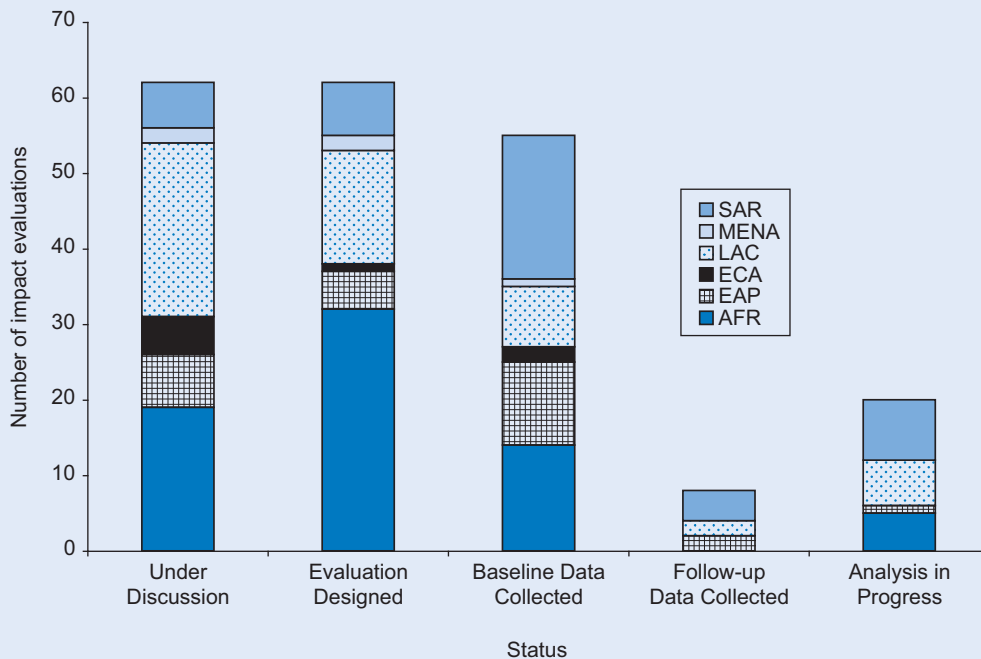
Appendix 3: Impact Evaluation Status by Region and Sector

Figure A3.1: Status of Ongoing Impact Evaluations by Category



Source: World Bank Impact Evaluation Database at: <http://www.worldbank.org/impacetevaluation>.

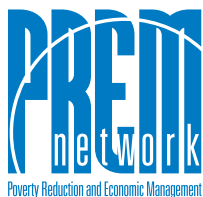
Figure A3.2: Status of Ongoing Impact Evaluations by Region



Source: World Bank Impact Evaluation Database at: <http://www.worldbank.org/impacetevaluation>.

Endnotes

1. See the recent debate on aid in the *Boston Review*: <http://bostonreview.net/ndf.html#Aid>.
2. All graphs and data cited are current as of July 2008.
3. All graphs and data cited are current as of July 2008.
4. “Ongoing” refers to the status of impact evaluations where, at minimum, the evaluation has been designed and not to evaluations that are just under discussion. See color coding in Figure 3.
5. Calculations based on information in Figure 3 and Appendix 2.
6. The only exception to this are evaluations where DECRC staff are involved, as they have a longer time horizon and are more likely to push the next TTL to continue the evaluation.
7. The *Doing Impact Evaluation* series presents general methodological guidance (e.g. Data for Impact Evaluations) as well as sectoral methods notes (e.g. Conducting Impact Evaluations in Urban Transport). The complete series can be found at <http://www.worldbank.org/impacetevaluation>.



This note series is intended to summarize good practices and key policy findings on PREM-related topics. The views expressed in the notes are those of the authors and do not necessarily reflect those of the World Bank. PREMnotes are widely distributed to Bank staff and are also available on the PREM Web site (<http://www.worldbank.org/prem>). If you are interested in writing a PREMnote, email your idea to Madjiguene Seck at mseck@worldbank.org. For additional copies of this PREMnote please contact the PREM Advisory Service at x87736. PREMnotes are edited and laid out by Grammarians, Inc.

Prepared for World Bank staff