Multilateral Development Banks:
Promoting Effectiveness and Fighting Corruption

Testimony to the Senate Committee on Foreign Relations
Ruth E. Levine, Director of Programs and Senior Fellow
Center for Global Development
March 28, 2006

Chairman Lugar, Ranking Minority Member Biden and Members of the Committee: I very much appreciate the opportunity afforded by this Committee and its staff to contribute to an essential and timely debate about how to improve the ability of development banks to support progress in low- and middle-income countries. What I will speak about today is based on work we have done over the past two years, under the auspices of the Evaluation Gap Working Group at the Center for Global Development. I would like my full testimony to be entered as part of the record, and I will summarize my major points.

To start with an obvious point: To succeed as institutions, multilateral development banks must succeed in their main business. In the near term, their main business is working with low- and middle-income governments to finance projects and programs that lead to better economic and social conditions than would have occurred without those projects or programs. The results from these investments should be valuable not only in absolute terms, but relative to the value of alternative uses of the funds. In the long-term, the banks' contributions must be still greater. They must help establish the fundamental conditions under which developing country governments can foster the welfare, productivity and prosperity of their people.

The ideal of the development enterprise, in which multilateral development banks play a special role, reflects the spirit of the well-known Chinese proverb: "Give a man a fish and you feed him for a day. Teach a man to fish and you feed him for a lifetime." In other words, development assistance is not solely or even primarily about the transfer of financial resources from wealthy countries to meet immediate survival needs of poorer nations; it is about providing the foundation of knowledge that constitutes the basis for long-term development. This is particularly true for development banks, who principally offer loan financing. Such resources are, by definition, appropriate only for investments oriented toward improved outcomes over the long-term.

In contrast, the success of development banks cannot be measured on the basis of whether they remain solvent, are on good terms with Congress and non-governmental organizations, have contented employees, or fight corruption at home and abroad. These probably are all necessary, but they are not sufficient. Success of development banks, as of other development agencies, rests on whether they can make the lives of those who are sometimes referred to as the "ultimate beneficiaries" better off, in a meaningful and sustained way.

Knowing whether or not the banks are succeeding in financing programs that directly improve people's lives is the core of accountability. Assessing this type of performance comes from impact evaluations, defined as evaluations that measure the results of an intervention in terms of changes in key variables (e.g. mortality, health status, school achievement, labor force status) that can be directly credited to the to the program itself, as distinguished from changes that are due to other factors. That is, they are evaluations that permit attribution of program-specific effects.

At the World Bank and other development banks, as in the field of development more broadly, there has been far more talk of "results" than measurement of them. Much emphasis has been placed on measuring and tracking inputs (such as commitments and disbursements) and assessing at the conclusion of a project whether the activities anticipated at the design stage have been completed. For example, were the schools built, teachers trained, blackboards procured? Did the government undertake the hoped-for changes in hospital finance or contracting with NGOs for delivery of basic health services? I can tell you from my experience within the World Bank and the InterAmerican Development Bank, and my observation from my current vantage point at the Center for Global Development, the institutional focus of monitoring performance remains very much on the "how much money did we move out the door?" and "what was financed?"

Distressingly, very little investment has been made in conducting rigorous impact evaluations that are necessary to tell us which interventions and approaches do and do not work in achieving the real goals of all that spending and program activity. Did those schools, teachers and blackboards result in more children attending and completing school than would have occurred in the absence of the project? Did the innovations in health finance and organization yield the results anticipated: less financial exposure of poor people when they became ill? Better use of essential health services, and ultimately better health? The failure to answer those basic questions in a systematic way leaves the banks, and the professionals and high-level decision makers within them, lacking feedback to foster improvement; it leaves Congressional overseers, taxpayers and others who wish to hold the institutions accountable for results with little to go on. I suspect those in this chamber are personally familiar with this deeply frustrating situation. I imagine that when you evaluate the development bank's performance, you would rather hear the results of serious impact evaluations of a sample of important programs than a recitation of amounts committed and disbursed, or number of textbooks procured.

Paradoxically, underinvestment in impact evaluation (and consequent undersupply of evidence about the relationship between specific types of investments and their effects) means that the banks have few chances of succeeding in the even more important goal of "teaching a man to fish": If we don't learn whether a program works in changing the well-being of beneficiaries, how can a government know if it is worth putting the money and effort into similar programs in the future? If we don't bother to measure the results that are direct consequences of the specific program, how can anyone make a credible case for this, or any other type of expenditure of public funds? If we don't learn whether and why our investments yield benefits, we have fundamentally failed.

Why So Little Impact Evaluation?

Several factors explain the lack of impact evaluation.

- Good impact evaluations require a degree of technical sophistication that has only emerged only recently among those who focus on international development, although it has been available to US domestic social programs since the 1970s. (Think, for example, of the landmark studies of Head Start, the tradition of excellent evaluations of within income support and job training programs in the US, and the Department of Education's laudable program evaluation initiatives today.) While many studies compare conditions before and after a project, such comparisons can be quite misleading without attention to other factors that might have also contributed to observed changes. Only by comparing observed changes among those who benefited from a project to some other control group is it possible to begin to disentangle how much of the effects can be attributed to the project or program itself. For example, when HIV prevention programs are evaluated, it is essential to measure the change in HIV incidence with both those participating in the program and a set of similar individuals or communities not exposed to program activities. Without such a comparison, before-after changes are impossible to interpret. A fall in HIV incidence in the population within the program – an observation that would likely be called "success" by the program implementers – might simply reflect declines due to non-program factors. A rise in HIV incidence among program participants – something that might be seen as "failure" – might actually reveal success if incidence is rising more slowly than in the general population. Without appropriate comparisons, we will never know. Separating out the changes due to projects from changes due to other factors is a complicated business that may require random assignment of beneficiaries or other methods. Fortunately, advances in research methods and increasing capacity around the world to conduct such impact evaluations are beginning to surmount these technical difficulties. Interestingly, we have seen how feasible this is through experiences in Mexico, where excellent design of PROGRESA, a conditional cash transfer program, illustrated the possibility and the value of introducing rigorous evaluation within the design a program as it scales up.

- Demand for the knowledge produced by impact evaluations tends to be distributed across many actors and across time. It is only at the moment of designing a new program, however, that anything can be effectively done to start an impact evaluation. At that precise moment, program designers want the benefit of prior research, yet have few incentives to invest in starting a new study. In contrast, they will get rewarded for quickly starting implementation, rather than doing the spadework to undertake a baseline study. Paradoxically, if they do not invest in a new study, the same program designers will find themselves in the exact same position four or five years later because the of the missed opportunity to learn whether or not the intervention has an impact.[1] Because information from impact evaluations is a public good, other institutions and governments that might have obtained valuable knowledge from the experience also lose when these investments in learning about impact are neglected.

- Evaluation simply is not seen as the central business of the development banks. When material and human resources are stretched, short-term operational demands will over-ride the longer-term, more strategic imperative of evaluation and learning. As one indication, resources spent to design and implement impact evaluations were not even recognized as a separate item in the World Bank's budgeting system until 2005. Most task managers at the development banks can tell very sad tales about watching their evaluation budgets disappear during negotiations with either management or borrowing governments.

- Those rare individuals within large bureaucracies who wish to undertake impact evaluations typically encounter daunting resistance. Program implementers may perceive evaluation as a threat, potentially leading to a cut-off of funding if results are not uniformly positive. At higher levels in an organization, managers who are responsive to demands by shareholders for "results, results, results" may prefer to promulgate anecdotes about success, without regard to the strength of evidence, rather than expose the genuine lessons of experience – including the occasions when results were poor. This behavior continues if it is tolerated by the institutions' constituencies and funders, including legislative bodies.

All of these reasons contribute to the situation observed today: For most types of programs, a body of scientific evidence about effectiveness is lacking. For almost all projects currently in operation or in the pipeline, virtually no credible information will be generated about program impact. The banks don't know whether they are succeeding, and they are not generating knowledge for the future. For lack of feedback, the banks continue to repeat failed approaches, and miss opportunities to expand upon successful ones. In short: We do not know how to fish. We are not learning how to fish. We have little hope of teaching others how to fish.

[1] O'Donoghue, T., Rubin, M., 1999. "Doing It Now or Later." The American Economic Review 89, 103-124.

What Can Be Done?

Fortunately, some have recognized this problem, care about solving it, and are trying hard to find a way to do so. Within the development banks, the Independent Evaluation Departments make heroic efforts to squeeze knowledge out of the experiences of projects that are conducted without baseline data, without comparison groups, frequently without any impact indicators at all. This is a difficult and often fruitless task, although it does generate the basis for improvements in the operational activities, which is very valuable. They deserve additional resources so that they can undertake more in-depth studies.

Separate from the Independent Evaluation Department (previously called Operations Evaluation Department), in the past couple of years, the World Bank has created a new initiative called the Development IMpact Evaluation (DIME) to increase the number of Bank projects with impact evaluation components, particularly in strategic areas and themes; to increase the ability of staff to design and carry out such evaluations, and to build a process of systematic learning on effective development interventions based on lessons learned from those evaluations. The regional development banks also have undertaken a limited set of impact evaluations, within either research or evaluation departments. These efforts should be recognized and provided with additional institutional resources so that they contribute to a cultural change.

But much more is required, both within and outside of the institutions. Indeed, a broader and bolder solution to the problem is needed. Three central elements are required for a lasting and genuine solution to the problem of lack of knowledge about what works.

First, we need to use good evaluation methods to get answers to important questions. This means identifying the enduring questions, a process that would be done best in true partnership between developing countries and the range of institutions that provide development finance. The World Bank has made a start by identifying a handful of thematic areas within its impact evaluation initiative. But the benefits of concentrating such studies around enduring questions across agencies and countries would be far greater.

Second, we need to use evaluation methods that yield answers. This means increasing the number of impact evaluations that use rigorous methods – such as random assignment and regression discontinuity – and applying them to a small number of programs from which the most can be learned.

Third, while the overall agenda should be developed by the “interested parties,” impact evaluations themselves need to be done independently of the major international agencies and borrowing country governments. Returning to the example of the fish, I think we’d all agree that it is better to ask an impartial judge to measure a fish with a standard ruler than to ask a fisherman to guess at the size of his catch! Independent evaluations would be more credible in the public eye, and less subject to inappropriate pressures within institutions to modify results or conclusions, or limit dissemination of unfavorable findings. We have learned about the value of independence in evaluation many times over in other fields, including medicine and social programs in our own country. The existence of an independent source of impact evaluation – geared to a longer time frame and toward learning – will avoid many of the inevitable pressures to focus on implementation alone, and restrict the communication of bad news to higher levels of management.

At the Center for Global Development, we have spent many months thinking through the reasons for the shortcoming in impact evaluation, assessing the demand for knowledge about program performance, learning about the efforts in the development banks, bilateral agencies and developing country governments related to evaluation, and developing options and recommendations. In addition to encouraging within-institution efforts to improve evaluation and create a culture of learning, we believe this may be a moment to call for a collective international initiative for leading-edge bilateral and multilateral development agencies to develop a shared impact evaluation agenda (working with interested developing countries), agree to methodological standards and, potentially, jointly fund both design and independent implementation of impact evaluations on enduring questions in international development. We have developed ideas about many of the specific functions for such an initiative, and believe that there are technically, politically, and financially feasible institutional options.

Conclusion

This Committee, and the US Congress more generally, can take a leadership role in fostering genuine, long-term success of multilateral development banks, as well as other agencies. This can be done by making three clear statements: First, that Congress values and demands the type of knowledge about program impact that comes from rigorous evaluation. Second, that Congress sees development agencies' success first and foremost in terms of whether the program experiences are yielding true *learning*, with relevant new knowledge being shared with partner governments, as a key ingredient for long-term, sustained development. Third, that Congress has an interest in exploring mechanisms to foster independent high quality impact evaluation across agencies.

I firmly believe that the development banks have tremendous untapped potential to contribute to improved outcomes over the long term – the healthier children, more productive adults, cleaner environment and lasting prosperity that we wish for all nations. From personal experience as well as recent research, I believe that they will not be firmly on the track toward fulfilling that potential until the development banks ask, and publicly answer, serious questions about the impact of their programs.