



ROOM DOCUMENT 3/B

IMPACT EVALUATION: ITS STATUS AND ITS FUTURE

This report has been prepared by the Independent Evaluation Group (IEG) of the World Bank for consideration at the 7th meeting of the DAC Network on Development Evaluation, 20 – 21 February 2008.

**7th meeting
20 – 21 February 2008**

IMPACT EVALUATION: ITS STATUS AND ITS FUTURE¹

Vinod Thomas, Director-General, Evaluation

The rising challenge of impact evaluation

Impact evaluation is now high on the development agenda. The World Bank's DIME initiative has helped initiate 120 impact studies over the past three years. A new international agency, 3IE, or the International Initiative for Impact Evaluation, is being created with the sole purpose of supporting impact evaluations. And the Network of Networks for Impact Evaluation, or NONIE, born out of a joint initiative of IEG and the DAC Evaluation Network Secretariat, in cooperation with the Evaluation Cooperation Group of the multilateral development institutions and the United Nations Evaluation Group, has made considerable progress in expanding awareness of impact evaluation.

The increased importance of this type of evaluation is linked to the focus on outcomes, as embodied in the Millennium Development Goals, and the need to demonstrate to donor nations the impact of the development projects they help finance. Impact evaluations have shown their value in several ways:

1. *Supporting or questioning a program:* Impact evaluations demonstrated the effectiveness of conditional cash transfer programs on school attendance, children's visits to clinics, and consumption poverty at a time when these programs were being dismissed by development practitioners. On the other hand, an impact evaluation of the Train and Visit extension system used in Kenyan agriculture found the system to have no impact on agricultural production, even though it had been promoted as a gold standard for extension services.
2. *Tracking cross-sectoral benefits:* Some of the strong development impacts come from interactions across sectors. An impact evaluation showed the linkage between a nutrition program in Bangladesh and secondary schooling for girls. Evaluation of the Mexican *piso firme* project that upgraded dirt floors to cement in slum housing revealed that the intervention also reduced the incidence of diarrhea, anemia, and parasitical infection in children. More recently impact evaluation has shown the education and health effects of rural electrification projects, even though the original emphasis of these projects was on infrastructure.
3. *Helping to depoliticize decision making:* Impact evaluations can help to focus the support for programs on the basis of evidence of the benefits, depoliticizing decisions to some extent. By building an evidence-based case for a program such as the Mexican *Oportunidades* program, they make it difficult for politicians to discontinue their support for the intervention after a change of government. This feature has important implications for decisions within development agencies and bureaucracies as well.

Of course, impact evaluation has important limitations and is not easily applied in the case of a good number of programs or policy situations we are involved in. For one thing, it cannot and need not be done in all cases, and is not intended to supplement or displace other evaluation tools. Furthermore, full-scale impact evaluation can be costly—between \$300,000 and \$500,000

¹ Howard White contributed to this note.

per study—though costs can be lowered substantially if existing data are used (the Inter-American Development Bank has carried out impact studies at an average cost of \$40,000). This is one reason why impact evaluation cannot reasonably be expected to be done for more than a relatively small percentage of operations. It is important, therefore, to select subjects carefully so as to build a collection of policy-relevant knowledge.

It also needs to be noted that there are cases where a full-blown, rigorous impact evaluation isn't necessary and where attribution is obvious. For example, where the installation of a pump in a village reduces the amount of time villagers spend walking to get water, a scaled-down version of an impact analysis, namely a simple before-after analysis, without a control group, would be sufficient.

Recent IEG activities in impact evaluation

IEG has conducted a small, but vigorous, program of impact evaluation-related activities over the past few years.

For example, in January 2008 IEG hosted three events related to impact evaluation. First, in our role as NONIE secretariat, we organized the second NONIE meeting, which was attended by more than 60 people. A separate briefing note has been circulated on NONIE. Second, we co-organized the conference “Making Smart Policy: Using Impact Evaluation to Inform Policy Making.” Finally, we launched IEG’s latest impact evaluation report *The Welfare Impact of Rural Electrification*, copies of which are available from IEG website².

The “Making Smart Policy” conference had over 400 participants, half from the World Bank and half from other organizations, including a substantial number from developing countries. Financial support from the Netherlands and United Kingdom made it possible to sponsor a number of developing country participants. The conference sought to move beyond the now rather staid debates on impact evaluation methods to discuss how impact studies can best influence policy. A number of conclusions emerged in this regard.

First, a study needs to be designed specifically to yield policy-relevant conclusions. Such a design often means a rather narrow study focus. But this need not preclude the possibility of generalization if the success or failure of the intervention is set in its proper context. This approach—drawing on the precepts of theory-based evaluation and realistic evaluation—means that a good impact study design also covers topics that would be covered in a process evaluation, but does so with the advantage of substantial quantitative (as well as qualitative) data resources.

The rural electrification impact study demonstrates how well-contextualized impact studies can yield policy-relevant conclusions. It also shows how rigorous impact evaluation quantifies the benefits, allowing cost-benefit analysis to be conducted. There has been a decline in the use of cost-benefit analysis: the data show the proportion of Bank projects reporting a rate of return has declined from 53 percent in 1985 to 21 percent in 2005. Such a decline does a disservice to the role economic analysis can play in informing policy; the new emphasis on quantitative impact evaluation can help reverse this trend.

² <http://ieg.worldbank.org>

The study process also matters. A good relationship with those implementing the intervention being evaluated is usually necessary to ensure access to data, though care must be taken to preserve the independence of the study. This relationship will also enhance the impact of the study findings—especially if preliminary findings are shared and discussed, so there are no surprises on either side when the final report is issued. Influence on government generally is greater when the message comes from an organization trusted by government, which usually means working in collaboration with a government agency or a local policy or research body that has government's ear.

Finally, the incentives for doing impact evaluations need to be improved. Such an incentive structure might come from linking rewards to proven impact, which would be in line with a results-based approach. However, the potential linkage between impact evaluation and the results agenda has yet to be realized.

Of course many of these lessons apply to all evaluations, not just impact evaluations. But the motivation for the current high level of attention to impact evaluation is the belief that they can provide a higher quality of evidence, and thus make a greater contribution to improving development effectiveness. The importance assigned by this conference, both to impact evaluation and to ensuring its relevance to policy, should move us closer to that goal.

Work on impact evaluation elsewhere in the World Bank

As in other agencies, the bulk of evaluation work at the World Bank takes place not in the independent evaluation department, but on the operational side. Accordingly, there are a variety of ongoing impact evaluation initiatives within the Bank. These include:

1. *The Development Impact Evaluation Initiative (DIME)*: Under leadership of the Bank's research department (DEC), this initiative promotes the embedding of (prospective) impact evaluations in project designs. Some 60 such studies are now underway, concentrated in selected areas: conditional cash transfers, education service delivery, HIV/AIDS prevention, local development, malaria control, rural roads, rural electrification, urban upgrading, and youth service programs.
2. *The IFC's Advisory Services Impact Evaluation Initiative*: Under the leadership of the Small and Medium Enterprise Department, IFC has undertaken an initiative to evaluate advisory services and technical assistance using experimental and quasi-experimental methods. Thus far, 32 such evaluations are under way covering all the business lines, from Value Addition to Firms, Access to Finance, and Infrastructure and Business Enabling Environment to Environmental and Social Sustainability.
3. *Spanish World Bank Trust Fund for Impact Evaluation*: Managed by the Human Development Network, this new three-year, €10.4 million program is dedicated to the evaluation of innovative programs to improve human development outcomes.
4. *Africa Impact Evaluation Initiative*: The World Bank's Africa Region is collaborating with internal and external partners in this program designed to mainstream impact evaluation in Bank-supported operations.

Lessons and challenges for the future

Lessons from experience suggest several issues to consider for the future of impact evaluation.

First, impact evaluation must be more than a passing fad. There have been calls for more monitoring and evaluation in the past, similar to those in today's development environment. The early 1980s, for example, saw a similar focus leading to the establishment of a department within the Bank devoted to M&E just for the agricultural sector. Much high-quality work was done, but the department closed after a few years and the quality of M&E in agricultural projects has almost certainly declined as a result, chiefly because M&E systems were over-designed with too little thought given to the use to which the data would be put. The system was not demand-driven and there was no ownership among the intended users. Both policy relevance and policy use suffered as a result. Ensuring policy relevance is therefore of great importance.

Second, impact evaluation needs to be timely. Lack of timely impact evaluation data can result in the continuation of bad programs. For example, a World Bank nutrition project in Bangladesh was scaled up before a thorough evaluation of a smaller-scale version of the project had been carried out. Once an evaluation was conducted, the project was found to have had limited impact and several design problems were identified.

Third, impact evaluations need to be increasingly factored in ex ante rather than conducted ex post. One rationale for waiting some time before doing an evaluation can be that impacts take a long time to appear. There are, however, several problems with this. First, after some years, poor institutional memory can mean that it is difficult to trace what the project actually did in different areas—a complaint, for example, in IEG's study of the Kampung Improvement Program in Indonesia for which data were collected five years after the project closed. Second, for evaluations financed by the project there is a practical constraint that the project must, by definition, still be open when the evaluation takes place, thus constraining the choice of timing. Prospective evaluation means putting in place a design now from which evaluation results will only be available in 5 to 7 years. Third, impact evaluations are cheaper and more rigorous when done prospectively because good baseline data can be collected, an appropriate control group identified, and key variables built into the project's monitoring, thereby lowering data collection costs. Finally, building in impact evaluations can encourage better project design and can facilitate real-time feedback.

Next, we need to go beyond looking under the lamp post—and to be driven by issues, not methods. Since impact evaluation is relatively new, there is a tendency to measure things that are easy to measure and that have been reviewed before, and with tools that are readily available. This is why there have been several rigorous studies of conditional cash transfer programs, job training, and nutrition interventions in a few countries. Of the over 200 studies already catalogued, more than half are in social protection, many for cash transfers. Just 2 percent are for agriculture, showing that there are clear gaps in coverage. Little progress seems to have been made on the seventh MDG dealing with the environment, but this remains hard to assess because it is also an area with poor data and little evaluation.

Finally, as already mentioned, impact evaluation design should be theory-based. It pays to examine the causal chain from inputs to outcomes. (i) IEG re-analyzed data from a nutrition project in Bangladesh, using a national survey to make a control group—using propensity score

matching since the project survey control group was weak. Very little impact was found because nutritional counseling was targeted to mothers of young children, whereas men do the food shopping in rural Bangladesh. Furthermore, the mother-in-law usually has the final say on many matters relating to childcare and diet, not the mother. So messages need to reach a broader audience and social workers need to be trained accordingly. (ii) In the Grameen Bank only the richest third of borrowers experienced a significant income gain. Our data from an impact evaluation of irrigation in Andhra Pradesh show something similar. The poorest 50 percent are more likely to use the money for consumption, whereas the better off nearly all use the loan for productive investments. Hence the better off experience an income gain, but not the poor—which won't be captured if the impact study just compares mean consumption levels. But the study also found that over half of household enterprises fail in the first two years—another reason to look at variations in the data, not just the mean.

A worthwhile goal would be that impact evaluation, together with its powerful complements of cost-benefit analysis, good M&E, and the judicious use of indicators, become an integral part of the way we work. Decisions on strategy would draw on their results or the summations of these efforts as part of the business process. We are still far from this scenario, but there is great promise—including in the fact that this event is taking place.

This is a turnaround moment— there is more financing, and greater conviction than ever before that rigorous evaluation can help confront the growing problems. We have better tools and stronger capacity than before that can be matched by greater political will to deliver. The DAC Evaluation Network has a vital role to play in exploiting these opportunities, taking forward the impact evaluation agenda in its own work, and promoting rigorous approaches within its agencies and among their partners. NONIE has provided a basis for starting such a program, and needs renewed and intensified support from the Network.