DAC Network on Development Evaluation

ROOM DOCUMENT 8

IMPACT EVALUATION – A DISCUSSION NOTE

Item IV: i

This note has been prepared by the World Bank for discussion at the meeting of the DAC Network on Development Evaluation on 2 – 3 June 2005.

3rd meeting
2 – 3 June 2005

IMPACT EVALUATION—A DISCUSSION NOTE

Impact evaluation is the systematic identification of the effects — positive or negative, intended or not — on individual households, institutions, and the environment caused by a given development activity such as a program or project. It is a type of evaluation which has received increasing attention in recent years. It is an important component of the armory of evaluation tools and approaches, albeit only one among a number. The purpose of this note is to provide an overview of impact evaluation, particularly of the more rigorous methods of impact evaluation. Lastly, its relevance to the work of the Operations Evaluation Department (OED) is considered.

Development interventions such as a project or program can be conceptualized as having a results chain — from the intervention's inputs, leading to its immediate outputs, and then to outcomes and final impacts. Evaluations can focus on all or part of this results chain.[1] Thus, evaluations which focus on efficiency, for example, are concerned with the relationship between inputs and outputs. Evaluations which include an analysis of final impacts are termed ***impact evaluations***. These focus on the effects — positive or negative, intended or not — of an intervention on individuals, institutions, and the environment. Impact can be conceptualized as the difference between what happened with the project or program and the situation if the intervention had not been made, i.e., the counterfactual situation.

Methods of Impact Evaluation

There are several methods or models of impact evaluation:

- Rapid assessment or review, conducted ex post. This method can encompass a range of approaches to endeavor to assess impact, such as participatory methods, interviews, focus groups, case studies, an analysis of beneficiaries affected by the project, and available secondary data;

- Ex-post comparison of project beneficiaries with a control group. With this method, multivariate analysis may be used to control statistically for differences in attributes between the two groups — this is one way of estimating the counterfactual situation;

- Quasi-experimental design, involving the use of matched control and project (beneficiary) groups. This method involves the use of a "non-equivalent" control group to match as closely as possible the characteristics of the project population –

[1] A range of types of M&E, including the advantages, disadvantages, cost, skills and time required to conduct each of them, is discussed in OED, *Monitoring & Evaluation: Some Tools, Methods and Approaches*, 2nd edition, World Bank, 2004.

either through propensity score matching or using a multivariate regression approach. This method often involves the use of large scale sample surveys, and sophisticated statistical analysis; and

- Randomized design. This involves the random assignment of individuals or households either as project beneficiaries, or as a control group which does not receive the service or good being provided by the project. This is also known as the experimental method, and is used in health research, for example, in areas such as evaluating the effectiveness of new drugs and medical procedures.

These four methods are discussed in more detail in the Annex.

Rigorous Impact Evaluation

The remainder of this note focuses on the last two impact evaluation methods, described here by the shorthand ***rigorous impact evaluation***.[2] The strong advantage of these two methods is that they are the most reliable for establishing causality — the relationship between a specific intervention and actual impacts — and for estimating the magnitude of impact attributable to the intervention. They are able to distinguish the impacts of the intervention from the influence of other, external factors. To do this it is important to have a good quality control group, either from randomization or using quasi-experimental methods to construct such a control. The advantage of a randomized approach is that it helps ensure that the measured impact is solely due to the intervention being evaluated.

This advantage has led to a spirited debate in recent times about the merits of rigorous impact evaluation. The Poverty Action Lab at the Massachusetts Institute of Technology (MIT) has become one vocal proponent,[3] and it was their comments which prompted a New York Times article and an editorial in The Lancet that the World Bank had rarely properly evaluated the impact of its projects.[4]

One disadvantage of randomization, however, is that it is not always feasible or appropriate. Randomization works best for discrete, homogenous interventions. It is less appropriate for complex project designs with many features unevenly implemented, or for national programs — for example, an entitlement program such as universal primary education — or for several aid modalities such as policy-based lending and budget support. There may also be ethical objections or political constraints. Thus a randomized

---

[2] See, for example, Judy Baker, *Evaluating the Poverty Impact of Projects: A Handbook for Practitioners*, World Bank, 2002, and Joseph Valadez and Michael Bamberger (eds.), *Monitoring and Evaluating Social Programs in Developing Countries: A Handbook for Policymakers, Managers, and Researchers*, World Bank, 1994.

[3] See www.povertyactionlab.com/

[4] The Lancet editorial is at: http://povertyactionlab.org/news/Lancet%20comment%20on%20WB1.pdf

approach would be inappropriate where there is knowledge about the effectiveness of the particular intervention being evaluated (such as anti-retroviral HIV/AIDS drugs) and it would be considered unethical to deny treatment to any individuals.[5]

Randomized evaluations are social experiments. Where these can not be done then a range of quasi-experimental approaches is available. In particular, the method of propensity score matching (PSM) tackles selection bias by matching project beneficiary households (or individuals) with households (or individuals) with similar characteristics, where these characteristics are identified as those affecting project participation.

These rigorous methods for impact evaluation — using randomized or quasi-experimental methods — should, if done well, also make use of other, more qualitative review methods. Such methods are particularly useful to gauge issues such as the views of key stakeholders concerning project operation and impact, and to help clarify issues such as the results chain or causality relationships of the intervention. A mixed method approach to evaluation helps achieve triangulation of evaluation findings.[6] Moreover, the analysis of impact should be embedded in a good understanding of context, ideally presenting an analysis which embraces the whole results chain.

Rigorous methods have other disadvantages. One is that they are usually very expensive — rigorous impact evaluations conducted by the World Bank have often cost from $200,000 to $900,000, depending on project size, complexity and data requirements. A second is that are often time-consuming, taking up to 2 years or more to complete; moreover, it can often take several years after a project has been in operation before its impacts start to emerge. These timing issues reduce the utility of rigorous impact evaluation when decision-makers require information quickly. Finally, rigorous methods are quite demanding in terms of the skills needed to conduct them; they require strong technical skills in social science research design, management, analysis and reporting.

Since impact evaluation is expensive it can only be done for a small percentage of projects. An important question is the extent to which lessons from any particular study can be transferred to other settings. The answer depends on the extent to which context matters. In cases where an intervention, such as immunization, varies little from setting to setting, then findings are likely to be readily transferable. But where the project outcome is found to depend strongly on the project context, then care should be taken to draw lessons for projects in similar contexts.

---

[5] See Martin Ravallion, "Comments on the paper by Duflo and Kremer", in G.K. Pitman, O. N. Feinstein, and G.K. Ingram (eds.), *Evaluating Development Effectiveness*, World Bank Series on Evaluation and Development Vol. 7, Transaction Publishers, 2005, pp. 240-7, and Howard White, "Challenges in evaluating development effectiveness", in *ibid*, pp. 33-54.

[6] Michael Quinn Patton has argued that rigorous impact evaluation, while an invaluable approach, does not constitute an evaluation "gold standard", and that other evaluative approaches also are necessary to answer particular issues. See, for example, "The debate about randomized controls as the gold standard in evaluation", video presentation to the National Institutes of Health, 21 September 2004, available at: http://videocast.nih.gov/PastEvents.asp

These features of impact evaluation make it particularly valuable as a tool for long-term studies and research. Moreover, there are public good arguments for the international development community to help fund such evaluations, thus contributing to what is becoming a growing "library" of impact evaluation findings, available to all countries and to the donor community.[7] The World Bank has been active in conducting rigorous impact evaluations. Many of these have involved the active participation of borrower governments and universities in their countries.[8] The Bank is preparing an online database of rigorous impact evaluations conducted by the Bank and others. [9] It also plans to extract lessons from the findings of such evaluations, in terms of their broader relevance to other countries, and to also make these widely available.

OED and Impact Evaluation

Role of OED

OED is an independent unit within the World Bank, and it reports directly to the Bank's Board of Executive Directors. OED assesses what works, and what does not; how a borrower plans to run and maintain a project; and the lasting contribution of the Bank to a country's overall development. The goals of evaluation are to learn from experience, to provide an objective basis for assessing the results of the Bank's work, and to provide accountability in the achievement of its objectives. It also improves Bank work by identifying and disseminating the lessons learned from experience and by framing recommendations drawn from evaluation findings.

The World Bank's operational staff conducts separate, self-evaluations of the Bank's projects as soon as the project is completed; currently, about 270 new projects are completed each year. These self-evaluations are then assessed independently by OED staff, using a rapid review approach based on standardized criteria.[10] OED's evaluation approach is termed objectives-based evaluation, focusing on whether a project's actual outcomes are likely to achieve its stated objectives. The specific criteria include:

- The ***relevance*** of the project's objectives in relation to country needs and institutional priorities;

---

[7] This argument has also been made by Esther Duflo and Michael Kramer, "Use of randomization in the evaluation of development effectiveness", in G.K. Pitman et. al, *ibid*, pp. 205-32.

[8] The Bank is preparing an inventory of rigorous impact evaluations conducted by the Bank and others in developing countries in recent years. This inventory will be made publicly available, and is currently projected to include over 110 such evaluations, of which about 50 have been conducted by the Bank itself (excluding OED, discussed below).

[9] A discussion of the World Bank's work in this area by Francois Bourguignon is available at: http://web.worldbank.org/WBSITE/EXTERNAL/NEWS/0,,contentMDK:20266730%7EmenuPK:34457%7EpagePK:34370%7EpiPK:34424%7EtheSitePK:4607,00.html

[10] These are summarized in http://www.worldbank.org/oed/oed_approach_summary.html.

- Its ***efficacy*** — the extent to which its development objectives have been (or are expected to be) achieved;

- Its ***efficiency*** — the extent to which its objectives have been (or are expected to be) achieved without using more resources than necessary;

- The ***sustainability*** of the project — the likelihood that its estimated net benefits will be maintained or exceeded over the life of the project;

- The ***institutional development impact*** — the extent to which the project improves the ability of a country to make better use of its resources; and

- The ***performance of the Bank and the borrower***, focusing on how good a job each partner has done at each stage of the project cycle.

OED's project ratings may well diverge from those of Bank operational staff. OED also subjects 25% of completed Bank projects to detailed field inspections. These project ratings are then used as a building block for the conduct of OED's higher-level evaluations — these comprise sector and thematic evaluations, country assistance evaluations, and evaluations of global programs.[11]

Rigorous Impact Evaluations Conducted by OED

As noted above, OED's mandate covers a broad range of areas of focus, and includes the evaluation of Bank work at the project, sector, thematic, country and global levels. These evaluations encompass all parts of the development results chain, including the relevance, efficacy, efficiency, sustainability, institutional development, and Bank and borrower performance dimensions of the Bank's development interventions. Measurement of the impact of the Bank's interventions is, in a sense, the bottom line. OED uses a range of evaluation types to measure impact, including rigorous impact evaluation — by conducting such rigorous impact evaluations itself, and also through use of the findings of such studies conducted by others. Examples of OED's rigorous impact evaluations are shown in Box 1.[12]

Box 1: OED Impact Evaluations in Ghana and Bangladesh

[11] Available publicly at: http://www.worldbank.org/oed/

[12] The Bangladesh evaluation will be completed shortly. The Ghana evaluation is available at: http://lnweb18.worldbank.org/oed/oeddoclib.nsf/DocUNIDViewForJavaSearch/928A136DEB347B3485256E8A0061BC8D/$file/report_28779_basic_education.pdf

OED has recently conducted two impact studies in partnership with the United Kingdom's Department for International Development (DFID). The first of these used a nationally representative household survey, as a follow-up to another 15 years earlier, which collected data on school quality and learning outcomes in Ghana. A steady increase in enrolments has been accompanied by large strides in learning outcomes. Three main factors are found to explain these improvements: economic growth; the better quality of schools, including greater availability of learning materials; and a 'multiplier effect' as more educated parents are more likely to send their children to school and to assist their learning. The World Bank is shown to have made a major contribution to these developments through a succession of projects which have constructed 8,000 classroom blocks around the country and supplied 35 million textbooks.

In Bangladesh, existing data sets were used to analyze the impact of interventions in a number of sectors on maternal and child health and nutrition. Within the health sector, immunization was shown to be a cost-effective means of averting early death. Training traditional birth attendants also saves infant lives, although the program has been abandoned in favor of promoting the use of skilled birth attendants. Providing stipends to encourage girls to attend secondary school has substantial health benefits and reduces fertility. Similar effects were found from rural electrification. However, a Bank-supported nutrition program based on growth monitoring, with nutritional counseling and supplementary feeding, had a disappointingly small impact on nutritional outcomes.

OED itself has conducted about 23 rigorous impact evaluations since 1980: 6 in agriculture; 2 in industry; 3 in education; 3 in water &sanitation; 2 in urban; 1 in transport; 2 in HNP; and 4 others.[13] In recent years, the cost of these has ranged from $400,000 to $500,000 each; they take 18-24 months from inception to completion.

Conducting such evaluations — as distinct from relying on the impact evaluations conducted by others — has several benefits to the work of OED. First, it provides OED with rigorous evaluation findings into development interventions of particular interest and priority for OED's work, such as basic education in Ghana, or Bangladesh maternal and child health.[14] Such findings provide a complement to OED's other evaluation findings, which use different evaluation types and methods. Second, in conducting rigorous impact evaluations it is often necessary to create or improve existing country data sets, and these also have a public good element, in that the data can be used by interested parties for other monitoring and evaluation work in the future. Third, OED has also used its work on rigorous impact evaluations as an additional vehicle to help build country capacities in this advanced approach.[15]

Lastly, by maintaining its long-standing program of rigorous impact evaluations, OED is able to maintain its expertise in this area. This ensures it is well placed to be able to

[13] Anju Kapoor, *Review of Impact Evaluation Methodologies Used by the Operations Evaluation Department Over Past 25 Years*, OED working paper, 2002.
[14] These evaluations are available at: http://www.worldbank.org/oed/ie/
[15] OED's work on ***evaluation capacity development*** is described at: http://www.worldbank.org/oed/ecd/

assess the quality of rigorous impact evaluations conducted by Bank Operations. Part of OED's mandate is to report annually to the Board of the Bank on the self-evaluation work of Bank Operations.[16] These annual reports will include an assessment of samples of the Bank's rigorous impact evaluations.[17] OED also intends to develop a set of good-practice standards for rigorous impact evaluation.

Finally, OED will continue its series of seminars in impact evaluation methods,[18] and sponsor training in this topic. A number of these seminars and training events will be sponsored jointly with the Bank's Poverty Analysis, Monitoring and Evaluation Thematic Group. These events will help ensure that OED staff can be discriminating consumers of rigorous impact evaluation — to help them to distinguish good from bad studies, and thus enable them to be able to interpret soundly the findings from such evaluations.

[16] For example, OED, *2003 Annual Report on Operations Evaluation*, World Bank, 2003.

[17] Unfortunately, it is easy for evaluators to do an impact evaluation badly, and to produce misleading findings and recommendations. One example is lack of a control group, so that only a before and after comparison is made. The danger with this approach is that other key variables, unrelated to the development intervention being evaluated, will have changed fortuitously over the period; failure to measure and account for these variables will often seriously distort the attributed results of the intervention

[18] Material from these presentations can be accessed on http://www.worldbank.org/oed/ie/ie_reports.html.

ANNEX

Impact Evaluation[19]

What is it?

Impact evaluation is the systematic identification of the effects—positive or negative, intended or not—on individual households, institutions, and the environment caused by a given development activity such as a program or project. Impact evaluation helps us better understand the extent to which activities reach the poor and magnitude of their effects on people's welfare. Impact evaluations can range from large scale sample surveys in which project populations and control groups are compared before and after, and possibly at several points during program intervention; to small-scale rapid assessment and participatory appraisals where estimates of impact are obtained from combining group interviews, key informants, case studies and available secondary data.

What can we use it for?

- Measuring outcomes and impacts of an activity and distinguishing these from the influence of other, external factors.
- Helping to clarify whether costs for an activity are justified.
- Informing decisions on whether to expand, modify or eliminate projects, programs or policies.
- Drawing lessons for improving the design and management of future activities.
- Comparing the effectiveness of alternative interventions.
- Strengthening accountability for results.

ADVANTAGES

- Provides estimates of the magnitude of outcomes and impacts for different demographic groups, regions or over time.
- Provides answers to some of the most central development questions—to what extent are we making a difference? What are the results on the ground? How can we do better?
- Systematic analysis and rigor can give managers and policy-makers added confidence in decision-making.

DISADVANTAGES

- Some approaches are very expensive and time-consuming, although faster and more economical approaches are also used.
- Reduced utility when decision-makers need information quickly.
- Difficulties in identifying an appropriate counter-factual.

COST

A number of World Bank impact evaluations have ranged from $200,000 - $900,000 depending on program size, complexity and data collection. Simpler and rapid impact evaluations can be conducted for significantly less that $100,000 and in some cases for as little as $10,000 - $20,000.

SKILL REQUIREMENTS

Strong technical skills in social science research design, management, analysis and reporting. Ideally, a balance of quantitative and qualitative research skills on the part of the evaluation team.

TIME REQUIRED

Can take up to 2 years or more. Rapid assessment evaluations can often be conducted in less than 6 months.

EXAMPLES OF IMPACT EVALUATION DESIGNS

Randomized evaluation designs, involving the collection of information on project and control groups at two or more points in time, provide the most rigorous statistical analysis of project impacts and the contribution of other factors. But in practice it is rarely possible to use these designs for reasons of cost, time, methodological or ethical constraints. Thus most impact evaluations use less expensive and less rigorous evaluation designs. The following table describes four approaches to impact evaluation designs in development evaluation. The first is an example of a randomized evaluation design, the second is a quasi-experimental design in which a "non-equivalent" control group is selected to match as

---

[19] Extract of OED's *M&E: Some Tools, Methods and Approaches* booklet, 2nd edition, 2004, pp.22-24.

closely as possible the characteristics of the project population; in the third example the project population is compared with a non-equivalent control group after the project has been implemented; and the fourth is a rapid assessment evaluation which combines group interviews, key informants, case studies and secondary data. Each successive model sacrifices methodological rigor, in return from which there are significant reductions in cost and time requirements.

***For more information*:**

- Baker, J. (2000). *Evaluating the Poverty Impact of Projects: A Handbook for Practitioners*. The World Bank, Washington, DC. http://www.worldbank.org/poverty/library/impact.htm
- World Bank website on impact evaluation: http://www.worldbank.org/poverty/impact/
- Roche, C. (1999) *Impact Assessment for Development Agencies: Learning to Value Change*. Oxfam, Oxford.

4 models of impact evaluation

Model	Design	Example	Indicative cost and time
1. Randomized pre-test post-test evaluation	Subjects (families, schools, communities etc) are randomly assigned to project and control groups. Questionnaires or other data collection instruments (anthropometric measures, school performance tests, etc) are applied to both groups before and after the project intervention. Additional observations may also be made during project implementation.	Water supply and sanitation or the provision of other services such as housing, community infrastructure etc where the demand exceeds supply and beneficiaries are selected by lottery. Example: Bolivia Social Fund.	1-5 years depending on time which must elapse before impacts can be observed. Cost can range from $50,000 - $1million depending on the size and complexity of the program being studied.
2. Quasi-experimental design with before and after comparisons of project and control populations.	Where randomization is not possible, a control group is selected which matches the characteristics of the project group as closely as possible. Sometimes the types of communities from which project participants were drawn will be selected. Where projects are implemented in several phases, participants selected for subsequent phases can be used as the control for the first phase project group.	These models have been applied in World Bank low-cost housing programs in El Salvador, Zambia, Senegal and the Philippines.	Cost and timing similar to Model 1.
3. Ex-post comparison of project and non-equivalent control group.	Data are collected on project beneficiaries and a non-equivalent control group is selected as for Model 2. Data are only collected after the project has been implemented. Multivariate analysis is often used to statistically control for differences in the attributes of the two groups.	Assessing the impacts of micro-credit programs in Bangladesh. Villages where microcredit programs were operating were compared with similar villages without these credit programs.	$50,000 upwards. The cost will usually be one third to one half of a comparable study using Models 1 or 2.
4. Rapid assessment ex-post impact evaluations.	Some evaluations only study groups affected by the project while others include matched control groups. Participatory methods can be used to allow groups to identify changes resulting from the project, who has benefited and who has not, and what were the project's strengths and weaknesses. Triangulation is used to compare the group information with the opinions of key informants and information available from secondary sources. Case studies on individuals or groups may be produced to provide more in-depth understanding of the processes of change.	Assessing community managed water supply projects in Indonesia.	$25,000 upwards (the Indonesia study cost $150,000). Some studies are completed in 1-2 months; others take a year or longer.