



Reconstructing Baseline Data for Impact Evaluation and Results Measurement



Michael Bamberger

Many international development agencies and some national governments base future budget planning and policy decisions on a systematic assessment of the projects and programs in which they have already invested. Results are assessed through Mid-Term Reviews (MTRs), Implementation Completion Reports (ICRs), or through more rigorous impact evaluations (IE), all of which require the collection of baseline data before the project or program begins. The baseline is compared with the MTR, ICR, or the posttest IE measurement to estimate changes in the indicators used to measure performance, outcomes, or impacts. However, it is often the case that a baseline study is not conducted, seriously limiting the possibility of producing a rigorous assessment of project outcomes and impacts. This note¹ discusses the reasons why baseline studies are often not conducted, even when they are included in the project design and funds have been approved, and describes strategies that can be used to “reconstruct” baseline data at a later stage in the project or program cycle.

Baseline Data Are Important for Assessing Project/Program Results and Impacts

Baseline data can come from the project’s monitoring and evaluation (M&E) system, rapid assessment studies, surveys commissioned at the start and end of the project, or from secondary data sources. Whatever the source, the availability of appropriate baseline data is always critical for performance evaluation, as it is impossible to measure changes without reliable data on the situation before the intervention began. Despite the importance of collecting good baseline data, there are a number of reasons why they are frequently not collected, and the purpose of this paper is to present a range of strategies that can be used for “reconstructing” baseline data when they are not available.

The strategies for reconstructing baseline data apply to both discrete projects and broader programs (the term “interventions” is used here to cover both), although they must sometimes

be adapted to the special characteristics of each. Projects often introduce new M&E systems customized to the project’s specific data needs, but often with significant start-up delays, which can be problematic for collecting baseline data. In contrast, ongoing programs can often build on existing M&E and other data collection systems as well as have access to secondary data and sampling frames, although these systems are often not sufficient for the purposes of evaluation and tend to be difficult to change. Nongovernmental organizations (NGOs), important development players in many countries, may face different issues with respect to baseline data for their activities.

Limited Access to Baseline Data in Many Projects/Programs

Although most interventions plan to collect baseline data for results monitoring and possibly impact evaluation, often data are not collected or collection is delayed until the intervention has been underway for some time. The reasons may

include a lack of awareness of the importance of baseline data, a lack of financial resources, or limited technical expertise. Even when management recognizes its importance, administrative procedures (for example, recruiting and training M&E staff, purchasing computers, or commissioning consultants) may create long delays before baseline data can be collected.

Data for Results Monitoring and Evaluation

M&E systems collect baseline information on indicators for measuring program outputs and outcomes for the target population. Impact evaluations collect similar information, but from both beneficiaries and a comparison group. Information is also collected on the social and economic characteristics of individuals, groups or communities; on contextual factors such as local economic conditions; and on political and organizational factors that might explain variations in outcomes and impacts among different project locations.

The World Bank and other development agencies incorporate this information into a Results-Based Monitoring and Evaluation System (RBME). Kusek and Rist (2004) describe a 10-step system for implementing RBME, 3 of which involve the creation of a baseline:

Step 2: Agreeing on the outcomes to monitor and evaluate

Step 3: Selecting key indicators to monitor outcomes and performance

Step 4: Collecting baseline data

Strategies for Reconstructing Baseline Data

This section presents some practical strategies for estimating (“reconstructing”) conditions of the project, and sometimes also the comparison group, at the time the intervention is launched. Most of these are economical, relatively simple to apply, and do not require too great an investment of time.

Timing of the baseline

Evaluations often implicitly assume that an intervention only starts to produce impacts after it officially begins, but, in fact, changes may occur long before this. For example, once it is known

that roads, water supply, or other services are to be provided to certain communities, speculators may begin to buy land and families may start to make improvements to their property. If the baseline is not conducted until the official program launch, many of these important changes may not be captured. Using techniques such as recall or key informant interviews to capture information on these early changes should be considered.

Using secondary data to reconstruct the baseline

There are many documentary sources that may provide information on the beneficiary population or comparison groups around the time the intervention began. Censuses covering areas such as population, agriculture, industry, education, and environment may be available. Other useful sources are household socioeconomic surveys, the largest of which are the Living Standards Measurement Surveys (LSMS), which have been conducted in at least 35 countries. When surveys are repeated periodically, it may be possible to find a reference point close to the intervention launch date. However, while many surveys have a large enough sample to generate a comparison group, the samples are often too small or do not contain sufficiently detailed information to generate a sample of the beneficiary population (particularly when this population is relatively small).

Ministries of education, health and agriculture, among others, publish annual reports that can provide baseline reference data, and they can sometimes provide information on particular schools, health centers, or other facilities in the target areas. Donor agencies, NGOs, and universities also conduct studies providing useful reference data. Birth and death certificates can be used to examine life expectancy, family size and common causes of death, while legal documents relating to marriage and divorce can provide information on, for example, the property rights of women. Mass media also provide information on issues concerning local schools, clinics, public transport, and so forth that can provide background information on conditions at the start of the intervention. Box 1 presents two examples where secondary data were used to reconstruct baseline data for matched project and comparison groups using propensity score matching.

Box 1. Using Secondary Sources to Reconstruct Baseline Data

The evaluation of the Nicaragua Emergency Social Investment Fund used the 1998 LSMS, conducted five years earlier, to identify baseline project and comparison communities for each project component (water supply and sanitation, health, education, and so forth). Propensity scores were then used to improve the match of the two samples (Pradhan and Rawlings 2002).

The World Bank Operations Evaluation Department (OED, now the Independent Evaluation Group [IEG]) evaluation of the impacts of the Bangladesh Integrated Nutrition Project used three separate secondary data sets to reconstruct the baseline and to monitor implementation progress. Each survey had strengths and weaknesses, some having more information on project implementation while others had more demographic and nutrition information. The study created a new comparison group using propensity score matching. Combining the data sources reduced bias and strengthened the validity of the impact estimates (OED 2005).

Source: Compiled by author.

Box 2. Using Project Administrative Data to Reconstruct the Baseline

A recent post-test multidonor evaluation of the Nepal Education for All Project used the project's education management information system to obtain sex-disaggregated data on school enrollment, repetition, and academic tests scores at the start of the program and at various key milestones during implementation (NORAD 2009).

The evaluation of the poverty impacts of the World Bank–financed Vietnam Rural Roads Project used administrative data collected at the canton level to understand the selection criteria used to determine where roads would be built and to monitor the quality and speed of construction (van de Walle 2009).

The evaluation of the feeder roads component of the Eritrea Community Development Fund used planning and feasibility studies commissioned by the implementing agency to obtain socioeconomic baseline data on the communities affected by the roads (unpublished national consultant report).

Source: Compiled by author.

There are a number of factors affecting the utility and validity of secondary data sources: the data cover the wrong reference period; key information is missing; information was not collected from the right people (for example, only the household head was interviewed); the sample does not cover the whole population of interest or is too small; or the information is not reliable or complete. These factors must always be assessed before utilizing any of these sources.

Using administrative data from the intervention

Many interventions collect monitoring and other kinds of administrative data that could be used to estimate baseline conditions for the target population (box 2). For example, socioeconomic data included in the application forms of people, communities, or organizations applying to participate or receive benefits; planning and feasibility studies; monitoring reports; and administrative records providing information such as changes in project eligibility criteria or the services provided

to particular beneficiaries. Sometimes the application forms for people not accepted can provide a comparison group of nonparticipants.

While administrative data are a potentially valuable source of baseline data, the data are often not available in a convenient format for analysis. Often the evaluator must work closely with program staff to ensure that administrative data are collected and filed in a usable format (discussed further later in this note). Often when the evaluator discovers that the expected administrative records have vanished or are not organized in a usable format, staff respond “No one told us that this information would be required for a future evaluation.” Better coordination between the evaluators and the program staff might have ensured the information would be available.

Recall

Recall techniques ask individuals or groups to provide information on their social or economic conditions, their access to services, or the conditions of their community at a particular point in time (for example, project launch) or over a particular period of time. *Recall* is used in poverty analysis,

demography, and income expenditure surveys (Deaton and Grosh 2000) to elicit information on behavior (for example, contraceptive usage or fertility) or economic status (household income or expenditure). Several comparative studies (for example, Deaton and Grosh [2002]; Belli, Stafford, and Alwin [2009]) have concluded that recall, when carefully designed and implemented, can be a useful estimating tool with predictable and, to some extent, controllable errors, and a potentially valuable way to reconstruct baseline data.

Recall can be applied through questions in surveys and individual or group interviews (box 3). In addition to collecting numerical data such as income or farm prices, recall can also be used to obtain estimates of major changes in the welfare conditions of the household, such as which children attended a school outside the village before the village school opened and the travel time and costs of getting there. Families can also provide information on questions such as access to health facilities and where they previously obtained water and how much it cost.

Box 3. Using Recall to Reconstruct Baselines

An evaluation of the gender impacts of a World Bank–financed rural roads project in Bangladesh asked women to recall the situation of their family before the upgrading of the roads and to compare this with their current situation on 20 indicators of economic status, access to water, quality of housing, and consumption of basic items. The differences were used to estimate project impact (unpublished study funded through the World Bank Gender and Transport Initiative [2000]).

An evaluation of the impacts of the village school construction component of the Eritrea Community Development Fund asked families to recall which of their children had attended school before the school was built in their village, how far the children had to travel, and the means and cost of transport. Families were able to clearly recall all of this information, and in the opinion of the researchers the information was reliable because the respondents had no incentives to distort their answers (unpublished consultant report).

Source: Compiled by author.

Recall always involves a risk of bias due to memory or distortion. *Unintentional* distortion occurs when, for example, people romanticize the past (“when I was young there was much less crime in the community”) or unintentionally adjust their response to what they think the researcher wants to hear. *Intentional* distortion occurs when, for example, families are reluctant to admit their children had not been attending school, or they might underestimate how much they spend on water to convince planners they are too poor to pay the water charges proposed in a new project. The reliability of recall data also depends on the nature of the outcome variable being studied. For example, families will usually be able to recall major events such as a death in the family or enrollment of a child in school, but it may be more difficult to obtain reliable responses on nutrition questions or changes in the frequency of diarrhea or other very common ailments.

A challenge in using recall is the absence of studies providing guidelines for estimating or adjusting for systematic bias. The most detailed research on this question was conducted on the recall of expenditures in national household income and expenditure surveys and studies on fertility. The income and expenditure studies identified some consistent biases that can be used to adjust estimates: “telescoping,” that is, reporting major expenditures as being more recent than they actually were, and underestimating small expenditures. Also, men and the better off are more likely to report they have been sick than are women and poorer people. Other areas where research on the validity and reliability of recall is available include: substance abuse, adolescent health research, assessment of stressful events, and time use. Belli, Stafford, and Alwin (2009) report that the reliability of recall is significantly enhanced when using the calendar method of life course research (in which topics of interest are linked to critical events in the life course of the subject: birth, death, marriage, enrollment in school, and changing employment) compared to conventional recall questions in a structured questionnaire.

Recall can sometimes provide better self-assessment estimates of behavioral changes and knowledge (for example, child care and nutrition,

leadership skills) than pre- and post-test comparisons. People often overestimate their behavioral skills or knowledge before entering a program because they do not understand the tasks being studied or the required skills. After completing the program, they may have a better understanding of these behaviors and provide a better assessment of their previous level of competency or knowledge and how much these have changed (Pratt, McGuigan, and Katzeva 2000).

Key informants

Key informants (box 4) can provide knowledge and experience on a particular agency and the population it serves, an organization (such as a trade union, women's group or a gang), or group (such as mothers with young children, sex workers, or landless farmers). For example, when evaluating a program to increase secondary school enrollment, key informants could include: school directors, teachers and other school personnel, parents of children who do and do not attend school, students, and religious leaders.

Key informants combine "factual" information with a particular point of view, and it is important to select informants with differing perspectives. For example, low- income and higher-income parents may have different opinions on programs to increase school enrollment, as may those from different ethnic or religious groups.

Group interview techniques for reconstructing baseline data

Focus groups are used in market research and program evaluation to obtain information on socioeconomic characteristics, attitudes, and behaviors of groups that share common attributes (Krueger and Casey 2000). Groups, usually with five to eight persons per group, are selected to cover different economic strata, as well as people who have and have not participated in the project or who received different services. The group moderator goes systematically through a checklist of questions making sure each person responds to every question. For the purposes of reconstructing baseline data, participants could be asked to provide information on, for example, conditions of their household, group, community, or agricultural production at some point in the past. *When properly designed and implemented*, focus groups ensure that all key

Box 4. Storeowners as Key Informants on Economic Trends in the Community

In the evaluation of urban housing programs in Latin America, local storeowners were a valuable source of information on changes in the economic conditions of the community. For example, an increase in the demand for meat was an indicator that households had more disposable income. Most stores provide credit and the level of default was another good indicator of changing community fortunes. Storeowners also have a good memory for trends in the sales of key items over long periods of time.

Source: Compiled by author.

sectors are sampled and that responses provide a representative snapshot of each group. However, readers of evaluation reports should be aware that focus groups are often used in development evaluation as a fast and economical way to obtain general information on the opinions of the target population with very little attention to participant selection or ensuring balanced participation in the discussion. Market research companies make extensive use of focus groups, developing sampling frames to select samples with the socioeconomic characteristics required by different clients. If funds are available, contracting a market research company to design and implement focus groups for a program evaluation could be considered.

Participatory assessment techniques (PRAs), originally meaning "participatory rural appraisal," is now used as a generic term for all participatory studies in which communities or groups report on their conditions, problems, and changes over time. Groups can provide estimates on things such as the volume and quality of water, crop production and sales, travel time and costs, and time use. PRAs are widely used with poor rural and urban communities with low literacy levels or where participants have difficulties in expressing complex ideas (such as changes in environmental conditions). PRAs include construction of charts, maps, or tables where the group agrees on the placement of familiar objects, such as stones or seeds, on a chart to illustrate trends, important events, magnitude, or causal patterns. Timelines, trend analysis, historical

transects, seasonal diagrams, and daily activity schedules can be used to assess changes over time or the situation at the baseline reference point (Kumar 2002).

These PRAs have several benefits. Respondents may feel more comfortable expressing themselves in a group with their peers, rather than in a one-on-one interview with an outside researcher. The group consensus can also provide a cost-effective way to obtain an approximate estimate of average travel time, volume and quality of water consumed, volume of agricultural production, and average crop prices rather than having to use a sample survey. Synergistic group interaction also generates new ideas that might not have come up in one-on-one interviews. There are also potential risks: the group may be dominated by a few vocal people; participants may defer to politically powerful, wealthier, or more educated group members; or the group facilitator may inadvertently direct the group toward certain decisions.

Applying the Reconstruction Strategies to Fill In Baseline Data for the RBME

M&E systems often take some time to get established, so there may be a period at the start of the intervention when monitoring data are not being collected. So when setting up the RBME, a first step should be to check: What are the key indicators on which baseline data are required? Which indicators are available and which are missing? Why are the data missing and how easily can the problems be overcome? Is any important information not being collected during the interim period before the monitoring system becomes fully operational? All of the techniques for reconstructing baseline data can be applied to filling in RBM baseline data gaps.

RBME systems are usually based on a program theory model that includes: how the program is intended to achieve its objectives, implementation and outcome indicators that should be measured, key assumptions to be tested, and the time horizon over which different results are to be achieved (Bamberger, Rugh, and Mabry 2006, chapter 9). Often the program theory model was not in fact defined or fully articulated at the start of the project. In these cases, the evaluator may need to work with the implementing agency and other

stakeholders to reconstruct the implicit program theory on which the program is based. Sometimes there is agreement among staff concerning the underlying theory model and all that is needed may be a short workshop to put this on paper. However, in other cases, staff may have difficulty articulating the model or there may be disagreements concerning the purpose of the program, how it will achieve its outcomes, and the critical assumptions on which it is based.²

Applying Baseline Reconstruction Strategies for Evaluating Outcomes and Impacts

There is a wide variety of evaluation designs for estimating project impacts and effects ranging from strong statistical designs with before-and-after comparisons of project and comparison groups, to statistically weaker quasi-experimental designs that may not include baseline data on the comparison or project groups, and nonexperimental designs that do not include a comparison group. Different baseline reconstruction strategies can be applied to different evaluation designs. For the weaker quasi-experimental designs and non-experimental designs where no baseline data have been collected for the project and/or the comparison group, all of the baseline reconstruction techniques discussed earlier could be considered. On the other hand, the stronger quasi-experimental and the experimental designs all include baseline data for both project and control groups. However, in most cases only quantitative data are collected (for example, the number of students enrolled in school or patients visiting health centers), and the design would be strengthened by complementing this with qualitative data such as the quality of services, women's participation in household decision making at the time the project began, and how different ethnic groups were received when they visited health clinics.

Quantitative and qualitative evaluations rely on different types of data and data collection procedures. When quantitative researchers collect primary data to reconstruct baselines, they are likely to incorporate recall questions into a structured questionnaire. In contrast, qualitative researchers use a wider range of

techniques, including key informants, in-depth individual interviews, focus groups, and PRAs. Both quantitative and qualitative research designs can benefit from incorporating mixed-method approaches to baseline reconstruction so as to combine depth of understanding with generalizability of the findings (Bamberger, Rao, and Woolcock 2010).

Special Issues: Reconstructing Baseline Data for Comparison Groups

Selecting a well-matched baseline comparison group presents special challenges. Participant selection procedures often result in project participants having special attributes that affect, and frequently increase, the probability of successful program outcomes. Often these attributes, termed “unobservables” or “omitted variables,” are not included in the baseline surveys. For example, in a microcredit program for women, many of the women who are successful in starting or expanding small businesses might come from households where they have more control over household decision making than is normally the case in their community, or they may have previous experiences with a small business. These characteristics might affect project outcomes, but this information will usually not be included in the baseline data. The following methods could be used to assess the importance of these omitted variables: key informant interviews (for example, staff of microcredit and other economic development programs); administrative data from the loan programs; focus groups with women participants and nonparticipants; in-depth interviews with participants and nonparticipants; and PRAs.

Operational Implications

Even when an agency is strongly committed to setting up an M&E system to generate the baseline data required for results-based management and impact evaluation, there are often other pressing staffing, organizational and financial matters, so there will often be considerable delays before the M&E systems are operational. The following are measures that can be taken to increase the likelihood that the M&E systems will be in place from the time of program launch:

- Define funding arrangements that avoid long delays in contracting monitoring unit staff and commissioning evaluation consultants.
- Begin recruiting M&E staff before intervention launch.
- Arrange for M&E staff to receive basic training before intervention launch.
- Early recruitment of an experienced M&E staffer.

Having staff on board who are familiar with the practical and technical problems faced when trying to reconstruct baseline data can avoid many of the problems that typically occur when generalist task managers attempt to handle these problems themselves.

There are a number of practical ways to enhance an agency’s ability to generate baseline data. Using evaluation funds to contract additional administrative staff may remove bottlenecks and facilitate good quality data collection. In other cases, baseline data on target households, communities, or organizations such as schools, health clinics, or agricultural cooperatives may not be organized or archived in a way that facilitates identification of a comparable sample one or two years later for repeat interviews. Discussions with agency staff at the planning stage could ensure that valuable data such as application forms that include socioeconomic data on households or communities applying to participate in a project or program, or feasibility studies for the selection of roads to be built or upgraded, are not discarded once beneficiaries have been selected or the sites for road improvements chosen. Effective coordination with agency staff is critical.

M&E systems compare progress at different points over the life of the project, and “baseline” data for these comparisons must be collected throughout the life of the project. So it is important to ensure M&E systems continue to provide good quality data. The following are recommendations that can help sustain M&E systems:

- Check the budget allocated to effective M&E systems in other organizations and ensure sufficient resources are allocated in the present program.
- Ensure that specific and adequate budget line items for M&E are approved and reauthorized when necessary in the relevant government budgets.
- Organize workshops for management and policy makers to explain the benefits of good

M&E data and explain how the costs of both monitoring and evaluation are calculated. Prepare case studies on how M&E systems were organized and used in other projects, and establish contact with these agencies through study tours, videoconferencing, or visits of resource persons.

- Ensure that stakeholders are actively involved in the planning and design of the M&E systems and that the systems respond to their information needs (Patton 2008).
- Use clients' preferred communication style for presenting evaluation findings so that stakeholders are able to use information generated from the M&E system and are motivated to support the continued collection of the data (Vaughan and Buss 1998; Patton 2008).
- A continuing evaluation capacity development (ECD) program is essential to ensure upgrading of the evaluation skills of agency and consultant staff involved with M&E.

The willingness of agency staff to continue to collect and deliver good quality data to the M&E unit is critical. How can staff be motivated to continue to produce this information month after month and year after year?

- Collection and transmission of M&E data should be simple and rapid.
- Provide evidence to staff that the information they collect is used. Staff should receive regular feedback on issues or questions arising from their data, and they should be asked for further information on examples of successes or unanticipated problems.
- Staff should receive recognition through personal thanks from headquarters, invitation to prepare an article for a newsletter, or a small prize from time to time.
- Provide evidence to staff showing that the data helps improve the quality of the programs. For example, the evaluation of the Uganda Education for All Program made extensive use of monitoring data in the follow-up evaluations at the district level. Local staff reported this was the first time they had seen their data being used and this gave them an incentive to improve the quality of data collection (Bamberger and Kirk 2009).
- Mackay (2007) argues that a strategy of incentives to develop and sustain an M&E system

requires *carrots* (for example, budgetary incentives and greater management autonomy to programs that use M&E well); *sticks* (for example, laws and regulations mandating M&E or withholding funding from agencies that fail to implement M&E); and *sermons* (for example, high-level endorsements of M&E importance).

Implications for National Planning Agencies

National sample surveys conducted at least once a year on topics such as income and expenditure, access to health or education, or agricultural production provide very valuable baseline data for results-based management and impact evaluation. Household income and expenditure surveys are one example that has proved very valuable. If these surveys can be used in the evaluation of several different development programs, they become very cost-effective and they also can provide a larger and methodologically more rigorous comparison group sample than an individual evaluation could afford. Regularly repeated surveys provide a very valuable longitudinal database that can control for seasonal variation and economic cycles.

The value of these surveys for results-based management and impact evaluation can be greatly enhanced if they are planned with this purpose in mind and in coordination with the agencies and donors who may use the surveys to generate baseline data and comparison groups. Some of the ways to enhance their utility include:

- Ensure the sample is sufficiently large and has a sufficiently broad regional coverage to generate subsamples covering particular target populations with sufficient statistical power to be used for major program evaluations.
- Include, in consultation with social sector agencies, core information on topics such as: school enrollment, access to health services, and participation in major development programs. This would facilitate selecting samples of participants and comparison groups for impact evaluations.
- Include one or more special modules in each round of the survey to cover the needs of a particular evaluation that is being planned.
- Document the master sampling frame to facilitate its use for selecting samples for particular evaluations.

Many of these approaches can only be considered for large and expensive evaluations or for studying issues that are of high priority to government agencies and/or donors. Also, national statistics offices are typically overburdened, so they can only be expected to help out when the program is particularly important or when special funding can be arranged to cover the costs of additional staff for data collection or analysis.

Conclusions

Good quality baseline data that measure the conditions of the target population and the matched comparison group are an essential component of effective monitoring, results-based management, and impact evaluation. Without this reference information, it is very difficult to assess how well a project or program has performed and how effectively it has achieved its objectives or results.

However, many projects and programs fail to collect all of the required baseline data. While some of the reasons for this can be explained by inadequate funding or technical difficulties in collecting the data (particularly for control groups), many of the causes could be at least partially corrected by better management and planning. Many reasons relate to administrative delays in releasing funds and recruiting and training staff and contracting consultants. While administrative procedures (such as those relating to personnel and consultants) are often difficult to change, ways could probably be found to reduce some of these delays. Other issues concern the relatively low priority that is often given to M&E, particularly when there are so many other urgent priorities during the early stages of a project or program.

Even with the best of intentions, these administrative challenges will never be completely resolved and there will continue to be many situations where the collection of baseline monitoring data is delayed and the commissioning of baseline studies for impact evaluations never takes place. Included in this note are a range of strategies, many of them relatively simple and cost-effective, for reconstructing baseline data when necessary. It is recommended that appropriate tools should be built into RBME and impact evaluation systems as contingency tools for reconstructing important baseline data. While some of the statistical techniques such as propensity score matching have been widely used and their strengths and

weaknesses are well understood, others such as recall or the systematic use of key informants have often been used in a somewhat ad hoc manner and more work is required to test, refine, and validate the methods. Finally, there are many potentially valuable sources of administrative data from the project itself that tend to be underutilized and more attention should be given to the development and use of these valuable and relatively accessible sources of information.

About the Author

Michael Bamberger has a PhD in Sociology from the *London School of Economics*. He worked for 23 years with the *World Bank* as advisor on monitoring and evaluation to the *Urban Development Department*, training coordinator for Asia and senior sociologist in the *Gender and Development Department*. Since retiring in 2001, he has worked as an evaluation consultant and evaluation trainer with 10 United Nations agencies, the World Bank, the Asian Development Bank, and a number of bilateral development agencies and developing country governments. He has published extensively on evaluation and is on the editorial board of several evaluation journals.

Notes

1. The author wishes to thank these colleagues from the Poverty Reduction and Equity Group: Jaime Saavedra (Acting Sector Director), Gladys Lopez Acevedo (Senior Economist), Keith Mackay (Consultant), Emmanuel Skoufias (Lead Economist), Philipp Krause (Consultant), and Helena Hwang (Consultant) for comments.
2. See Bamberger, Rugh, and Mabry (2006, 179–82) for a discussion of the different strategies for reconstructing a program theory model.

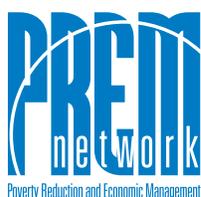
References

- Bamberger, M., and A. Kirk. 2009. *Making Smart Policy: Using Impact Evaluation for Policy Making: Case Studies on Evaluations That Influenced Policy*. PREM Thematic Group for Poverty Analysis, Monitoring and Impact Evaluation, Doing Impact Evaluation Series No. 14, World Bank, Washington, DC.
- Bamberger, M., Rugh, J., and L. Mabry. 2006. *Real World Evaluation: Working under Budget, Time, Data and Political Constraints*. Thousand Oaks, CA: Sage Publications.
- Belli, F., F. Stafford, and D. Alwin. 2009. *Calendar and Time Diary Methods in Life Course Research*. Thousand Oaks, CA: Sage Publications.

- Bourguignon, F. 2009. "Toward an Evaluation of Evaluation Methods: A Commentary on the Experimental Approach in the Fields of Employment, Work, and Professional Training." *Journal of Development Effectiveness* 2 (3): 310–19.
- Deaton, A., and M. Grosh. 2000. "Consumption." In *Designing Household Survey Questionnaires for Developing Countries: Lessons from 15 Years of the Living Standards Measurement Study*, Vol. 3, ed. M. Grosh and P. Glewwe, 91–134. Washington, DC: World Bank.
- Gibson, J. 2006. "Statistical Tools and Estimation Methods for Poverty Measures Based on Cross-Sectional Household Surveys." In *Handbook on Poverty Statistics: Concepts, Methods and Policy Use*, 128–205. United Nations.
- Krueger, R., and M. Casey. 2000. *Focus Groups: A Practical Guide for Applied Research*, 3rd Edition. Thousand Oaks, CA: Sage Publications.
- Kumar, S. 2002. *Methods for Community Participation: A Complete Guide for Practitioners*. Rugby, Warwickshire: Practical Action.
- Kusek, J., and R. Rist. 2004. *Ten Steps to a Results-Based Monitoring and Evaluation System*. Washington, DC: World Bank.
- Mackay, K. 2007. *How to Build M&E Systems to Support Better Government*. Washington, DC: World Bank Publications, Independent Evaluation Group.
- OED (Operations Evaluation Department). 2005. *Maintaining Momentum to 2015? Impact Evaluation of Interventions to Improve Maternal and Child Health and Nutrition in Bangladesh*. Washington, DC: World Bank.
- Patton, M. 2008. *Utilization-Focused Evaluation*, 4th Edition. Thousand Oaks, CA: Sage Publications.
- Pradhan, M., and L. Rawlings. 2002. "The Impact and Targeting of Social Infrastructure Investments: Lessons from the Nicaraguan Social Fund." *World Bank Economic Review* 16 (2): 275–95.
- Pratt, C., W. McGuigan, and A. Katzeva. 2000. "Measuring Program Outcomes: Using Retrospective Pretest Methodology." *American Journal of Evaluation* 21 (3):341–49.
- Van de Walle, D. 2009. "The Poverty Impact of Rural Roads Projects." *Journal of Development Effectiveness* 1 (1): 15–36.
- Vaughan, R., and T. Buss. 1998. *Communicating Social Science Research to Policymakers*. Thousand Oaks, CA: Sage Publications.
- White, H. 2006. "Impact Evaluation: Experience of the Independent Evaluation Group of the World Bank." IEG, Evaluation Capacity Development Series. The World Bank.

Further Reading

- Gorgens, M., and J. Z. Kusek. 2010. *Making Monitoring and Evaluation Systems Work: A Capacity Development Toolkit*. Washington, DC: World Bank.
- Khandker, S., G. Koolwal, and H. Samad. 2009. *Handbook on Impact Evaluation: Quantitative Methods and Practices*. Washington, D.C: World Bank.
- Pretty, J., I. Guijt, J. Thompson, and I. Scoones. 1995. *A Trainer's Guide for Participatory Learning and Action*. London. International Institute for Environment and Development.
- Silverman, D. 2004. *Qualitative Research: Theory, Method and Practice*, 2nd Edition. Thousand Oaks, CA: Sage Publications.
- Teddlie, C., and A. Tashakkori. 2008. *Foundations of Mixed Methods Research: Integrating Quantitative and Qualitative Approaches in the Social and Behavioral Sciences*. Thousand Oaks, CA: Sage Publications.



This notes series is intended to summarize good practices and key policy findings on PREM-related topics. The views expressed in the notes are those of the authors and do not necessarily reflect those of the World Bank. PREMnotes are widely distributed to Bank staff and are also available on the PREM Web site (<http://www.worldbank.org/prem>). If you are interested in writing a PREMnote, email your idea to Madjiguene Seck at mseck@worldbank.org. For additional copies of this PREMnote please contact the PREM Advisory Service at x87736.

This series is for both external and internal dissemination