



INSTITUTIONALIZING IMPACT EVALUATION WITHIN THE FRAMEWORK OF A MONITORING AND EVALUATION SYSTEM



INSTITUTIONALIZING IMPACT EVALUATION WITHIN THE FRAMEWORK OF A MONITORING AND EVALUATION SYSTEM



Acknowledgements

This booklet was prepared by Michael Bamberger and was jointly sponsored by the World Bank's Independent Evaluation Group (IEG) and the Poverty Analysis, Monitoring and Impact Evaluation Thematic Group of the World Bank. Valuable comments on earlier drafts of this booklet were provided by a number of Bank staff, including Keith Mackay, Manuel Fernando Castro, Nidhi Khattri, Emmanuel Skoufias, Muna Meky, and Arianna Legovini. Howard White, Executive Director of the International Initiative for Impact Evaluation (3IE), also provided useful comments. The task managers of the booklet were Keith Mackay and Nidhi Khattri. The publication draws extensively on *How to Build M&E Systems to Support Better Government*.

The generous financial support of the Norwegian Agency for Development Cooperation (Norad) is gratefully acknowledged.

Knowledge Programs & Evaluation Capacity Development
Independent Evaluation Group

ISBN-13: 978-1-60244-101-9

ISBN-10: 1-60244-101-4

Copyright 2009

Independent Evaluation Group

The International Bank for Reconstruction and Development/The World Bank

700 19th Street, N.W.

Washington, DC 20431, U.S.A.

All rights reserved.

Manufactured in the United States of America.

The opinions expressed in the report do not necessarily represent the views of the World Bank or its member governments. The World Bank does not guarantee the accuracy of the data included in this publications and accepts no responsibility whatsoever for any consequence of their use.

Table of Contents

Acronyms and Abbreviations	4
1. Overview	5
2. A Brief Introduction to Impact Evaluation.....	9
3. Institutionalizing Impact Evaluation	14
4. Creating Demand for Impact Evaluation	27
5. Capacity Development for Impact Evaluation.....	30
6. Data Collection and Analysis for Impact Evaluation	34
7. Promoting the Utilization of Impact Evaluation	38
8. Conclusions.....	44
References.....	47

Tables

1. Models of IE.....	12
2. Incentives for IE—Some Carrots, Sticks, and Sermons	28
3. IE Skills and Understanding Required by Different Stakeholder Groups	31
4. Examples of the Kinds of Influence Impact Evaluations Can Have	40

Figure

1. Three Pathways for the Evolution of Institutionalized IE Systems.....	16
--	----

Boxes

1. Colombia: Moving from the Ad Hoc Commissioning of IE by the Ministry of Planning and Sector Ministries toward Integrating IE into the National M&E System (SINERGIA).....	17
2. Mexico: Moving from an Evaluation System Developed in One Sector toward a National Evaluation System (SEDESOL).....	18
3. Africa Impact Evaluation Initiative.....	19
4. Chile: Rigorous IE Introduced as Part of an Integrated Whole-of-Government M&E System	21

Acronyms and Abbreviations

AIM	Africa Impact Evaluation Initiative
DIME	Development Impact Evaluation Initiative
ECD	Evaluation capacity development
IE	Impact evaluation
M&E	Monitoring and evaluation
NGO	Non-governmental organization
OECD-DAC	Organisation for Economic Co-operation and Development, Development Assistance Committee
PROGRESA	Programa de Educación, Salud y Alimentación (Health, Nutrition, and Education Program, Mexico)
SEDESOL	Secretaría de Desarrollo Social (Secretariat for Social Development, Mexico)
SINERGIA	Sistema Nacional de Evaluación de Resultados de la Gestión Pública (National System for Evaluation of Public Sector Performance, Colombia)

1. Overview

With the growing emphasis on the assessment of aid effectiveness and the need to measure the results of development interventions, it is no longer acceptable for governments, official development agencies, and nongovernmental organizations (NGOs) to simply report how much money has been invested or what outputs have been produced. Parliaments, finance ministries, funding agencies, and the general public are demanding to know how well development interventions achieved their intended objectives, how results compared with alternative uses of these scarce resources, and how effectively they contributed to broad development objectives such as the Millennium Development Goals and the eradication of poverty.

These demands have led to an increase in the number and sophistication of impact evaluations (IEs). In the most favorable cases, the evaluations have improved the efficiency and effectiveness of ongoing programs, helped formulate future policies, strengthened budget planning and financial management, and provided a rigorous and transparent rationale for the continuation or termination of particular programs.¹ However, many IEs are selected in an ad hoc and opportunistic manner, often depending on the availability of funds or the interest of donors; and although they may have made important contributions to the program or policy being evaluated, their potential contribution to broader development strategies was often not fully achieved.

Many funding agencies and evaluation specialists have tended to assume that once a government has seen the benefits of a few

¹ Two World Bank publications have discussed the many different ways in which impact evaluations have contributed to development management (Bamberger, Mackay, and Ooi 2004 and 2005; World Bank 2008). The Development Impact Evaluation (DIME) initiative Web site includes many useful reference papers on IE methodology and studies. <http://go.worldbank.org/1F1W42VYV0>.

well-designed IEs, the process of building a systematic approach for identifying, implementing, and using evaluations at the sector and national levels will be relatively straightforward. However, many countries with decades of experience of project and program IE have made little progress toward institutionalizing the selection, design, and utilization of this type of evaluation.

This booklet describes the progress being made in the transition from individual IE studies to building a systematic approach to identifying, implementing, and using evaluations at sector and national levels whereby IE is seen as an important budgetary planning, policy formulation, management, and accountability tool. The institutionalization of IE has been achieved in a relatively small number of developing countries, mainly in Latin America, but many countries have already started or expressed interest in the process of institutionalization. This paper reviews this experience in order to draw lessons on the benefits of an institutionalized approach to IE, the conditions that favor it, the challenges limiting progress, and some of the important steps in the process of developing such an approach.

Progress toward institutionalization of IE can be assessed in terms of the following characteristics:

- *Country-led.* It is country led and managed by a central government or a major sectoral agency.
- *Strong “buy-in” from key stakeholders.* There is strong acceptance from the agencies being evaluated, parliament, policymakers, budget planners, and the public and strong support of a powerful central government agency (usually finance or planning), which manages its implementation. Although the system needs “champions,” particularly in its early stages, it is essential that it be depoliticized so that it is not seriously affected by national elections or changes in the administration.
- *Existence of legislation or strong administrative directives requiring program evaluation.*
- *Well-defined procedures and methodologies.* Processes for the selection of the programs or policies to be evaluated in a given year, and for the commissioning, conduct, dissemination, and use of the evaluations are clearly defined and widely understood. A set of standard evaluation methodologies has also been developed. Institutional

mechanisms ensure that findings and recommendations are seriously considered and implemented, with follow-up on whether recommendations have been implemented.

- *IE that is integrated into sector and national monitoring and evaluation (M&E) systems that generate much of the data used in the IE studies.* IE is one of a set of evaluation tools that respond to different information needs and management/policy guidance at different stages of project, program, and sectoral development cycles.
- *IE that is integrated into national budget formulation and development planning.* IE is recognized as an important tool for budget planning and financial management, policy formulation, and program management. Funding is guaranteed within the national budget and does not depend on the current interest of external funding agencies or particular ministries.
- *Openness and accountability.* The government is open to evaluation findings and will not suppress findings it does not like; results are put in the public domain and debated in parliament and the press; and data are made publicly available for further analysis.
- *Independence of the evaluation function.* This must be guaranteed by law but also respected in practice.
- *Evaluation capacity development.* Capacity has been developed to commission, design, conduct, manage, and use IEs. There is both a strong *demand* for the evaluations and an adequate *supply* of technical expertise and the organizational capacity to conduct evaluations and analyze data. Data-collection and analysis capacity has also been developed in planning agencies so that data sets, such as household income and expenditure or demographic and health surveys, are available for use as baseline data or for selecting control/comparison groups for the IE studies.

Institutionalization is a continuum and may work better in some sectors of government than in others. Certain studies may be used more effectively than others, and certain research methodologies will be stronger than others. Experience has also shown that there is no single path toward institutionalization of IE and no single best way in which it should be organized. An effective system must be compatible with political and public administration systems and consistent with national data-collection and analysis capacity.

Finally, although we are focusing on IE in this publication, it is important to emphasize that IE is only one of many types of evaluation that planners and policymakers use. A successful IE program can only be achieved when it is part of a broader M&E system. It would not make sense, or even be possible, to focus exclusively on IE without building up the monitoring and other data-collection systems on which IE relies. Although IEs are often the highest profile (and most expensive) evaluations, they only provide answers to certain kinds of questions; for many purposes, other kinds of evaluation will be more appropriate. Consequently, the need is to institutionalize a comprehensive M&E system that provides a menu of evaluations to cover all the information needs of managers, planners, and policymakers.

2. A Brief Introduction to Impact Evaluation

The importance of IE for development management

The primary purpose of an IE is to estimate the magnitude and distribution of changes in outcome and impact indicators among different segments of the target population and to assess the extent to which these changes can be attributed to the interventions being evaluated. In other words, is there convincing evidence that the intervention being evaluated has contributed to its intended objectives? IE can be used to assess the impacts of projects (a limited number of clearly defined and time-bound interventions, with a start and end date, and a defined funding source), programs (broader interventions that often comprise a number of projects, with a wider range of interventions and a wider geographical coverage and often without an end date), and policies (broad approaches designed to strengthen or change how government agencies operate or to introduce major new economic, fiscal, or administrative initiatives). IE methodologies were originally developed to assess the impacts of precisely defined interventions (similar to the project characteristics described above); an important challenge is how to adapt these methodologies to evaluate the multicomponent, multidonor sector- and country-level support packages that are becoming the central focus of development assistance.

A well-designed IE helps managers, planners, and policymakers avoid continued investment in programs that are not achieving their objectives, avoid eliminating programs that either are or potentially could achieve their objectives, ensure that benefits reach all sectors of the target population, ensure that programs are implemented in the most efficient and cost-effective manner and that they maximize both the quantity and the quality of the services and benefits they provide, and provide a decision tool for selecting the best way to invest scarce development resources. Without access to a good IE, there is an increased risk of reaching wrong decisions on whether programs should continue or be terminated and how resources should be allocated.

Defining IE and the different reasons for commissioning it²

The World Bank PovertyNet Web site defines IE as an evaluation that

... assesses changes in the well-being of individuals, households, communities or firms that can be attributed to a particular project, program, or policy. The central impact evaluation question is what would have happened to those receiving the intervention if they had not in fact received the program. Since we cannot observe this group both with and without the intervention, the key challenge is to develop a counterfactual—that is, a group which is as similar as possible (in observable and unobservable dimensions) to those receiving the intervention. This comparison allows for the establishment of definitive causality—attributing observed changes in welfare to the program, while removing confounding factors.³

The Organisation for Economic Co-operation and Development's Development Assistance Committee (OECD-DAC), on the other hand, defines impact as "positive and negative, primary and secondary, long-term effects produced by a development intervention, directly or indirectly, intended or unintended (OECD-DAC 2002, p. 24). OECD-DAC does not recommend a particular methodology for conducting an IE but defines impacts as long-term effects; the PovertyNet definition recommends a particular methodology (the definition of a counterfactual, based on a pre-test/post-test project/control group comparison) but does not specify a time horizon over which impacts should be measured.

IE is only one of several types of evaluation that provide information to policymakers, planners, and managers at different stages of a project or program. Although many impacts cannot be fully assessed until an intervention has been operating for several years, planners and policymakers cannot wait three or five years before receiving feedback; consequently, many IEs are combined with formative or process evaluations designed to provide preliminary findings on whether a

² For extensive coverage of the definition of IE and a review of the main quantitative analytical techniques, see the World Bank's DIME Web site: <http://go.worldbank.org/T5QGNHFO30>. For an overview of approaches used by IEG, see White (2006), and for a discussion of strategies for conducting IE (mainly at the project level) when working under budget, time, and data constraints, see Bamberger (2006a).

³ See "What Is Impact Evaluation?" at <http://go.worldbank.org/2DHMCRFFT2>.

program is on track to achieve its intended outcomes. These designs are strengthened if a program theory model is used to define key milestones and indicators at each stage of the program cycle.⁴

The most common IE designs

There is no one design that fits all IE. The best design will depend on what is being evaluated (a small project, a large program, or a nationwide policy); the purpose of the evaluation; budget, time, and data constraints; and the time horizon (is the evaluation designed to measure medium- and long-term impacts once the project is completed or to make initial estimates of potential future impacts at the time of the midterm review and the implementation completion report?). IE designs can also be classified according to whether they are commissioned at the start of the project, during implementation, or when the project is already completed; and according to their level of methodological rigor (Table 1). Designs range from randomized control trials and strong quasi-experimental designs through less robust designs where one or more of the pre-test/post-test surveys on the project or comparison groups were eliminated. Table 1 does not include nonexperimental designs without a comparison group among the IE designs, although their value for other types of program evaluation is fully recognized.⁵

Experience suggests there are relatively few situations in which the most rigorous evaluation designs can be used.⁶ It is important for policymakers and planners to keep this in mind because much of

⁴ A logic model defines the processes and stages through which an intervention is expected to achieve its intended products, outcomes, and impacts. It also identifies the external (contextual) factors, the institutional incentives and constraints, and the characteristics of target populations that affect outcomes. The model should also indicate the time horizons over which the different effects are expected to be achieved: This is important because there are often pressures to conduct the “impact” evaluation when the project/loan closes, and this is usually too early to evaluate impacts, so all that can be assessed will be outcomes or perhaps only outputs (Clark, Sartorius, and Bamberger 2004).

⁵ However, some of the examples cited in Chapter 7 to illustrate evaluation influence and utilization do use a broader definition of IE.

⁶ Although it is difficult to find statistics, based on discussions with development evaluation experts, this report estimates that randomized control trials have been used in only 1–2 percent of IEs; that strong quasi-experimental designs are used in less than 10 percent, probably not more than 25 percent include baseline surveys, and at least 50 percent and perhaps as high as 75 percent do not use any systematic baseline data.

Table 1. Models of IE

Model	Design	Example	Indicative cost and time
1. Randomized pre-test post-test evaluation	Subjects (families, schools, communities, and so forth) 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, and the like, where the demand exceeds supply and beneficiaries are selected by lottery. For example: Bolivia Social Fund—Randomized designs have been used in health and education projects and sometimes for selecting beneficiaries of conditional cash transfer programs such as PROGRESA in Mexico.	1–5 years, depending on time that must elapse before impacts can be observed. Cost can range from \$50,000 to \$1 million, 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 that matches the characteristics of the project group as closely as possible. Where possible, the project and comparison groups will be matched statistically using techniques such as propensity score matching. In other cases it may be necessary to rely on judgmental matching. 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 (pipeline design).	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. Expost comparison of project and nonequivalent control group	Data are collected on project beneficiaries and a nonequivalent 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. These designs can be strengthened when secondary data permit the reconstruction of baseline data.	Assessing the impacts of microcredit programs in Bangladesh. Villages where microcredit programs were operating were compared with similar villages without these credit programs.	\$50,000+. The cost will usually be one-third to one-half of a comparable study using Models 1 or 2.

Source: Adapted from Clark, Sarforius, and Bamberger (2004, Table 4). The category of rapid, ex post IEs has been excluded, as it does not satisfy the criteria for a quantitative IE defined early in this chapter.

the evaluation literature focuses on the small number of cases where strong designs have been used, and much less guidance is available on how to strengthen the methodological rigor of the majority of IEs that are forced by budget, time, data, or political constraints to use methodologically weaker designs.

Deciding when an IE is needed

IE may be required when policymakers or implementing agencies need to make decisions or obtain information on one or more of the following:

- To what extent and under what circumstances could a successful pilot or small-scale program be replicated on a larger scale or with different population groups?
- What has been the contribution of the intervention supported by a single donor or funding agency to a multidonor or multiagency program?
- Did the program achieve its intended effects, and was it organized in the most cost-effective way?
- What are the potential development contributions of an innovative new program or treatment?

IE may be justified when decisions have to be made about the continuation, expansion, or replication of a program and when the benefits of the evaluation (for example, money saved by making a correct decision or avoiding an incorrect one) exceed the costs of conducting the evaluation. An expensive IE that produces important improvements in program performance can be highly cost-effective; even minor improvements in a major program may result in significant savings to the government. Of course, it is important to be aware that there are many situations in which an IE is not the right choice and where another evaluation design is more appropriate.

3. Institutionalizing Impact Evaluation

Defining institutionalization of IE and why it is important

Institutionalization of IE at the sector or national level occurs under the following conditions:

- It is country-led and managed by a central government or a major sectoral agency.
- There is strong “buy-in” from key stakeholders.
- There are well-defined procedures and methodologies.
- IE is integrated into sectoral and national M&E systems that generate much of the data used in the IE studies.
- IE is integrated into national budget formulation and development planning.
- There is a focus on evaluation capacity development (ECD).

Institutionalization is a process, and at any given point it is likely to have advanced further in some areas or sectors than in others. The way in which IE is institutionalized and used will also vary from country to country, reflecting different political and administrative systems and traditions and historical factors such as strong donor support for programs and research in particular sectors. It should be pointed out that the institutionalization of IE is a special case of the more general strategies for institutionalizing an M&E system, and many of the same principles apply. As we pointed out in Chapter 1, IE can only be successfully institutionalized as part of a well-functioning M&E system.

Although the benefits of well-conducted IE are now widely recognized as useful for operational and financial management and for accountability, in practice many IE have been of limited value for policymaking and national budget planning because they were selected and funded in a somewhat ad hoc and opportunistic way that was dependent on the interests of donor agencies or individual ministries. The value of IE to policymakers and budget planners can be greatly

enhanced once it becomes part of a national or sector IE system. This requires an annual plan for selection of the government's priority programs on which important decisions have to be made concerning continuation, modification, or termination and where the evaluation framework permits the comparison of alternative interventions in terms of potential cost-effectiveness and contribution to national development goals. The examples presented in the following sections illustrate the important benefits that have been obtained in countries where significant progress has been made toward institutionalization.

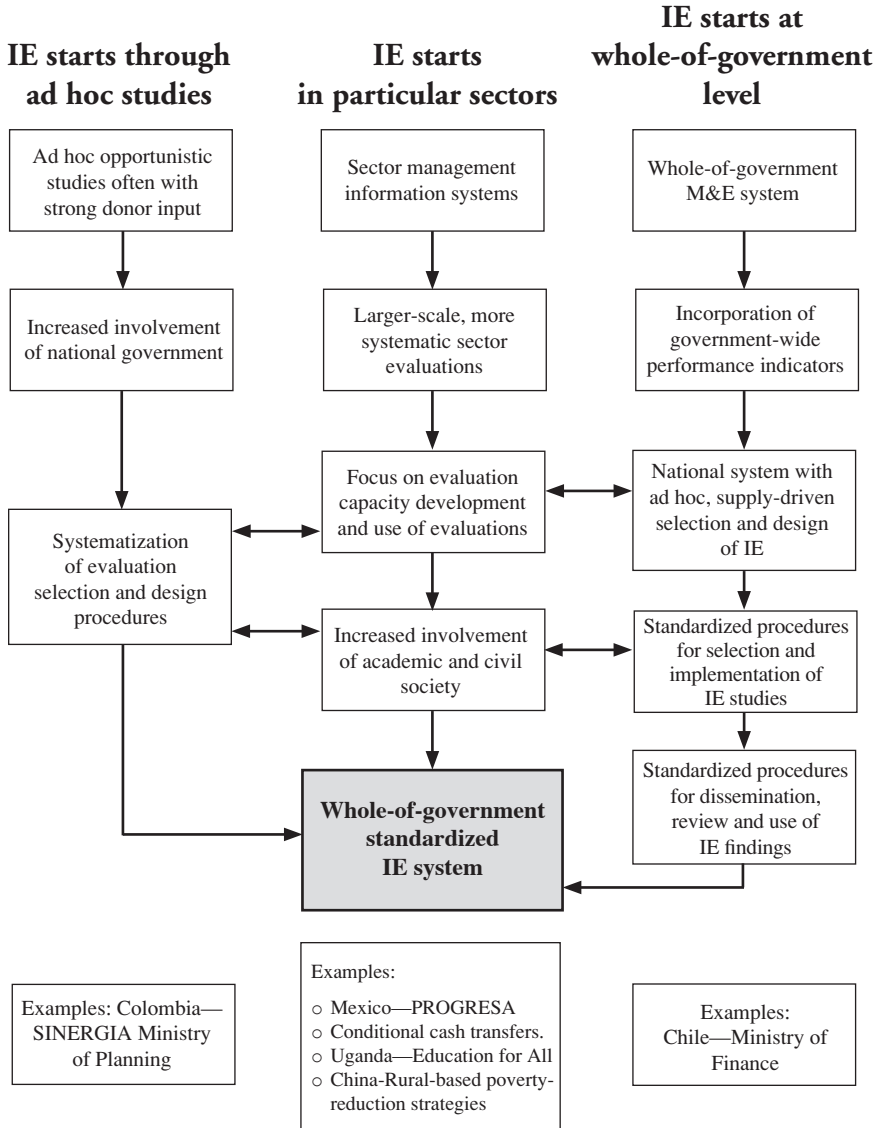
Alternative pathways to the institutionalization of IE

There is no single strategy that has always proved successful in the institutionalization of IE. Countries that have made progress in this area have built on existing evaluation experience, political and administrative traditions, and the interest and capacity of individual ministries, national evaluation champions, or donor agencies. Although some countries—particularly Chile—have pursued a national M&E strategy that has evolved over a period of more than 30 years, most countries have responded in an ad hoc manner as opportunities have presented themselves.

Figure 1 identifies three alternative pathways to the institutionalization of IE that can be observed. The first (the ad hoc or opportunistic approach) evolves from individual evaluations that were commissioned to take advantage of available funds or from the interest of a government official or a particular donor. Often evaluations were undertaken in different sectors, and the approaches were gradually systematized as experience was gained in selection criteria, effective methodologies, and how to achieve both quality and utilization. A central government agency—usually finance or planning—is either involved from the beginning or becomes involved as the focus moves toward a national system. Colombia's national M&E system, SINERGIA, is an example of this pathway (Box 1).

The second pathway is where IE expertise is developed in a priority sector supported by a dynamic government agency and with one or more champions, where there are important policy questions to be addressed and strong donor support. Once the operational and policy value of these evaluations has been demonstrated, the sectoral

Figure 1. Three Pathways for the Evolution of Institutionalized IE Systems



Box 1. Colombia: Moving from the Ad Hoc Commissioning of IE by the Ministry of Planning and Sector Ministries toward Integrating IE into the National M&E System (Secretaría de Desarrollo Social, SINERGIA)

In Colombia the Ministry of Planning is responsible for managing the National System for Evaluation of Public Sector Performance (SINERGIA). The most visible and heavily utilized component is the subsystem for monitoring progress against a total of 320 country development and presidential goals.

Although IE was initiated in 1999, these goals have evolved since 2000 to be commissioned and managed from SINGERGIA for a wide range of priority government programs. To date, SINERGIA has played a major role in the selection of the programs to be evaluated. Initially it was a somewhat ad hoc process—partly determined by the interest of international funding agencies. As the program of IE evolved, the range of methodologies was broadened and technical criteria in the selection of programs to be evaluated were formalized through policy documents (with more demand-side involvement from the agencies managing the programs being evaluated) and in how the findings are used. Most of the IEs carried out use rigorous econometric evaluation techniques. A World Bank loan is supporting the strengthening of the system with specific activities aiming to further institutionalize IE.

Source: IEG (2007, pp. 31–36).

experience becomes a catalyst for developing a national system. The evaluations of the health, nutrition, and education conditional cash transfer programs in Mexico (PROGRESA) are an example of this approach (Box 2).

Experience suggests that IE can evolve at the ministerial or sector level in one of the following ways. Sometimes IE begins as a component built into an existing ministry or sector-wide M&E system. In other cases it is part of a new M&E system being developed under a project or program loan funded by one or several donor agencies. It appears that many of these IE initiatives have failed because they tried to build a stand-alone M&E system into an individual project or program when no such system existed in other parts of the ministry or executing

Box 2. Mexico: Moving from an Evaluation System Developed in One Sector toward a National Evaluation System (SEDESOL)

In Mexico a series of rigorous evaluations of the Progresá Conditional Cash Transfer programs were conducted over a number of years. The evaluations convincingly demonstrated the effectiveness of conditional cash transfers as a way to improve the welfare (particularly education and health) of large numbers of low-income families. The evaluations are considered to have been a major contributing factor in convincing the new government that came to power in 2002 to continue these programs, which has been started by the previous administration. The evaluations also served to convince policy makers of the technical feasibility and policy value of rigorous IEs and contributed to the passing of a law by Congress in 2007 mandating the evaluation of all social programs. This law also created the National Commission for the Evaluation of Social Programs, which was assigned the responsibility for regulating the development of monitoring and evaluation functions in the social sectors. A similar continuity was achieved in Colombia, where progress is also being made toward a national M&E system (see Box 1).

Source: IEG (2007, p. 56).

agency. This has not proved an effective way to design an IE system, both because some of the data required for the IE are to be generated by the M&E system that is still in process of development, and also because the system is “time bound,” with funding ending at the closing of the project loan—which is much too early to assess impacts.

In many other cases, individual IEs are developed as stand-alone initiatives where either no M&E system exists or, if such a system does exist, it is not utilized by the IE team, which generates its own databases. Stand-alone IE can be classified into evaluations that start at the beginning of the project and collect baseline data on the project and possibly a comparison group; evaluations that start when the project is already under way—possibly even nearing completion; and those that are not commissioned until the project has ended.

The evaluations of the national Education for All program in Uganda offer a second example of the sector pathway (World Bank 2008). These evaluations have broadened interest in the existing

Box 3. Africa Impact Evaluation Initiative

The Africa Impact Evaluation Initiative (AIM) is a program of the World Bank's Africa Region using a sector approach to generating and supporting IEs. AIM currently houses umbrella thematic initiatives in education, HIV, malaria, and community-driven development, each coordinated by a team that provides organizational and technical advisory services to the participating country IE teams. AIM is currently supporting 90 experimental or quasi-experimental evaluations in 20 countries in Africa.

The stated goals of AIM are to build government capacity to implement IEs and to provide evidence on the effectiveness of different interventions, to use the findings for making decisions, and to support learning across countries within the Region. The program aims to promote the dissemination of the findings and lessons in an easily understood format through the AIM Web site, seminars, workshops, and government presentations.

For more information: <http://worldbank.org/afr/impact>

national M&E system (the National Integrated M&E System, or NIMES) and encouraged various agencies to upgrade the quality of the information they submit. The World Bank Africa Impact Evaluation Initiative (AIM) is an example of a much broader regional initiative—designed to help governments strengthen their overall M&E capability and systems through sectoral pathways—that is currently supporting some 90 experimental and quasi-experimental IEs in 20 countries in the areas of education, HIV, malaria, and community-driven development (see Box 3). Similarly, at least 40 countries in Asia, Latin America, and the Middle East are taking sectoral approaches to IE with World Bank support. A number of similar initiatives are also being promoted through recently created international collaborative organizations such as the Network of Networks for Impact Evaluation (NONIE)⁷ and the International Initiative for Impact Evaluation (3IE).⁸

⁷ <http://www.worldbank.org/ieg/nonie/index.html>.

⁸ <http://www.3ieimpact.org>.

According to Ravallion (2008), China provides a dramatic example of the large-scale and systematic institutionalization over more than a decade of IE as a policy instrument for testing and evaluating potential rural-based poverty-reduction strategies. In 1978 the Communist Party's 11th Congress adopted a more pragmatic approach whereby public action was based on demonstrable success in actual policy experiments on the ground:

A newly created research group did field work studying local experiments on the de-collectivization of farming using contracts with individual farmers. This helped convince skeptical policymakers ... of the merits of scaling up the local initiatives. The rural reforms that were then implemented nationally helped achieve probably the most dramatic reduction in the extent of poverty the world has yet seen (Ravallion 2008, p. 2).

The third pathway is where a planned and integrated series of IEs was developed from the start as one component of a whole-of-government system, managed and championed by a strong central government agency, usually the ministry of finance or planning. Chile is a good example of a national M&E system in which there are clearly defined criteria and guidelines for the selection of programs to be evaluated, their conduct and methodology, and how the findings will be used (Box 4).

Guidelines for institutionalizing IEs at the national or sector level

As discussed earlier, IEs often begin in a somewhat ad hoc and opportunistic way, taking advantage of the interest of key stakeholders and available funding opportunities. The challenge is to build on these experiences to develop capacities to select, conduct, disseminate, and use evaluations. Learning mechanisms, such as debriefings and workshops, can also be a useful way to streamline and standardize procedures at each stage of the IE process. It is helpful to develop an IE handbook for agency staff summarizing the procedures, identifying the key decision points, and presenting methodological options (DFID 2005). Following are important steps in developing an IE system.

Box 4. Chile: Rigorous IE Introduced as Part of an Integrated Whole-of-Government M&E System

The government of Chile has developed over the past 14 years a whole-of-government M&E system with the objective of improving the quality of public spending. Starting in 1994, a system of performance indicators was developed; rapid evaluations of government programs were incorporated in 1996; and in 2001 a program of rigorous IEs was incorporated. There are two clearly defined IE products. The first are rapid ex post evaluations that follow a clearly defined and rapid commissioning process, where the evaluation has to be completed in less than 6 months for consideration by the ministry of finance as part of the annual budget process. The second are more comprehensive evaluations that can take up to 18 months and cost \$88,000 on average.

The strength of the system is that it has clearly defined and cost-effective procedures for commissioning, conducting, and reporting of IEs, a clearly defined audience (the Ministry of Finance), and a clearly understood use (the preparation of the annual budget). The disadvantages are that the focus of the studies is quite narrow (only covering issues of interest to the Ministry of Finance) and the involvement and buy-in from the agencies being implemented is typically low. Some have also suggested that there may be a need to incorporate some broader and methodologically more rigorous IEs of priority government programs (similar to the PROGRESA evaluations in Mexico).

Source: IEG (2007, pp. 25–30).

- Conduct an initial diagnostic study to understand the context in which the evaluations will be conducted.⁹ This should include an assessment of evaluation resources within the organization and also those that can be drawn from other government agencies, donors, and local consultants; the nature of the programs to be evaluated; the kinds of evaluation issues to be addressed; and the likely approaches that will be required. The diagnostic study should take

⁹ For a comprehensive discussion of diagnostic studies and their importance, see IEG (2007, chapter 12).

account of local capacity, and where this is lacking, it should define what capacities are required and how they can be developed (see Chapter 5). A key consideration is whether a particular IE will be a single evaluation that probably will not be repeated or whether there will be a continuing demand for such IEs. In the former case, the concern is to define the most cost-effective way to ensure a quality evaluation. In the latter case, however, the ministry or agency must consider how to strengthen its internal capacity to commission, implement, or manage future evaluations.

- Define the appropriate option for planning, conducting and/or managing the IE, such as:
 - ▶ *Option 1:* Most IE will be conducted by the government ministry or agency itself. Technical assistance may either be obtained on an ad hoc basis or long-term support may be provided through a technical assistance agreement with either a resident adviser or regular visits from local or foreign consultants. This option may be viable for an agency with sufficient resources and technical capacity to manage the IEs, which are often complex and time consuming. Video-conferencing and the Internet have increased the possibility for just-in-time technical support from international experts.
 - ▶ *Option 2:* IE will be planned, conducted, and/or managed by a central government agency, with the ministry or sector agency only being consulted when additional technical support is required.
 - ▶ *Option 3:* IE will be managed by the sector agency but subcontracted to local or international consultants.
 - ▶ *Option 4:* The primary responsibility will rest with the donor agencies.
- Define a set of standard and transparent criteria for the selection of the IE to be commissioned each year. Criteria may include (a) the size of the program, (b) how long it has been operating (or how long since it has been evaluated), (c) problems and issues that require study, (d) whether important decisions have to be made

on continuation or modification of the program, and (e) whether the continued need for long-standing programs is required to be reassessed. Criteria must also be defined for prioritizing competing government programs that could be evaluated using IE.

- Define guidelines for the cost of an IE and how many IEs should be funded each year. Different levels of funding may be defined for rapid IE, standard IE, and in-depth evaluations of selected priority projects (see Box 4 for Chile's approach).
- Clarify who will define and manage the IE (and overall evaluation) agenda. Develop cooperation agreements with external funding agencies so that they can actively contribute without driving the agenda.
- Define where responsibility for IE is located within the organization and ensure that this unit has the necessary authority, resources, and capacity to manage the IEs. Where the ministry or sector agency has limited responsibility for the IEs, but where IEs may be required periodically over a number of years, management should decide whether to launch an ECD program (see Chapter 5).
- Conduct a stakeholder analysis to identify key stakeholders and to understand their interest in the evaluation and how they might be become involved. Ensure all stakeholders are actively involved in the identification, design, dissemination, and use of the evaluations.¹⁰
- A steering committee may be required to ensure that all stakeholders are consulted. It is important, however, to define whether the committee only has an advisory function or is also required to approve the selection of evaluations.¹¹
- Ensure that users continue to be closely involved throughout the process. This goes beyond the traditional stakeholder analysis, which only ensures that users are consulted during the planning stage. The stakeholder analysis and active user involvement is itself an effective form of demand generation.

¹⁰ Patton (2008) provides guidelines for promoting stakeholder participation.

¹¹ Requiring steering committees to approve evaluation proposals or reports can cause significant delays as well as sometimes force political compromises in the design of the evaluation. Consequently, the advantages of broadening ownership of the evaluation process must be balanced against efficiency.

- Develop strategies to ensure effective dissemination and use of the evaluations (see Chapter 7). The strategy should include a set of incentives for stakeholders to participate in the collection of information, dissemination, and use of the information and so these agencies support the evaluations of their programs. Administrative procedures should be put in place to ensure wide dissemination and to ensure that there is a formal process to review the findings and recommendations of each IE. There also needs to be a follow-up process to monitor the actions that have been taken.
- Develop an IE handbook to guide staff through all stages of the process of an IE: identifying the program or project to be evaluated, commissioning, contracting, designing, implementing, disseminating, and using the IE findings. A key step that often requires strengthening is the process of commissioning evaluations and the preparation of the terms of reference. For many organizations these terms focus mainly on the administrative procedures and do not provide sufficient guidance on the preferred methodology and required quality standards of the evaluations.
- Develop a list of prequalified consulting firms and consultants eligible to bid on requests for proposals. Prequalification considers the financial stability of the firms/individuals, their qualifications and experience, and their capacity to handle large contracts. Although prequalification has the advantage of ensuring a certain level of quality and experience and can speed up the process of commissioning studies, it is important to ensure that selection criteria do not limit the professional background of consultants to certain disciplines (such as economics), thereby excluding other consultants who could contribute to broadening the research approaches used in the evaluations. Efforts could be made to ensure that NGOs and other sectors of civil society are involved, both as stakeholders and as potential consultants to conduct at least some parts of the evaluations.

Integrating IE into sector and/or national M&E and other data-collection systems

The successful institutionalization of IE will largely depend on how well the selection, implementation, and use of IE are integrated into sector

and national M&E systems and national data-collection programs. This is critical for several reasons. First, much of the data required for an IE can be obtained most efficiently and economically from the program M&E systems. This includes information on the following:

- How program beneficiaries (households, communities, and so forth) were selected and how these criteria may have changed over time.
- How the program is being implemented (including which sectors of the target population do or do not have access to the services and benefits), how closely this conforms to the implementation plan, and whether all beneficiaries receive the same package of services and of the same quality.
- The proportion of people who drop out, the reasons for this, and how their characteristics compare with people who remained in the program.
- How program outputs compare with the original plan.

Second, IE findings that are widely disseminated and used provide an incentive for agencies to improve the quality of M&E data they collect and report, thus creating a virtuous circle. One of the factors that often affects the quality and completeness of M&E data is the fact that overworked staff may not believe that the M&E data they collect are ever used, so there is a temptation to devote less care to the reliability of the data. For example, Ministry of Education staff in Uganda reported that the wide dissemination of the evaluations of the Education for All program made them aware of the importance of carefully collected monitoring data, and it was reported that the quality of monitoring reporting improved significantly (World Bank 2008).

Third, access to monitoring data makes it possible for the IE team to provide periodic feedback to managers and policymakers of interim findings that could not be generated directly from the IE database. This increases the practical and more immediate utility of the IE study and overcomes one of the major criticisms that clients express about IE—namely that there is a delay of several years before any results are available.

Fourth, national household survey programs such as household income and expenditure surveys, demographic and health surveys,

education surveys, and agricultural surveys provide very valuable sources of secondary data for strengthening methodological rigor of IE design and analysis (for example, the use of propensity score matching to reduce sample selection bias). Some of the more rigorous IEs have used cooperative arrangements with national statistical offices to piggy-back information required for the IE onto an ongoing household survey or to use the survey sample frame to create a comparison group that closely matches the characteristics of the project population. Piggy-backing can also include adding a special module. Although piggy-backing can save money, experience shows that the required coordination can make this much more time consuming than arranging a stand-alone data-collection exercise.

4. Creating Demand for Impact Evaluation¹²

Efforts to strengthen the governance of IE and other kinds of M&E systems are often viewed as technical fixes—mainly involving better data systems and the conduct of good quality evaluations (IEG 2007). Although the creation of evaluation capacity needed to provide high-quality evaluation services and reports is important, these supply-side interventions will have little effect unless there is sufficient demand for quality IE.

Demand for IE requires that quality IEs are seen as an important policy and management tool in one or more of the following areas: (a) to assist resource-allocation decisions in the budget and planning process; (b) to help ministries in their policy formulation and analytical work; (c) to aid ongoing management and delivery of government services; and (d) to underpin accountability relationships.

Creating demand requires that there be sufficiently powerful incentives within a government to conduct IE, to create a good level of quality, and to use IE information intensively. A key factor is to have a public sector environment supportive of the use of evaluation findings as a policy and management tool. If the environment is not supportive or is even hostile to evaluations, raising awareness of the benefits of IE and the availability of evaluation expertise might not be sufficient to encourage managers to use these resources. Table 2 suggests some possible carrots (positive incentives), sticks (sanctions and threats), and sermons (positive messages for key figures) that can be used to promote the demand for IE.

The incentives are often more difficult to apply to IE than to promoting general M&E systems for several reasons. First, IEs are only conducted on selected programs and at specific points in time; consequently, incentives must be designed to encourage use of IE in appropriate circumstances but not to encourage its overuse—for example, where an IE would be premature or where similar programs

¹² This chapter adapts the discussion by IEG (2007) on how to create broad demand for M&E to the specific consideration here of how to create demand for IE.

Table 2. Incentives for IE—Some Carrots, Sticks, and Sermons

Carrots	Sticks	Sermons
<ul style="list-style-type: none"> ▪ Work to build ownership by clients and other stakeholders at all stages of the IE. ▪ Award prizes—high-level recognition of good or best practice IE. ▪ Take a collegiate approach to IE among key ministries. ▪ Provide financial and other incentives to implementing agencies to design administrative reporting systems so that they can easily be used as baseline and process monitoring data for subsequent IEs. ▪ Provide additional funding to ministries to conduct IE. ▪ Provide technical assistance and funding support (via loans) from donors to governments. ▪ Present IE findings in an easily understandable format. ▪ Assist agencies and consultants in conducting IE (help desks, manuals, and other resources). ▪ Ensure that data providers understand how the data will be used and the importance of providing accurate data to enable IEs to be conducted. ▪ Provide IE training programs for managers and staff. ▪ Identify and highlight good practice examples of IE planning and implementation. ▪ Establish a government-wide network of IE staff. ▪ Provide financial support and technical assistance for government IE from multilateral and bilateral donors. 	<ul style="list-style-type: none"> ▪ Enact laws, decrees, or regulations mandating the planning, conduct, and reporting of IE. ▪ Penalize noncompliance with agreed IE recommendations. ▪ Withhold part of funding from ministries/agencies that fail to conduct IE. ▪ Highlight adverse IE information in reports to Parliament/Congress and disseminate widely. ▪ Highlight poor-quality IE planning, data systems, performance indicators, IE techniques, and IE reporting. A supreme audit institution—a central ministry such as finance or the president’s office, and possibly internal audit—can play this role. ▪ Highlight IE findings concerning government performance to civil society. 	<ul style="list-style-type: none"> ▪ Solicit high-level statements of endorsement of IE by the president, ministers, heads of ministries, deputies, etc. ▪ Hold awareness-raising seminars and workshops to demystify IE, provide comfort about its “do-ability,” and explain what’s in it for participants. ▪ Use actual examples of influential IEs to demonstrate their utility and cost-effectiveness. ▪ Explain to service managers and staff how IE can help them deliver better services to clients. ▪ Pilot some IEs to demonstrate their usefulness. ▪ Hold conferences/seminars on good practice IE systems in particular ministries, in other countries, and so forth. ▪ Establish a network of officials working on IE—helps showcase good practice examples in ministries, demonstrates their feasibility, and helps encourage quality standards.

Source: Adapted from IEG (2007, Table 11.1).

have already been subject to an IE. Second, as flexibility is required in the choice of IE designs, it is not meaningful to propose standard guidelines and approaches, as can often be done for M&E. Finally, many IEs are contracted to consultants so agency staff involvement (and consequently their buy-in) is often more limited.

It is important to actively involve some major national universities and research institutions. In addition to tapping this source of national evaluation expertise, universities—through teaching, conferences, research, and consulting—can play a crucial role in raising awareness of the value and multiple uses of IE. Part of the broad-based support for the PROGRESA programs and their research agenda was because they made their data and analysis available to national and international researchers on the Internet. This created a demand for further research and refined and legitimized the sophisticated methodologies used in the PROGRESA evaluations. Both Mexico's PROGRESA and Colombia's *Familias en Accion* recognized the importance of dissemination (through high-profile conferences, publications, and working with the mass media) in demonstrating the value of evaluation and creating future demand.

5. Capacity Development for Impact Evaluation

The successful institutionalization of IE requires an ECD plan to strengthen the capacity of key stakeholders to fund, commission, design, conduct, disseminate, and use IE. On the supply side, this involves:

- Strengthening the supply of resource persons and agencies able to deliver high-quality and operationally relevant IEs.
- Developing the infrastructure for generating secondary data that complement or replace expensive primary data collection. This requires the periodic generation of census, survey, and program-monitoring data that can be used for constructing baseline data and information on the processes of program implementation.

Some of the skills and knowledge can be imparted during formal training programs, but many others must be developed over time through gradual changes in the way government programs and policies are formulated, implemented, assessed, and modified. Many of the most important changes will only occur when managers and staff at all levels gradually come to learn that IE can be helpful rather than threatening, that it can improve the quality of programs and projects, and that it can be introduced without introducing an excessive burden of work.

An effective capacity-building strategy must target at least five main stakeholder groups: agencies that commission, fund, and disseminate IEs; evaluation practitioners who design, implement, and analyze IEs; evaluation users; groups affected by the programs being evaluated; and public opinion. Users include government ministries and agencies that use evaluation results to help formulate policies, allocate resources, and design and implement programs and projects.

Each of the five stakeholder groups requires different sets of skills or knowledge to ensure that their interests and needs are addressed and that IEs are adequately designed, implemented, and used. Some of the broad categories of skills and knowledge described in Table 3

Table 3. IE Skills and Understanding Required by Different Stakeholder Groups

Group	Examples	Evaluation skills and understanding
Funding agencies	<ul style="list-style-type: none"> ▪ Finance ministries and departments ▪ Donor agencies ▪ Foundations ▪ International and large national NGOs 	<ul style="list-style-type: none"> ▪ Defining when IEs are required ▪ Assessing evaluation consultants ▪ Assessing proposals ▪ Estimating IE resource requirements (funds, time, professional expertise) ▪ Preparing terms of reference
Evaluation practitioners	<ul style="list-style-type: none"> ▪ Evaluation units of line ministries ▪ Evaluation departments of ministries of finance and planning ▪ Evaluation units of NGOs ▪ Evaluation consultants ▪ Universities 	<ul style="list-style-type: none"> ▪ Defining client information needs ▪ Adapting theoretically sound designs to real-world budget, time, data, and political constraints ▪ Understanding and selecting from among different evaluation designs ▪ Developing mixed-method approaches ▪ Defining program theory model ▪ Collecting and analyzing data ▪ Sampling and survey design ▪ Supervising
Evaluation users	<ul style="list-style-type: none"> ▪ Central government agencies (finance, planning, and so forth) ▪ Line ministries and executing agencies ▪ NGOs ▪ Foundations ▪ Donor agencies 	<ul style="list-style-type: none"> ▪ Assessing the validity of quantitative evaluation designs and findings ▪ Assessing the adequacy and validity of qualitative and mixed-method designs
Affected populations	<ul style="list-style-type: none"> ▪ Community organizations ▪ Farmers' organizations ▪ Trade associations and business groups ▪ Trade unions and workers organizations 	<ul style="list-style-type: none"> ▪ Helping define when evaluations are required ▪ Negotiating with evaluators and funding agencies on the content, purpose, use, and dissemination of evaluations—ensuring that beneficiaries have a voice ▪ Asking the right questions ▪ Understanding and using evaluation findings ▪ Conducting participatory evaluations
Public opinion	<ul style="list-style-type: none"> ▪ The general public ▪ The academic community ▪ Civil society 	<ul style="list-style-type: none"> ▪ Knowing how to get evaluations done ▪ Conducting participatory evaluations ▪ Making sure the right questions are asked ▪ Understanding and using evaluation findings

include understanding the purpose of IEs and how they are used; how to commission, finance, and manage IEs; how to design and implement IEs; and the dissemination and use of IEs.

The active involvement of leading national universities and research institutions is also critical for capacity development. These institutions can mobilize the leading national researchers (and also have their own networks of international consultants), and they have the resources and incentives to work on refining existing and developing new research methodologies. Through their teaching, publications, conferences, and consulting, they can also strengthen the capacity of policymakers to identify the need for evaluation and to commission, disseminate, and use findings. Universities, NGOs, and other civil society organizations can also become involved in action research.

An important but often overlooked role of ECD is to help ministries and other program and policy executing agencies design “evaluation-ready” programs and policies. Many programs generate monitoring and other forms of administrative data that could be used to complement the collection of survey data, or to provide proxy baseline data in the many cases where an evaluation started too late to have been able to conduct baseline studies. Often, however, the data are not collected or archived in a way that makes them easy to use for evaluation purposes—often because of simple things such as the lack of an identification number on each participant’s records. Closer cooperation between the program staff and the evaluators can often greatly enhance the utility of project data for the evaluation.

In other cases, slight changes in how a project is designed or implemented could strengthen the evaluation design. For example, there are many cases where a randomized control trial design could have been used, but the evaluators were not involved until it was too late.

There are a number of different formal and less-structured ways evaluation capacity can be developed, and an effective ECD program will normally involve a combination of several approaches. These include formal university or training institute programs ranging from one or more academic semesters to seminars lasting several days or weeks; workshops lasting from a half day to one week; distance learning and online programs; mentoring; on-the-job training, where evaluation skills are learned as part of a package of work skills; and as part of a community development or community empowerment program.

Identifying resources for IE capacity development

Technical assistance for IE capacity development may be available from donor agencies, either as part of a program loan or through special technical assistance programs and grants. The national and regional offices of United Nations agencies, development banks, bilateral agencies, and NGOs can also provide direct technical assistance or possibly invite national evaluation staff to participate in country or regional workshops. Developing networks with other evaluators in the region can also provide a valuable resource for exchange of experiences or technical advice. Video-conferencing now provides an efficient and cost-effective way to develop these linkages.

There are now large numbers of Web sites providing information on evaluation resources. The American Evaluation Association, for example, provides extensive linkages to national and regional evaluation associations, all of which provide their own Web sites.¹³ The Web site for the World Bank Thematic Group on Poverty Analysis, Monitoring and Impact Evaluation (DIME) provides extensive resources on IE design and analysis methods and documentation on more than 100 IEs. The IEG Web site also provides extensive resource material on M&E (including IE) as well as links to IEG project and program evaluations.

¹³ www.worldbank.org/ieg/ecd.

6. Data Collection and Analysis for Impact Evaluation¹⁴

Data for an IE can be collected in one of four ways (White 2006):

- From a survey designed and conducted for the evaluation.
- By piggy-backing an evaluation module onto an ongoing survey.
- Through a synchronized survey in which the program population is interviewed using a specially designed survey, but information on a comparison group is obtained from another survey designed for a different purpose (for example, national household survey).
- The evaluation is based exclusively on secondary data collected for a different purpose, but that includes information on the program and/or potential comparison groups.

The evaluation team should always check for the existence of secondary data or the possibility of coordination with another planned survey (piggy-backing) before deciding to plan a new (and usually expensive) survey. However, it is very important to stress the great benefits from pre-test/post-test comparisons of project and control groups in which new baseline data, specifically designed for the purposes of this evaluation, are generated. Though all the other options can produce methodologically sound and operationally useful results and are often the only available option when operating under budget, time, and data constraints, the findings are rarely as strong or useful as when customized data can be produced. Consequently, the other options should be looked on as a second best rather than as equally sound alternative designs. One of the purposes of institutionalization of IE is to ensure that baseline data can be collected.

¹⁴ For a comprehensive review of data collection methods for IE, see <http://go.worldbank.org/T5QGNHFO30>.

Organizing administrative and monitoring records in a way that will be useful for the future IE

Most programs and projects generate monitoring and other kinds of administrative data that could provide valuable information for an IE. However, there is often little coordination between program management and the evaluation team (who are often not appointed until the program has been under way for some time), so much of this information is either not collected or not organized in a way that is useful for the evaluation. When evaluation information needs are taken into consideration during program design, the following are some of the potentially useful kinds of evaluation information that could be collected through the program at almost no cost:

- Program planning and feasibility studies could often provide baseline data on both the program participants and potential comparison groups.
- The application forms of families or communities applying to participate in education, housing, microcredit, or infrastructure programs could provide baseline data on the program population and (if records are retained on unsuccessful applicants) a comparison group.
- Program monitoring data can provide information on the implementation process and in some cases on the selection criteria.¹⁵

It is always important for the evaluation team to coordinate with program management to ensure that the information is collected and archived in a way that will be accessible to the evaluation team at some future point in time. It may be necessary to request that small amounts of additional information be collected from participants (for example conditions in the communities where participants previously lived or experience with microenterprises) so that previous conditions can be compared with subsequent conditions.

¹⁵ For example, in a Vietnam rural roads project, monitoring data and project administrative records were used to understand the criteria used by local authorities for selecting the districts where the rural roads would be constructed.

Reconstructing baseline data

The ideal situation for an IE is for the evaluation to be commissioned at the start of the project or program and for baseline data to be collected on the project population and a comparison group before the treatment (such as conditional cash transfers, introduction of new teaching methods, authorization of micro-credits, and so forth) begins. Unfortunately, for a number of reasons, an IE is frequently not commissioned until the program has been operating for some time or has even ended. When a post-test evaluation design is used, it is often possible to strengthen the design and analysis by obtaining estimates of the situation before the project began. Techniques for “reconstructing” baseline data are discussed in Bamberger (2006b).¹⁶

Using mixed-method approaches to strengthen quantitative IE designs

Most IEs are based on the use of quantitative methods for data collection and analysis. These designs are based on the collection of information that can be counted and ordered numerically. The most common types of information are structured surveys (household, farm production, transport patterns, access to and use of public services, and so forth); structured observation (for example, traffic counts, people attending meetings, and the patterns of interaction among participants); anthropometric measures and measures of health and illness (intestinal infections and so on); and aptitude and behavioral tests (literacy and numeracy, physical dexterity, visual perception). Quantitative methods have a number of important strengths, including the ability to generalize from a sample to a wider population and the use of multivariate analysis to estimate the statistical significance of differences between the project and comparison group. These approaches also strengthen quality control through uniform sample selection and data-collection procedures and by extensive documentation on how the study was conducted.

¹⁶ Techniques include using monitoring data and other project documentation and identifying and using secondary data, recall, and participatory group techniques such as focus groups and participatory appraisal.

However, from another perspective, these strengths are also weaknesses, because the structured and controlled method of asking questions and recording information ignores the richness and complexity of the issues being studied, the context in which data are collected or in which the programs or phenomena being studied operate. An approach that is rapidly gaining in popularity is mixed-method research that seeks to combine the strengths of both quantitative and qualitative designs. Mixed methods recognize that an evaluation requires both *depth of understanding* of the subjects and the programs and processes being evaluated and *breadth of analysis* so that the findings and conclusions can be quantified and generalized. Mixed-method designs can potentially strengthen the validity of data collection and broaden the interpretation and understanding of the phenomena being studied. It is strongly recommended that all IE should consider using mixed-method designs, as all evaluations require an understanding of both qualitative and quantitative dimensions of the program.

7. Promoting the Utilization of Impact Evaluation

Despite the significant resources devoted to program evaluation, there is widespread concern that—even for evaluations that are methodologically sound—the utilization of evaluation findings is disappointingly limited (Bamberger, Mackay, and Ooi 2004). The barriers to evaluation utilization also affect institutionalization, and overcoming the former will contribute to the latter. There are a number of reasons why evaluation findings are underutilized and why the process of IE is not institutionalized:

- Lack of ownership
- Lack of understanding of the purpose and benefits of IE
- Bad timing
- Lack of flexibility and responsiveness to the information needs of stakeholders
- Wrong question and irrelevant findings
- Weak methodology
- Cost and number of demands on program staff
- Lack of local expertise to conduct, review, and use evaluations
- Communication problems
- Factors external to the evaluation
- Lack of a supportive organizational environment.

There are additional problems in promoting the use of IE. IE will often not produce results for several years, making it difficult to maintain the interest of politicians and policymakers, who operate with much shorter time-horizons. There is also a danger that key decisions on future program and policy directions will already have been made before the evaluation results are available. In addition, many IE designs are quite technical and difficult to understand.

The different kinds of influence and effects that an IE can have

When assessing evaluation use, it is important to define clearly what is being assessed and measured. For example, are we assessing evaluation use—how evaluation findings and recommendations are used by policymakers, managers, and others; evaluation influence—how the evaluation has influenced decisions and actions; or the consequences of the evaluation? Program or policy outcomes and impacts can also be assessed at different levels: the individual level (for example, changes in knowledge, attitudes or behavior); the designed or implementation level; the level of changes in organizational behavior; and the national-, sector-, or program-level changes in policies and planning procedures.

Program evaluations can be influential in many different ways, not all of which are intended by the evaluator or the client. Table 4 illustrates some of the different kinds of influence that IEs can have.

Ways to strengthen evaluation utilization

Understanding the political context. It is important for the evaluator to understand as fully as possible the political context of the evaluation. Who are the key stakeholders, and what are their interests in the evaluation? Who are the main critics of the program, what are their concerns, and what would they like to happen? What kinds of evidence would they find most convincing? How can each of them influence the future direction of the program? What are the main concerns of different stakeholders with respect to the methodology? Are there sensitivities concerning the choice of quantitative or qualitative methods? How important are large sample surveys to the credibility of the evaluation?

Timing of the launch and completion of the evaluation. Many well-designed evaluations fail to achieve their intended influence because they were completed either too late (the critical decisions have already been made on future funding or program directions) or too early (before the questions being addressed are on the policymakers' radar screen).

Deciding what to evaluate. A successful evaluation will focus on a limited number of critical issues and hypotheses based on a clear understanding of the clients' information needs and of how the evaluation findings

Table 4. Examples of the Kinds of Influence Impact Evaluations Can Have¹⁷

Type of influence	Examples
Providing strong empirical evidence to convince new administrations to continue major programs that were initiated by the previous administration.	The <i>Familias en Acción</i> evaluation in Colombia and the PROGRESA evaluation in Mexico
Helping national budget agencies assess future funding requirements for major government programs.	Chile
Alerting public service agencies to problems of which they were not fully aware or had not considered important.	India: citizens' report cards*
Providing objective, quantitative data that civil society can use to pressure agencies to improve services.	India: citizens' report cards*
Demonstrating the economic and social benefits of village management of services, and alerting technical agencies to the need to incorporate vulnerable groups and address community conflicts caused by limited access to services.	Indonesia: village water supply evaluation*
Providing political cover to government to take a politically sensitive decision to eliminate subsidies, thus mitigating negative consequences for influential "losers" from the policy change.	Pakistan: wheat-flour ration shop evaluation*
Developing methodologies to systematically document problems in the use of public funds that were widely suspected but that have not been possible to document.	Uganda: Education Expenditure Tracking Surveys*

will be used. What do the clients need to know and what would they simply like to know? How will the evaluation findings be used? How precise and rigorous do the findings need to be?

Basing the evaluation on a program theory (logic) model. A program theory (logic) model developed in consultation with stakeholders is a good way to identify the key questions and hypotheses the evaluation should address. It is essential to ensure that clients and stakeholders and the evaluator share a common understanding with respect to the problem the program is addressing, what its objectives are, how they will be achieved, and what criteria the clients will use in assessing success.

¹⁷ IE in these examples is defined not just in terms of methods but also in terms of beneficiary outcomes.

Creating ownership of the evaluation. One of the key determinants of evaluation utilization is the extent to which clients and stakeholders are involved throughout the evaluation process. Do clients feel that they “own” the evaluation, or do they not know what the evaluation will produce until they receive the final report? The use of formative evaluation strategies that provide constant feedback to key stakeholders on how to use the initial evaluation findings to strengthen project implementation is also an effective way to enhance the sense of ownership. Communication strategies that keep clients informed and avoid their being presented with unexpected findings (that is, a “no surprises” approach) can create a positive attitude for the evaluation and enhance utilization.

Defining the appropriate evaluation methodology. A successful evaluation must develop an approach that is methodologically adequate to address the key questions and that is also understood and accepted by the client. Many clients have strong preferences for or against particular evaluation methodologies, and one of the factors contributing to underutilization of an evaluation may be client disagreement with, or lack of understanding of, the evaluation methodology.

*Using process analysis and formative evaluation.*¹⁸ Even when the primary objective of an evaluation is to assess program outcomes and impacts, it is important to “open-up the black box” to study the process of program implementation. Process analysis (the study of how the project is actually implemented) helps understand why certain expected outcomes have or have not been achieved; why certain groups may have benefited from the program and others have not; and to assess the causes of outcomes and impacts. Process analysis also provides a framework for assessing whether a program that has not achieved its objectives is fundamentally sound and should be continued or expanded (with certain modifications) or whether the program model has not worked—at least not in the contexts where it has been tried so far. Process analysis can suggest ways to improve the performance of an

¹⁸ “An evaluation intended to furnish information for guiding program improvement is called a formative evaluation (Scriven 1991) because its purpose is to help form or shape the program to perform better” (Rossi, Lipsey, and Freeman, 2004, p. 34).

ongoing program, encouraging evaluation utilization as stakeholders can start to use these findings long before the final IE reports have been produced.

Evaluation capacity development is an essential tool to promote utilization because it not only builds skills, but it also promotes evaluation awareness (Chapter 5).

Communicating the findings of the evaluation. Many evaluations have little impact because the findings are not communicated to potential users in a way that they find useful or comprehensible. The following are some guidelines for communicating evaluation findings to enhance utilization:

- Clarify what each user wants to know and the amount of detail required. Do users want a long report with tables and charts or simply a brief overview? Do they want details on each project location or a summary of the general findings?
- Understand how different users like to receive information. In a written report? In a group meeting with a slide presentation? In an informal, personal briefing?
- Clarify if users want hard facts (statistics) or whether they prefer photos and narrative. Do they want a global overview, or do they want to understand how the program affects individual people and communities?
- Be prepared to use different communication strategies for different users.
- Pitch presentations at the right level of detail or technicality. Do not overwhelm managers with technical details, but do not insult professional audiences by implying that they could not understand the technicalities.
- Define the preferred medium for presenting the findings. A written report is not the only way to communicate findings. Other options include verbal presentations to groups, videos, photographs, meetings with program beneficiaries, and visits to program locations.
- Use the right language(s) for multilingual audiences.

Developing a follow-up action plan. Many evaluations present detailed recommendations but have little practical utility because the recommendations are never put into place—even though all groups might have expressed agreement. What is needed is an agreed action plan with specific, time-bound actions, clear definition of responsibility, and procedures for monitoring compliance.

8. Conclusions

Developing countries and aid organizations are facing increasing demands to account for the effectiveness and impacts of the resources they have invested in development interventions. This has led to an increased interest in more systematic and rigorous evaluations of the outcomes and impacts of the projects, programs, and policies they fund and implement. A number of high-profile and methodologically rigorous IEs have been conducted in countries such as Mexico and Colombia, and many other countries are conducting IEs of priority development programs and policies—usually with support from international development agencies.

Though many of these evaluations have contributed to improving the programs they have evaluated, much less progress has been made toward institutionalizing the processes of selection, design, implementation, dissemination, and use of IEs. Consequently, the benefits of many of these evaluations have been limited to the specific programs they have studied, and the evaluations have not achieved their full potential as instruments for budget planning and development policy formulation. This publication has examined some of the factors limiting the broader use of IE findings, and it has proposed guidelines for moving toward the institutionalization of IE.

Progress toward institutionalization of IE in a given country can be assessed in terms of six dimensions: (a) Are the studies country led and managed? (b) Is there strong buy-in from key stakeholders? (c) Have well-defined procedures and methodologies been developed? (d) Is IE integrated into sector and national M&E systems? (e) Is IE integrated into national budget formulation and development planning? and (f) Are there programs in place to develop evaluation capacity?

IE must be understood as being only one of the types of evaluation required at different stages of the project, program, or policy cycle. IE must be institutionalized as part of an integrated M&E system and not as a stand-alone initiative. A number of different IE designs are available, ranging from complex and rigorous experimental and quasi-experimental designs to less rigorous designs when working under budget, time, or data constraints or when the questions to be addressed do not merit the use of more rigorous and expensive designs.

Countries can move toward institutionalization of IE along one of at least three pathways. The first pathway begins with evaluations selected in an opportunistic or ad hoc manner and then gradually develops systems for selecting, implementing, and using the evaluations (for example, the SINERGIA M&E system in Colombia). The second pathway develops IE methodologies and approaches in a particular sector that lay the groundwork for a national system (for example, Mexico); the third is established from the beginning as a national system although it may be refined over a period of years or even decades (Chile). Chapter 3 identifies the actions required to institutionalize the IE. It is emphasized that IE can only be successfully institutionalized as part of an integrated M&E system and that efforts to develop a stand-alone IE system are ill advised and likely to fail.

Conducting a number of rigorous IEs in a particular country does not guarantee that ministries and agencies will automatically increase their demand for more. In fact, a concerted strategy has to be developed for creating demand for IE as well as other types of evaluation. Though it is essential to strengthen the supply of evaluation specialists and agencies able to implement evaluations, experience suggests that creating the demand for evaluations is equally if not more important. Generating demand requires a combination of incentives (carrots), sanctions (sticks), and positive messages from key figures (sermons). A key element of success is that IE be seen as an important policy and management tool in one or more of the following areas: providing guidance on resource allocation, helping ministries in their policy formulation and analytical work, aiding management and delivery of government services, and underpinning accountability.

ECD is a critical component of IE institutionalization. It is essential to target five different stakeholder groups: agencies that commission, fund, and disseminate IEs; evaluation practitioners; evaluation users; groups affected by the programs being evaluated; and public opinion. Although some groups require the capacity to design and implement IE, others need to understand when an evaluation is needed and how to commission and manage it. Still others must know how to disseminate and use the evaluation findings. An ECD strategy must give equal weight to all five groups and not, as often happens, focus mainly on the researchers and consultants who will conduct the evaluations.

Many IEs rely mainly on the generation of new survey data, but there are often extensive secondary data sources that can also be used. Although secondary data have the advantage of being much cheaper to use and can also reconstruct baseline data when the evaluation is not commissioned until late in the project or program cycle, they usually have the disadvantage of not being project specific. A valuable, but frequently ignored source of evaluation data is the monitoring and administrative records of the program or agency being evaluated. The value of these data sources for the evaluation can be greatly enhanced if the evaluators are able to coordinate with program management to ensure monitoring and other data are collected and organized in the format required for the evaluation. Where possible, the evaluation should use a mixed-method design combining quantitative and qualitative data-collection and analysis methods. This enables the evaluation to combine the breadth and generalizability of quantitative methods with the depth provided by qualitative methods.

Many well-designed and potentially valuable evaluations (including IEs) are underutilized for a number of reasons, including lack of ownership by stakeholders, bad timing, failure to address client information needs, lack of follow-up on agreed actions, and poor communication and dissemination. An aggressive strategy to promote utilization is an essential component of IE institutionalization.

References

- Bamberger, M. 2008a. "Reconstructing Baseline Data for Program Impact Evaluation and Results-Based Management." Draft report, World Bank Institute, Washington, DC.
- . 2008b. "Enhancing the Utilization of Evaluation for Evidence-Based Policy Making." In *Bridging the Gap: The Role of Monitoring and Evaluation in Evidence-Based Policy Making*, ed. M. Segone. Geneva: UNICEF.
- . 2006a. *Conducting Quality Impact Evaluations under Budget, Time, and Data Constraints*. Washington, DC: World Bank.
- . 2006b. "Evaluation Capacity Building." In *Creating and Developing Evaluation Organizations: Lessons Learned from Africa, Americas, Asia, Australasia and Europe*, ed. M. Segone. Geneva: UNICEF.
- Bamberger, M., K. Mackay, and E. Ooi. 2005. *Influential Evaluations: Detailed Case Studies*. Independent Evaluation Group. Washington, DC: World Bank.
- . 2004. *Influential Evaluations: Evaluations that Improved Performance and Impacts of Development Programs*. Washington, DC: World Bank.
- Bamberger, M., J. Rugh, and L. Mabry. 2006. *Real World Evaluation: Working under Budget, Time, Data and Political Constraints*. Thousand Oaks, CA: Sage Publications.
- Clark, Mari, Rolf Sartorius, and Michael Bamberger. 2004. *Monitoring and Evaluation: Some Tools, Methods and Approaches*. Independent Evaluation Group Evaluation Capacity Development. Washington, DC: World Bank.
- DFID (Department for International Development, UK). 2005. Guidance on Evaluation and Review for DFID Staff. <http://www.dfid.gov.uk/aboutdfid/performance/files/guidance-evaluation.pdf>.
- IEG (Independent Evaluation Group). 2007. *How to Build M&E Systems to Support Better Government*. Washington, DC: World Bank.
- OECD-DAC (Organisation of Economic Co-operation and Development, Development Advisory Committee). 2002. *Glossary of Key Terms in Evaluation and Results Based Management*. Paris: OECD.
- Patton, M.Q. 2008. *Utilization-Focused Evaluation* (4th ed.). Thousand Oaks, CA: Sage Publications.

- Picciotto, R. 2002. "International Trends and Development Evaluation: The Need for Ideas." *American Journal of Evaluation* 24 (2): 227–34.
- Ravallion, M. 2008. Evaluation in the Service of Development. Policy Research Working Paper No. 4547, World Bank, Washington, DC.
- Rossi, P., M. Lipsey, and H. Freeman. 2004. *Evaluation: a Systematic Approach* (7th ed.). Thousand Oaks, CA: Sage Publications.
- Scriven, M. 1991. *Evaluation Thesaurus* (4th ed.). Newbury Park, CA: Sage Publications.
- White, H. 2006. "Impact Evaluation: The Experience of the Independent Evaluation Group of the World Bank." Working Paper No. 38268, World Bank, Washington, DC.
- Wholey, J.S., J. Scanlon, H. Duffy, J. Fukumoto, and J. Vogt. 1970. *Federal Evaluation Policy: Analyzing the Effects of Public Programs*. Washington, DC: Urban Institute.
- World Bank. 2008. "Using Impact Evaluations for Policymaking." PREM Thematic Group for Poverty Analysis, Monitoring, and Impact Evaluation Doing Impact Evaluation Series Paper No. 13, World Bank, Washington, DC.

Additional Resources on Monitoring and Evaluation

- World Bank Independent Evaluation Group
<http://www.worldbank.org/ieg>
- World Bank Independent Evaluation Group—Impact Evaluation
<http://www.worldbank.org/ieg/ie>
- World Bank—Impact Evaluation
<http://www.worldbank.org/impactevaluation>
- Building government monitoring and evaluation systems
<http://www.worldbank.org/ieg/eccd>
- Monitoring and Evaluation News
<http://www.mande.co.uk>



THE WORLD BANK GROUP

1818 H Street, N.W.

Washington, D.C. 20433, U.S.A.

Telephone: 202-477-1234

Facsimile: 202-477-6391

Internet: www.worldbank.org

Independent Evaluation Group

Knowledge Programs and Evaluation Capacity Development (IEGKE)

E-mail: ieg@worldbank.org

Telephone: 202-458-4497

Facsimile: 202-522-3125

