

American Journal of Evaluation

<http://aje.sagepub.com>

The Evaluation of International Development Programs: A View from the Front

Michael Bamberger

American Journal of Evaluation 2000; 21; 95

DOI: 10.1177/109821400002100108

The online version of this article can be found at:
<http://aje.sagepub.com/cgi/content/abstract/21/1/95>

Published by:



<http://www.sagepublications.com>

On behalf of:

American Evaluation Association

Additional services and information for *American Journal of Evaluation* can be found at:

Email Alerts: <http://aje.sagepub.com/cgi/alerts>

Subscriptions: <http://aje.sagepub.com/subscriptions>

Reprints: <http://www.sagepub.com/journalsReprints.nav>

Permissions: <http://www.sagepub.com/journalsPermissions.nav>

Citations <http://aje.sagepub.com/cgi/content/refs/21/1/95>

The Evaluation of International Development Programs: A View from the Front

MICHAEL BAMBERGER

ABSTRACT

International program evaluation is a booming business, with important and challenging evaluations of development programs being conducted in almost every country in the developing world. However, many U.S. domestic evaluation practitioners are not yet familiar with this field. Evaluators of international development programs normally must operate in a very different environment than one would expect to find when evaluating U.S. programs. These differences are discussed and a number of promising developments and methodological approaches are described here. I conclude by suggesting a number of areas in which a closer exchange of experiences between U.S. evaluation practitioners and their colleagues from developing countries could be mutually beneficial.

INTRODUCTION

Over the past few years the American Evaluation Association (AEA) has sought to broaden its membership by attracting more evaluators from Europe, Australasia, and developing countries. This effort received a strong impetus from the international focus of the AEA conference in Vancouver in 1995. Since then the number of conference sessions on the practice of evaluation in developing countries has increased steadily. For example, at the Chicago conference in November 1998, a three-day pre-session workshop on "Program Evaluation in Developing Countries" drew some 30 participants from 12 countries. The purpose of this article is to stimulate interest in the field of international program evaluation. Toward this end, I review the current approaches in the evaluation of donor-funded projects and programs in developing countries and describe the social and political contexts within which these evaluations are conducted. Although a systematic comparison with U.S. domestic evaluation practice is not made, the reader should be able to appreciate the areas of similarity and difference between international evaluation and U.S. domestic evaluation practice.

Michael Bamberger • Gender and Development Group, The World Bank, 1818 H Street, N.W., Washington, DC 20433; Tel.: (202) 473-6438; Fax: (202) 522-3237; E-mail: jbamberger@worldbank.org.

American Journal of Evaluation, Vol. 21, No. 1, 2000, pp. 95–102. All rights of reproduction in any form reserved. ISSN: 1098-2140 Copyright © 2000 by American Evaluation Association.

In the first section, I describe the context within which evaluations of international development programs are conducted and identify a set of common issues and challenges faced by many different agencies and evaluation practitioners. Although none of these issues are unique to international development, when taken together, they mean that evaluators of international development programs normally must operate in a very different environment than they would expect to find when evaluating U.S. programs. In the second section, I identify a number of promising developments and approaches in international development evaluation, and discuss how they are affecting the quality and utility of international evaluations.

THE CONTEXT FOR EVALUATIONS OF INTERNATIONAL DEVELOPMENT PROGRAMS

For the purposes of this discussion, development programs are defined as all social and economic programs in developing countries funded by multilateral and bilateral development agencies or by international non-government organizations (NGOs). These funding agencies normally require that their programs be evaluated. The evaluations may be conducted by the funding agency, the national agency administering the program being evaluated, or international or national consultants. The evaluation activities may be limited to specific projects or programs, or they may seek to develop national evaluation capacity to replicate the methods.

The term *evaluation* is used differently by different agencies and authors. Some distinguish between *monitoring* activities, which are conducted during project or program implementation to assess the efficiency and effectiveness with which inputs are used to achieve intended outputs, and *evaluation* activities, which assess the extent to which projects or programs have achieved their intended objectives and have produced their intended changes and benefits in the target populations. In other cases the term *evaluation* is used more broadly to cover both of these functions.

Despite the diversity of the field of international development, a number of common evaluation approaches can be identified, and a number of common issues and challenges face most of the agencies and evaluation practitioners working in this field.

Trends and Challenges in Evaluating International Development Programs

1. In many developing countries, donor agencies continue to be the main sponsors of evaluation. Consequently, donors' information priorities and evaluation methodologies continue to exert considerable influence on how evaluation is practiced and used.
2. The majority of evaluation resources continue to be devoted to monitoring project implementation and to the production of immediate outputs. In contrast, very little attention and few resources are devoted to the systematic assessment of whether policies, programs, or projects achieve their intended impacts and benefits for the target populations. This is equally true for official multilateral and bilateral development agencies, governments, and NGOs. One consequence is that much of the discussion about the impacts of development policies, programs, and projects is based on limited evidence. In some cases generalizations are made from local-level, and not necessarily representative, case studies. In others, observed changes over time are imputed to be caused by certain policies or programs, usually with very little comparative data from control populations not affected by the programs.

3. There is increased concern about giving voice to the poor and other groups affected by development programs and policies. This had led to the widespread use of participatory evaluation methods. Many of these approaches use qualitative evaluation methods, thereby resulting in active debates concerning the merits of different research paradigms and how qualitative methods can be integrated with conventional quantitative methods.
4. There is a growing interest in *thematic evaluations* that examine the impacts of development strategies on particular groups such as women or indigenous populations, or on development issues such as the environment. Many evaluators believe that special evaluation methodologies must be developed to ensure that the perspectives and reality of these populations or groups are understood and fully reflected in policy and program design.
5. There is an increasing interest in assessing the social and economic impacts of economic development policies, and particularly broad economic reform policies such as structural adjustment. Many of these efforts are constrained by the lack of adequate databases. This has led to efforts to identify and use longitudinal data sets, a number of which have become available in recent years.
6. Efforts continue to develop national evaluation capacity. While progress has been made, the results of many of these efforts have been disappointing, due in part to the pressures on international development agencies to respond to the information needs of their own parliaments and funding agencies, rather than to design systems responding to the needs of the host country (Valadez & Bamberger, 1994). There is a need for greater coordination between and/or among international agencies. Often a number of different agencies are supporting projects in the same sector, with each requiring similar monitoring and evaluation data. However, different donors require data in different formats, which greatly complicates the work of the often small national evaluation agencies. If the information needs of the different development agencies were consolidated and their financial and technical support coordinated, the demands on national agencies could be reduced and the quality of their work improved.

PROMISING APPROACHES AND FUTURE CHALLENGES

Increased Availability of Longitudinal Databases

One of the greatest difficulties facing impact evaluation research has been the lack of longitudinal databases that can be used to compare the situation before and after an intervention. This has been particularly problematic for the assessment of the impacts of macro-economic policies because the cost of generating national data sets is extremely high. Over the past few years a number of major investments in the creation of national longitudinal data sets have begun to bear fruit. One of the largest and most notable examples is the Living Standards Measurement Study (LSMS), which has developed a standard set of socioeconomic survey instruments that have now generated data sets on living standards in over 40 countries. Many of these studies have been repeated so that longitudinal data sets are becoming available and have been used to assess the household-level impacts of economic reform, particularly in Africa. In the LSMS, a set of standard questionnaires and specialized modules covering a broad range of indicators of living standards (health, education, income,

employment, access to basic services, etc.) were developed by the World Bank in cooperation with other UN agencies. These have now been applied in several dozen countries by national statistical agencies (Deaton, 1997). In some of these countries (e.g., the Ivory Coast, Ghana, and Peru) panels of households were interviewed at two or more points in time, while in others comparable samples were interviewed. These data sets are now available to researchers inside and outside the World Bank. LSMS data have been used to conduct a large number of descriptive studies on changes in poverty and living standards. They are now starting to be used in impact evaluation. While panel surveys offer a number of advantages for evaluating household-level changes and impacts (as well as problems relating to the effects of drop-outs on sample representativeness), it is possible to use both panel and comparable sample longitudinal data in conducting impact evaluations. The LSMS studies continue to be one of the best data sets for national and comparative cross-country impact evaluations of macroeconomic policies.

One of the first applications has been to assess the impacts of economic and administrative reform programs (often referred to as structural adjustment). The LSMS data have been used to compare living standards before and after the introduction of the reform programs in a number of countries (e.g., Ghana and the Ivory Coast). The LSMS and other national data sets also have been used in comparative worldwide studies to assess the impacts of economic growth on poverty (Deininger & Squire, 1995; Demery, Chen, & Vishwanath, 1995; Ravallion & Chen, 1996; The World Bank, 1996). These data sets make it possible, in most cases for the first time, to conduct a systematic and rigorous assessment of the impacts of macro-economic policies.

The use of these data sets is not limited to the evaluation of macro-economic policies. Examples of other potential applications include the assessment of trends in the incidence and distribution of poverty, and cross-country comparisons to test the association between female-headed households and poverty.

There are also several sectors, such as agriculture and education, in which longitudinal data sets are potentially available, but where they have been little utilized for evaluation purposes. In agriculture, extensive data on production, prices, and employment are often collected during project design and the information is often updated periodically. An initiative is now underway among the U.S. Department of Agriculture (USDA), Food and Agriculture Organization (FAO), and the World Bank to strengthen these databases for evaluation purposes in three pilot studies in African countries. In education, large-scale data sets are often generated on school attendance, continuation, repetition, and performance for purposes of administering subsidies, scholarships, and other kinds of targeting benefits. For example, in Bangladesh data have been collected on all of the secondary school population as part of the administration of the Female Secondary Schools Scholarship program. The information has been updated annually over the past three years and can be compared with national educational data collected several years earlier.

Participatory Evaluation

In recent years a wide range of participatory methods have been developed to give voice to the intended project beneficiaries (or affected groups) in the identification, design, and management of projects. These methods include rapid rural appraisal (RRA), beneficiary assessment, stakeholder analysis, and a wide range of social assessment methods. The *Participation Tool Kit* (Narayan, 1997) presents a wide range of participatory evaluation

methods and documents how they have been used in different sectors and regions—mainly to consult with beneficiaries during project design. However, the same methods can be used equally well to obtain beneficiaries' perspectives on the impacts of the projects and policies once they become operational.

The 1998 social assessment study of the World Bank-financed Angola Social Assistance Fund (FAS) illustrates the use of participatory methods (Bamberger, 1999). The participatory assessment studied the effectiveness of all community stakeholders' (including women) involvement in the selection and design of projects (water supply, schools, health posts), and evaluated the communities' perception of the projects' impacts. A similar study was completed for the Eritrean Community Development Fund in late 1998 (Bamberger, 1999).

One promising area for future research would involve promoting greater cross-fertilization between the growing empowerment evaluation work in the U.S. (Fetterman, Kaftarian, & Wandersman, 1996) and the international field. The areas in which empowerment evaluation have been applied in the U.S. (e.g., HIV prevention, education, strengthening community involvement in local government, community health initiatives, strengthening women's organizations) are all of major concern in the international field and with appropriate modifications U.S. evaluation methods could easily be applied. The related field of cluster evaluation (Sanders, 1997) also provides a useful way to involve local stakeholders in the evaluation and control of projects which are implemented in multiple sites.

Integrating Quantitative and Qualitative Evaluation Methods

Many economists and quantitative researchers who are interested in the potential insights that can be obtained from qualitative methods are, at the same time, concerned with the apparent lack of statistical rigor in the selection of households and communities. Some participatory researchers challenge the validity or appropriateness of quantitative methods, arguing that the concerns for statistical rigor are inappropriate and that the qualitative research paradigm methods are more suited for understanding the concerns of the poor (or of women or indigenous peoples). Nevertheless, many NGOs take concerns about quantitative rigor seriously because they want their study findings to be taken into consideration by national and international policy makers, most of whom come from a quantitative, and often economic, tradition. A challenge for these evaluation practitioners will be to develop guidelines for the integration of quantitative and qualitative methods so that each approach can draw on the strengths of the other.

One of the promising attempts to reconcile quantitative and qualitative evaluation methods is the Structural Adjustment Participatory Research Initiative (SAPRI). Countries undergoing economic reform and structural adjustment programs are participating in a collaborative research project to integrate participatory and conventional economic survey methods in the evaluation of the impacts of the economic reform programs. Bangladesh, Ecuador, Ghana, Hungary, Mali, Uganda, and Zimbabwe are currently participating and several other countries might join. (See <http://www.worldbank.org/research/sapri/index.htm>.) In each country a wide range of NGOs is cooperating with the World Bank and the national governments in the design and implementation of methodologies combining quantitative and qualitative methods to evaluate the social and economic impacts of the structural adjustment policies. The field work is conducted in each country by an independent research institute, so it will not be exclusively a World Bank study or an NGO study.

While the practical utility of this collaborative approach has been demonstrated, a

number of difficulties have been identified. Many of the national NGOs and women's organizations that would like to participate in the studies are not familiar with the kinds of policies and programs being evaluated or with the research methods used by World Bank economists. Consequently, the time allowed for consultations has often been insufficient for the orientation and involvement of the local organizations. Another problem is that many of the local organizations do not fully understand the need for more rigorous quantitative methods to ensure that observed problems or changes can be attributed to the programs. The conclusion is that, while this kind of collaboration has proved very constructive, the time required and the level of resources needed for full consultation and orientation probably need to be increased (Bamberger, 2000).

Quasi-Experimental Impact Evaluations

Despite the relatively widespread belief that quasi-experimentation is neither feasible nor useful in international program evaluation because of the costs, methodological difficulties, and long delays in producing results (Valadez & Bamberger, 1994), it has proved possible to apply quasi-experimentation in a cost-effective manner in a number of studies. Valadez and Bamberger (1994) report several successful applications of these methods in the World Bank. Examples include:

- A longitudinal study, using a comparison group design with interviews conducted at three points in time, with project beneficiaries and three comparison groups to evaluate the impacts of a housing program on the income and expenditures of low-income households in El Salvador. Project beneficiaries were compared with families in the three main kinds of housing settlements from which participants had been selected. The study was able to identify a number of social and economic impacts of the program, such as more stable employment, greater female labor force participation, and increased self-respect by families owning their own home.
- Studying the impacts of educational television in the Philippines by using a design which compares students in three groups of schools before and after the introduction of television. One group of schools used television, another group combined television with teacher training, and the third group received no intervention. The study was able to assess the gains from the use of television and also the added value from the additional investment in teacher training.
- Studies in Peru used a quasi-experimental design with an intervention and control group to assess the fertility impacts of different numbers of medical visits following the insertion of an IUD (Dennis & Boruch, 1989).

Others note the opportunities for using experimental designs. Newman, Rawlings, and Gertler (1994) cite seven studies in Bolivia, Colombia, Indonesia, Mexico, Nicaragua, Taiwan, and Turkey to demonstrate the feasibility of rigorous impact evaluation studies in a number of settings. Programs evaluated include radio programs to teach mathematics in Nicaragua, early childhood education in Turkey, alternative service delivery systems for family planning, cognitive abilities of malnourished children in Colombia, "Sesame Street" in Mexico, educational services assigned by lottery in Bolivia, and resource mobilization in Indonesia. Newman et. al. point out that the need to ration service provision in many

developing countries, often using lotteries and random assignment, provides many opportunities to use randomized experimental designs.

Finally, a greater effort should be invested in developing rapid, methodologically sound qualitative evaluation methods that could be used to complement quantitative methods. Advances have been made in the use of small samples that can produce statistically valid findings more rapidly and economically (Valadez & Bamberger, 1994). The Operations Evaluation Department (OED) of the World Bank has conducted rapid and economical ex-post impact studies that combine the exploitation of existing data sets with rapid sample surveys, tracer studies to locate individuals or families who have left the project area, key informant interviews, and a number of qualitative methods (Valadez & Bamberger, 1994). These approaches have been used successfully, for example, to evaluate the impacts of irrigation and resettlement projects.

CONCLUSION

One of the major differences between evaluations in the U.S. and those in developing countries is the major role played by international donor agencies as promoters, executors, and consumers of the evaluations conducted in developing countries. A key challenge is to reconcile the information requirements of U.S. and European governments and funding agencies, and the preferred evaluation approaches in these countries, with the development information needs, research traditions, and social and political contexts in the host countries. Some of the most interesting developments in evaluation design and utilization concern participatory evaluation as a way to give voice to the powerless; rapid, cost-effective impact assessments; assessment of the impacts of development on women; and the recent availability of longitudinal country socioeconomic data sets. In areas such as these, a closer exchange of experiences between U.S. evaluation practitioners and their colleagues from developing countries could be mutually beneficial.

ACKNOWLEDGMENT

An earlier version of this paper was presented at the AEA conference in San Diego in November 1987. The views expressed are those of the author and should not be attributed to the World Bank.

REFERENCES

- Bamberger, M. (1999). *Evaluating the participation of women in development projects: The case of Social Development Funds*. Paper presented at the AEA conference, Orlando, FL, November 4.
- Bamberger, M. (Ed.) (2000). *Integrating quantitative and qualitative research in development projects*. Directions in Development Series. Washington, DC: The World Bank.
- Deaton, A. (1997). *The analysis of household surveys: A microeconomic approach to development policy*. Washington, DC: The World Bank.
- Deininger, K., & Squire, L. (1995). *Measuring income inequality: A new data base*. Washington, DC: The World Bank, Policy Research Department.

- Demery, L., Chen, B., & Vishwanath, B. (1995). *Poverty, inequality and growth*. ESP Discussion Paper No. 70. Washington, DC: The World Bank.
- Dennis, M., & Boruch, R. (1989). Randomized experiments for developing and testing projects in developing countries. *Evaluation Review*, 13, 292–309.
- Fetterman, D., Kaftarian, S., & Wandersman, A. (Eds.) (1996). *Empowerment evaluation: Knowledge and tools for self-assessment and accountability*. Thousand Oaks, CA: Sage.
- Narayan, D. (1997). *Participation toolkit*. Washington, DC: The World Bank.
- Newman, J., Rawlings, L., & Gertler, P. (1994). Using randomized control designs in evaluating social sector programs in developing countries. *World Bank Research Observer*, 9(2), 181–198.
- Ravallion, M., & Chen, S. (1996). *What can new survey data tell us about recent changes in living standards in developing and transitional economies?* Washington, DC: The World Bank, Policy Research Department.
- Sanders, J. (1997). Cluster evaluation. In E. Chelimsky & W. Shadish (Eds.), *Evaluation for the 21st century*. Thousand Oaks, CA: Sage.
- Valadez, J., & Bamberger, M. (1994). *Monitoring and evaluating social programs in developing countries: A handbook for policymakers, managers and researchers*. Washington, DC: The World Bank, Economic Development Institute.
- The World Bank. (1996). *Poverty reduction and the World Bank: Progress and challenges in the 1990s*. Washington, DC.