Chapter 3 Monitoring and Evaluation

Chapter 3
Monitoring and Evaluation

Giovanna Prennushi, Gloria Rubio, and Kalanidhi Subbarao

Tables

Figures

Boxes

Technical Notes *(see Annex C, p. 433)*

Case Studies *(see Annex C, p. 436)*

3.1 Introduction

This chapter assists countries in developing a system to monitor and evaluate whether a poverty reduction strategy is effective in reducing poverty. How do we know if a poverty reduction strategy is effective?

First, a poverty monitoring system is needed to track key indicators over time and space and to determine if they change as a result of the strategy. Section 3.2 of the chapter therefore discusses setting up a poverty monitoring system: how to define key indicators, track them over time, and determine what changes have taken place. Many countries already have poverty monitoring systems in place, so the task is to assess their adequacy and strengthen them as necessary. Experience shows that elements such as the tracking of public expenditures and outputs and quick monitoring of household well-being need special attention. Participatory data collection methods and qualitative information give a different perspective and should not be overlooked.

Second, rigorous evaluations should be done selectively to assess the impact on poverty of interventions that are key components of the strategy. Section 3.3 discusses the decision to conduct a rigorous impact evaluation and explains its design and implementation, including necessary data for different methodologies.

Other types of evaluation, such as assessing the process of formulating a poverty reduction strategy, can also be useful. Section 3.4 briefly discusses this topic, as thus far only limited experience exists. This section also briefly discusses another challenging topic: evaluating the impact of poverty reduction strategies in general as opposed to the impact of specific components of a strategy, such as programs or single policies. The key point is that a solid monitoring system will provide the basic data necessary to conduct such evaluations, should the need arise in the future.

Both monitoring and evaluation activities need to be carried out by institutions that are competent and that have strong links to key decisionmakers, if they are to be useful in the design and implementation of a poverty reduction strategy. Much monitoring and evaluation takes place without adequate development of in-country capacity and without strong links to key decisionmaking processes; thus precious opportunities to learn what works and what does not are lost, sometimes along with funds. Section 3.5 offers guidance on building capacity, particularly strengthening the processes that provide policymakers and others with feedback on the impact of policies and programs. A key message of this section is that dissemination of results is critical for use. Results that are not widely disseminated, through mechanisms tailored to different groups in civil society, will not be used, and the resources spent in getting such results will be wasted.

Nongovernmental actors—research institutions, civil society organizations, special-interest and advocacy groups, and others—have an important role to play in the design of the monitoring and evaluation system, in actually carrying out monitoring and evaluation activities, and in using the results. Section 3.6 discusses the role of these actors.

A Guide to Web Resources at the end of the chapter contains references to Web and other sources of information. Technical notes and case studies provide more detail on specific topics and country examples.

3.2 Setting Up a Poverty Monitoring System

To know if a poverty reduction strategy is effective in reducing poverty, it is necessary to set in place a system to monitor progress. This section discusses the features of such a system and issues encountered frequently during implementation.

3.2.1 Defining goals, indicators, and targets

Before a monitoring system can be set up to assess whether a poverty reduction strategy is effective in reducing poverty, it is necessary to agree on which poverty reduction goals the strategy wants to achieve, select key indicators, and set targets for such indicators.

There are probably many possible definitions of these terms, but the following are used in this book:

Goals are the objectives a country or a society wants to achieve; they are often expressed in nontechnical, qualitative terms, such as "eradicate hunger" or "reduce poverty."

Indicators are the variables used to measure progress toward the goals. For example, progress toward eradicating hunger could be measured by looking at the number of families who say they are not able to have three meals a day all 12 months of the year.

Targets are the quantified levels of the indicators that a country or society wants to achieve at a given point in time—for example, a target of all families being able to eat three meals a day all 12 months of the year by 2015.

Example: The Millennium Development Goals

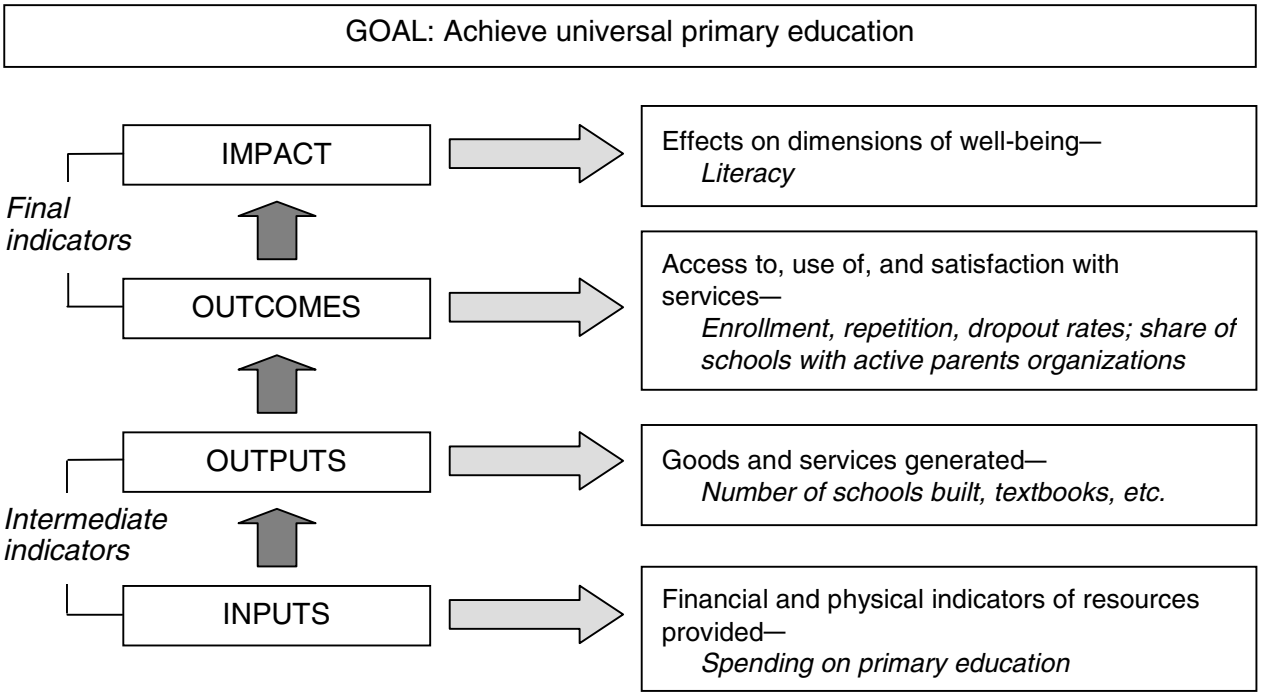
The Millennium Development Goals (MDGs) provide an example of the types of goals, indicators, and targets that can be used to monitor progress. Following various international conferences of the 1990s and the work on the International Development Goals, over 150 Heads of State gathered at the Millennium Summit in September 2000 in New York agreed on a set of goals to monitor progress in poverty reduction (box 3.1).

3.2.2 Selecting indicators

Once a set of goals has been agreed on through participatory processes, the next step is to identify indicators—also in a participatory way—to measure progress toward those goals.[1]

As shown in figure 3.1, indicators can be broadly classified into two categories: intermediate and final. When an indicator measures the effect of an intervention on individuals' well-being, we call it a "final" indicator. For example, literacy may be considered one of the dimensions of well-being, so an indicator measuring it—say, the proportion of people of a certain age who can read a simple text and write their name—would be a final indicator. Sometimes final indicators are divided into "outcome" and "impact" indicators. Impact indicators measure key dimensions of well-being such as freedom from

Figure 3.1. Types of Indicators



Box 3.1. Millennium Development Goals, Indicators, and Targets

Goals and Targets		*Indicators**
Goal 1: Eradicate extreme poverty and hunger		
Target 1:	Halve, between 1990 and 2015, the proportion of people whose income is less than one dollar a day	**1.** Proportion of population below $1 per day **2.** Poverty gap ratio (incidence x depth of poverty) **3.** Share of poorest quintile in national consumption
Target 2:	Halve, between 1990 and 2015, the proportion of people who suffer from hunger	**4.** Prevalence of underweight children (under 5 years of age) **5.** Proportion of population below minimum level of dietary energy consumption
Goal 2: Achieve universal primary education		
Target 3:	Ensure that, by 2015, children everywhere, boys and girls alike, will be able to complete a full course of primary schooling	**6.** Net enrollment ratio in primary education **7.** Proportion of pupils starting grade 1 who reach grade 5 **8.** Literacy rate of 15- to 24-year-olds
Goal 3: Promote gender equality and empower women		
Target 4:	Eliminate gender disparity in primary and secondary education preferably by 2005 and to all levels of education no later than 2015	**9.** Ratio of girls to boys in primary, secondary and tertiary education **10.** Ratio of literate females to males of 15- to 24-year-olds **11.** Share of women in wage employment in the nonagricultural sector **12.** Proportion of seats held by women in national parliament
Goal 4: Reduce child mortality		
Target 5:	Reduce by two-thirds, between 1990 and 2015, the under-5 mortality rate	**13.** Under-5 mortality rate **14.** Infant mortality rate **15.** Proportion of 1-year-old children immunized against measles
Goal 5: Improve maternal health		
Target 6:	Reduce by three-quarters, between 1990 and 2015, the maternal mortality ratio	**16.** Maternal mortality ratio **17.** Proportion of births attended by skilled health personnel
Goal 6: Combat HIV/AIDS, malaria, and other diseases		
Target 7:	Have halted by 2015, and begun to reverse, the spread of HIV/AIDS	**18.** HIV prevalence among 15- to 24-year-old pregnant women **19.** Contraceptive prevalence rate **20.** Number of children orphaned by HIV/AIDS
Target 8:	Have halted by 2015, and begun to reverse, the incidence of malaria and other major diseases	**21.** Prevalence and death rates associated with malaria **22.** Proportion of population in malaria risk areas using effective malaria prevention and treatment measures **23.** Prevalence and death rates associated with tuberculosis **24.** Proportion of TB cases detected and cured under DOTS (Directly Observed Treatment Short Course)
Goal 7: Ensure environmental sustainability		
Target 9:	Integrate the principles of sustainable development into country policies and programs and reverse the loss of environmental resources	**25.** Change in land area covered by forest **26.** Land area protected to maintain biological diversity **27.** GDP per unit of energy use (as proxy for energy efficiency) **28.** Carbon dioxide emissions (per capita) [Plus two figures of global atmospheric pollution: ozone depletion and the accumulation of global warming gases]
Target 10:	Halve, by 2015, the proportion of people without sustainable access to safe drinking water	**29.** Proportion of population with sustainable access to an improved water source
Target 11:	By 2020, to have achieved a significant improvement in the lives of at least 100 million slum dwellers	**30.** Proportion of people with access to improved sanitation **31.** Proportion of people with access to secure tenure [Urban/rural disaggregation of several of the above indicators may be relevant for monitoring improvement in the lives of slum dwellers]

* Some indicators, particularly for goal 7, remain under discussion. Additions or revisions to the list may be made in the future.

hunger, literacy, good health, empowerment, and security. Outcome indicators capture access to, use of, and satisfaction with public services, such as use of health clinics and satisfaction with the services received; access to credit; representation in political institutions and so on. These are not dimensions of well-being in themselves, but are closely related.

When an indicator measures a factor that determines an outcome or contributes to the process of achieving an outcome, we call it an "input" or "output" indicator, depending on the stage of the process—in other words, an "intermediate" indicator. For example, many things may be needed to raise literacy levels: more schools and teachers, better textbooks, and so on. A measure of public expenditures on classrooms and teachers would be an input indicator, while measures of classrooms built and teachers trained would be output indicators. What is important is that inputs and outputs are not goals in themselves; rather, they help to achieve the chosen goals.

Outputs differ from outcomes because they are fully under the control of the agency that provides them; so, for example, the number of schools built is an output, because it is directly under the control of education or other public authorities, while the number of children going to the schools is an outcome, because it depends on the behavior of children and their families. Table 3.1 illustrates goals and some of their corresponding intermediate and final indicators.

Although the main objective of the monitoring system is to track progress in poverty outcomes and impacts, both final (outcome and impact) and intermediate indicators (input and output) should be tracked.[2] Monitoring final indicators helps to judge progress toward the goals set. But final indicators are the result of several factors, many of which are outside the control of policymakers and program administrators. Intermediate indicators, on the other hand, generally change as a result of actions by the government and other agents. Moreover, final indicators generally change slowly over time, while intermediate indicators change more rapidly, giving an indication, if not on what is happening

Table 3.1. Examples of Final and Intermediate Indicators

Goal	*Intermediate indicator (input and output)*	*Final indicator (outcome and impact)*
Reduce extreme poverty and expand economic opportunities for the poor.	• Expenditure on infrastructure • Expenditure on and number of beneficiaries of job training programs • Percentage of roads in good and fair condition	• Incidence of extreme poverty: percentage of population whose consumption falls below the poverty line • Poverty gap ratio • Income/expenditure of the poorest 20 percent of the population as a share of the total income/expenditure of the whole population • Unemployment/under-employment rate • Percentage of the poor population with access to microcredit programs
Enhance the capabilities of poor men and women.	• Expenditure on primary education as a share of national income • Expenditure on primary health care as a share of national income • Percentage of schools in good physical condition • Pupil-teacher ratio • Number of doctors per 100,000 inhabitants	• Literacy rates • Learning achievement • Dropout and repetition rates • Net enrollment in primary education • Percentage of population below the poverty line with access to health care facilities • Infant, child, and under-five mortality rate • Maternal mortality rate • Malnutrition rate
Reduce the vulnerability of the poor.	• Expenditure on safety net programs • Percentage of poor households/individuals receiving transfers from the government	• Variability of household consumption • Percentage of AIDS orphans protected

with well-being, at least what is happening with some of its determinants. This can make it possible to take corrective action while a program is being implemented. Finally, information on intermediate indicators is often easier to collect (we will return to this point below when discussing sources of data).

The most useful intermediate indicators are those that refer to key determinants of impact or outcome and that vary across areas or groups or over time. For example, in a country where all schools have more or less the same teacher-to-student ratio, the teacher-to-student ratio would not be a very useful intermediate indicator to monitor differences in quality of education across regions (although it could still be useful to monitor changes over time).

Final and intermediate indicators should be complemented with other selected indicators to measure overall country performance and account for the context in which the poverty reduction strategy is being implemented. For example, indicators measuring exogenous factors that are likely to impinge on outcome indicators such as rainfall or external demand for a country's goods should be included in the monitoring system.

In general, good indicators share a number of features. Box 3.2 summarizes some of these common features.

The choice of indicators is clearly dependent on the types of data that are available in a country, as well as on what can be feasibly monitored given resource and capacity constraints; in fact, the process of selecting indicators should start from an analysis of what is available and what is feasible, and indicators that are not yet available should be included in the monitoring system only if it is realistic to set up a mechanism to collect and analyze data on such indicators.

For the intermediate and final indicators that have been selected in practice, see case studies 1 and 2, which provide examples of the indicators used to monitor the effectiveness of the poverty reduction strategy in Uganda and Tanzania.

3.2.3 Disaggregating indicators

The decision on the level of disaggregation of indicators is as important as the choice of indicators itself. These are in a sense "joint decisions" that are usually considered at the outset, based on existing data sources and on the goals that a strategy aims to achieve. Indicators can be disaggregated along various dimensions, including location, gender, income level, and social group (based on ethnicity, religion, tribe, caste). Aggregate, country-level indicators are useful, as they give an overall picture of where a country stands in comparison with others. However, aggregate indicators tend to mask significant differences across areas, gender, or social groups, and it is hard to design good policies and programs to reduce poverty without a disaggregated picture that captures these differences.

The appropriate type and level of disaggregation depend on country conditions and the indicator itself. Here are some examples.

A basic type of disaggregation is by *geographic areas* including urban/rural, administrative units and geoclimatic zones. Calculating disaggregated *urban* and *rural* indicators is common, and essential, but not always sufficient. Smaller cities often tend to be more similar to rural areas than to megacities, for example, in terms of the importance of agriculture as a source of livelihood. So it may be useful to

Box 3.2. Features of Good Indicators

A good indicator

- **is a direct and unambiguous measure of progress—more (or less) is unmistakably better;**
- **is relevant—it measures factors that reflect the objectives;**
- **varies across areas, groups, over time, and is sensitive to changes in policies, programs, and institutions;**
- **is not easily diverted by unrelated developments and cannot be easily manipulated to show achievement where none exists; and**
- **can be tracked (better if already available), is available frequently, and is not too costly to track.**

disaggregate further among urban areas by size of settlement or at least to distinguish megacities from the rest. Similarly, the capital city often tends to have different characteristics: higher average income, better availability of services, a larger share of employment in services, and so on. Thus it may be useful to construct separate indicators for the capital.

Most countries are divided into *administrative units*—states, regions, provinces, districts, municipalities, villages, and so on—and these can be used as a basis of disaggregation. Ideally, there would be indicators for each administrative level with decisionmaking power over resources, or to which resources are allocated. In practice, however, the availability of data and resource constraints will determine the lowest feasible level of disaggregation.

A third type of geographic disaggregation is by *geoclimatic zones*. Most countries have a number of geographic zones characterized by different soils, rainfall, topography, and, consequently, different agricultural practices, settlement patterns, ease of access, and so on.

Another basic type of disaggregation is by *gender.* Appropriate gender indicators measure factors that vary by gender and take into account the impact of biological differences. For example, life expectancy tends to be higher for women, so a lower life expectancy for women than for men is usually an indication that women may be suffering severe health risks at childbirth. See chapter 10, "Gender," for more information.

Disaggregating by *income, consumption,* or *asset ownership* level is a common way to see how indicators vary across the population. It is usually preferable to a simple poor–nonpoor disaggregation, as it captures the fact that many household and individual characteristics vary along a continuum. There are often significant differences among those classified as poor, and those just below the poverty line generally have very similar characteristics to those just above it. So it is desirable to divide the population into groups of equal size rather than simply into poor and nonpoor. Some commonly used groupings based on income and consumption level are the following:

Name	Number of groups	Share of the population (percentage)
Deciles	10	10
Quintiles	5	20
Quartiles	4	25
n^{th} percentile	n	100/n

Disaggregating indicators by, for example, quintiles is important to monitor whether improvements reach the worse-off as well as the better-off. Nationwide average targets, such as those of the MDGs, can often be reached with different degrees of improvement for different groups.[3] If improving the well-being of the poorest is important, then tracking indicators disaggregated by quintile is essential.

In most countries there are significant differences across *socially defined groups,* whether along ethnic, tribal, religious, or other lines. The definition of the relevant groups will naturally vary across countries.

Finally, it is important to recognize that disaggregating indicators by areas, groups, and the like usually has political consequences and must be done carefully. Furthermore, monitoring indicators disaggregated by administrative area almost always requires complementary efforts to build capacity for monitoring and analysis in the decentralized administrative units, a point highlighted in case study C.1 on Uganda.

3.2.4 Setting targets

Once indicators are selected, it is useful to assess baseline values and set quantitative targets for at least some of them. Baseline values can be obtained from existing data, if they are of reasonable quality and not too old.[4] Where data for an indicator do not yet exist, the first available estimate, if it comes within a reasonable amount of time, or a preliminary estimate subject to revisions, can be used as the baseline.

Setting targets is a complex task. We offer some general guidelines here; additional guidance on the technical aspects of setting targets for different indicators can be found in chapter 4, "Development Targets and Costs."

First, targets should be selected on the basis of the current situation and what is attainable in a given country at a given time. Even if a country chooses goals consistent with the MDGs (see box 3. 1), the indicators and targets selected may not be the same. The target of achieving universal primary school enrollment obviously is not relevant for a country where this has already been achieved.

Second, targets may be set at different levels of disaggregation. In addition to national-level targets, specific targets can be set for certain regions or groups. For example, for most countries, educational targets are not very useful unless they are differentiated by gender, and for large countries such as Brazil and India, geographic targets make good sense.

Third, the inclusion of qualitative and subjective factors in goal setting is important. Many factors that affect quality of life cannot be easily quantified but are not for this reason less important. Where feasible, qualitative and subjective indicators could be added—for example, whether or not people perceive themselves as being poor.

Fourth, as a general rule, improvements become more difficult as levels improve. For example, it is generally more difficult to reduce income poverty from 10 percent to 0 than from 40 percent to 30 percent, because the target group generally becomes more difficult to reach.

Fifth, if a particular indicator has continuously worsened in the recent past, it may not be realistic to set a target indicating a substantial improvement in the short term. Most likely, it will take some time for that indicator to stabilize and start improving.

Finally, it is essential to consider the resource implications of the selected targets and their feasibility. Resources may have to be shifted from some sectors and programs toward activities that are in line with the selected targets. See chapter 4, "Development Targets and Costs" for a more detailed description of the costing of targets.

Figure 3.2 summarizes the steps involved in selecting indicators and setting targets and points to documents providing guidance on each step.

3.2.5 Determining data requirements

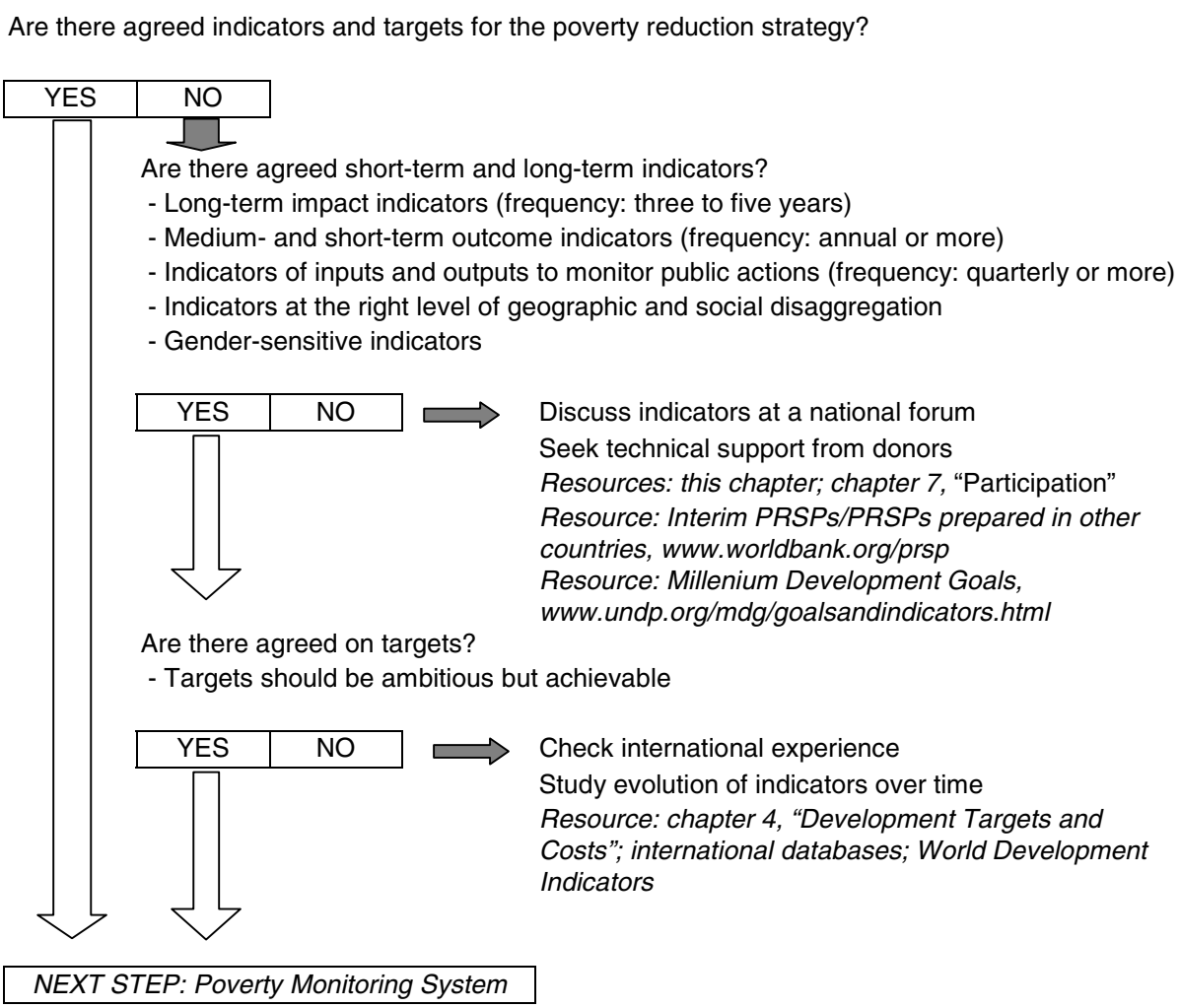
As mentioned, both intermediate and final indicators should be tracked. So a good poverty monitoring system would include data on both categories of indicators. These would be collected through a number of different instruments and by different agencies. This last point is important: the fact that a good poverty monitoring system requires data on different indicators does not mean that one agency needs to be in charge of all data collection, which would be neither desirable nor efficient.

Data on intermediate indicators are usually collected by the treasury or finance ministry and sectoral ministries at the central and local level through financial and management information systems. These systems collect data on public expenditures in various sectors and on activities and outputs produced by such expenditures. For example, the treasury or finance ministry will collect data on expenditures in education, while the education ministry will have data on schools built, textbooks purchased, scholarships provided, training activities, and so on. Data from administrative records usually exist in countries, although there may be problems with their accuracy, timeliness, and comprehensiveness. Data on the number of staff in key sectors come from sectoral ministries or the ministry in charge of public administration.

Information on outcome and impact indicators normally needs to be collected from beneficiaries through household or individual surveys and participatory methods. Because of the need to collect information directly from households and individuals, outcome and impact data are costlier to collect and require more time. Particular attention is needed to obtain reliable information from women and possibly other groups, such as children, the elderly, or excluded minorities, who may not be easily reached or feel comfortable responding to interviewers.

Why is it necessary to collect data on access to and use of services from households in addition to using data from administrative records? Why, for example, are household surveys needed to determine how many children are attending school? Why are enrollment data from the management information

Figure 3.2. Selecting Indicators and Setting Targets



systems (MIS) from the education ministry not enough? First, data collected from households are more reliable: households have fewer incentives to report school attendance incorrectly than program administrators and local officials, whose budget allocations and incentives may depend on achieving enrollment targets. Second, household surveys and participatory studies generally collect other information from households, such as income or consumption, education status of the parents and employment status, or reasons not to attend school; this additional information makes it possible to analyze the causes of trends in enrollment rates. This is not to say that MIS data on use of services are not useful, only that they should be checked against and complemented by information collected directly from households.

A good monitoring system should also include data on external factors that may influence the effectiveness of the poverty reduction strategy, such as weather or external market factors. Table 3.2 summarizes collection instruments, agencies usually responsible, and the level of disaggregation for different indicators.

For a more detailed discussion of various data collection instruments, see chapter 1, "Poverty Measurement and Analysis," and chapter 5, "Strengthening Statistical Systems."

Note that data from these various sources are complementary, not substitutes for one another. Having very good household-level data on consumption and incomes will not be sufficient to understand trends in poverty outcomes; accurate and timely data on public expenditures and public services are needed as well. The increased attention that poverty reduction strategies place on final indicators should not reduce attention to intermediate indicators, or shift resources away from tracking them.

Table 3.2. Data for Monitoring and Sources

Type	*Indicator*	*Instrument*	*Agency*	*Level*
Input	Public finance data: revenues, expenditures by category Human resources	Budget documents; actual expenditure data Expenditure tracking surveys Payroll data	Ministries of finance and planning and public administration; sectoral ministries; public accounting and auditing agencies	National and various subnational administrative levels
Output	Outputs of public expenditures: infrastructure, services provided	Administrative and management information systems Community surveys	Sectoral ministries; project implementation units; local administrations and local service providers	National and various subnational administrative levels; facilities (schools, clinics, etc.)
Outcome	Access to, use of, and satisfaction with services	Priority and quick monitoring surveys; multi-topic household surveys; qualitative studies	Central statistical agency; local service providers; others	Households and individuals; facilities (schools, clinics, etc.); communities
Outcome/ Impact	Household consumption and income; living conditions; social indicators; household priorities; perceptions of well-being	Household budget/ expenditure/ income surveys; single-topic surveys (for example, labor force surveys); multi-topic household surveys (such as Living Standard Measurement Surveys and Demographic and Health Surveys); qualitative studies	Central statistical agency	Households and individuals; communities
Other	National accounts: gross domestic product, consumption, investment, exports, imports, etc. Consumer and producer prices	System of national accounts, trade statistics	Central statistical agency; central bank	National (largest subnational levels in some cases)
Other	Climatic data: temperature, rainfall, water flows, etc.	Direct measurement	National weather agency; others	As detailed as possible

3.2.6 Determining the frequency of monitoring

The decision on how frequently a given indicator needs to be monitored depends on a careful assessment of the tradeoff between the desirability of recent data and the cost of collection, much like the decisions on which indicators to track and at what level of disaggregation. Data on input indicators, such as public expenditures, are tracked at least annually and, in most cases, more often (monthly or quarterly) as part of budget tracking mechanisms. Data on outputs are most often available on an annual basis, but it is highly desirable to have information on key outputs midway through the budget year to inform midcourse corrections and decisions on budget allocations for the following year. Data on some outcome indicators should also be available annually. Data on impacts, on the other hand, are usually not available annually, both because it is costly to collect and analyze household survey and participatory data and impact indicators do not usually change rapidly.

Table 3.3 indicates the desirable frequency of collection for the various indicators listed in the previous table.[5]

3.2.7 Elements of poverty monitoring systems that often need attention

Most countries already have monitoring systems in place to track most, if not all, the indicators needed to monitor the effectiveness of poverty reduction strategies. So what more needs to be done? Recent experience in countries that are developing and implementing poverty reduction strategies points to the need to devote attention early on to some key elements of the system.

Table 3.3. Frequency of Data Collection

Type	*Indicator*	*Instrument*	*Frequency*
Input	Public finance data: revenues, expenditures by category	Budget documents; actual expenditure data	Monthly or quarterly where possible; at least yearly
	Human resources	Expenditure tracking surveys Payroll data	
Output	Outputs of public expenditures: infrastructure, services provided	Administrative and management information systems Community surveys	Possibly every six months; at least yearly
Outcome	Access to, use of, and satisfaction with services	Priority and quick-monitoring surveys; multi-topic household surveys; qualitative studies	Yearly where possible
Outcome/ Impact	Household consumption and income; living conditions; social indicators; household priorities; perceptions of well-being	Household budget/ expenditure/income surveys; multi-topic household surveys; qualitative studies	Every three to five years
Other	National accounts: Gross domestic product, consumption, investment, exports, imports, etc.	System of national accounts, trade statistics	Monthly or quarterly where possible (trade statistics, for example); at least yearly
	Consumer and producer prices		Monthly or quarterly price collection; consumer price index basket updated at least every five years
Other	Climatic data: temperature, rainfall, water flows, etc.	Direct measurement	Daily where possible

Frequent problems in tracking intermediate indicators are the following:

- ***Actual expenditure data are not timely.*** In many countries actual expenditure data are available only with a significant time lag. This is less problematic for recurrent expenditures (especially salary, but also nonsalary), where actual expenditures are often fairly close to budgeted amounts, but can seriously limit a country's ability to track capital expenditures that are often quite different from budgeted amounts. Programs to improve expenditure tracking at the central and decentralized levels—for example, through the establishment of well-designed reporting formats and computerization—can improve the timeliness of expenditure data.[6]
- ***Input data (expenditures and human resources) cannot be easily related to outputs, so it is hard to estimate the cost of providing services.*** For example, a large share of expenditures in education is for "general administration," and it is not clear how much of this supports primary versus secondary or tertiary education. So the cost of providing, for example, a year of schooling to a primary school child cannot be estimated accurately. Solving this problem requires moving towards activity-based costing, where all expenditures are related to specific activities and outputs. This is done extensively only in a small number of countries, but in most countries there is scope to move in this direction.[7]
- ***Disaggregated spending data are unavailable or inaccurate.*** Without data disaggregated at the level of the facilities or agencies that provide services, it is hard to assess whether public funds reach the facilities or not. Where local government accounts are not available or are of poor quality, expenditure tracking surveys can be conducted. In Uganda, spending data for 1991–95 collected from a random sample of public schools revealed that less than 30 percent of the funds intended for nonsalary public spending actually reached schools because district administrations kept and used the rest of the funds. This finding led to the decision to inform the public on allocations and to implement changes in spending procedures. The survey instruments and methodologies used are available and can be applied elsewhere.[8]

In tracking outcomes and impact, other issues have emerged:

- ***It takes a long time to process data from household surveys and make them available for analysis.*** Data entry, cleaning, and organization often take years. This need not be: there are ways to shorten the process considerably. For example, data entry can be carried out in the field or in de-

centralized field offices concurrently with data collection; there are even experiments to eliminate paper questionnaires completely and enter data directly on disk. Data cleaning can be speeded up considerably by using precoded questionnaires and data entry programs that identify entry errors and inconsistencies between variables (for example, a mother who is younger than one of her children). Moreover, when data entry takes place while in the field, errors can be corrected through recall or re-interviewing.[9]

- ***There is a need to introduce quick monitoring tools to gather information from households on an annual (or more frequent) basis.*** Even when data from household surveys are processed and made available quickly, these surveys still take time to conduct (especially if data are collected over the course of a year to capture seasonal patterns) and may be too costly to be conducted every year. How can changes in household and individual well-being be tracked more frequently? There are now quick-monitoring tools that have been tested in different countries and can be applied fairly easily—the Core Welfare Indicators Questionnaire (CWIQ) is a good example (see box 3.3). Other examples are the citizen scorecards piloted in Bangalore, India (see case study C.3) and the user surveys piloted in Uganda that complemented the expenditure tracking surveys cited above.[10]

3.3 Designing Impact Evaluations

Poverty monitoring provides crucial information to assess overall progress in achieving poverty reduction goals and to understand changes over time and space. However, complementary tools such as impact evaluations are required to inform policymakers and the public on which public actions have been effective and which ones have not worked so well in reducing poverty. An impact evaluation assesses the changes in well-being that can be attributed to a particular program or policy. Information generated by impact evaluations informs decisions on whether to expand, modify, or eliminate a particular policy or program and is used in prioritizing public actions. It is a decisionmaking tool for policymakers and increases public scrutiny of programs.

There are other types of evaluations such as process evaluation and theory-based evaluations that are also important for improving management performance and should be conducted depending on the evaluation question at hand (see technical note C.1). However, it is important to note that these evaluations do not estimate the magnitude of effects and assign causation. Such a causal analysis is essential for understanding the effectiveness of alternative program interventions in reducing poverty and thus for designing appropriate poverty reduction strategies.

Some of the questions addressed in impact evaluations are the following:

- Do key policies/programs in the poverty reduction strategy achieve the intended goal?
- Can the changes in poverty outcomes be explained by those programs, or are they the result of some other intervening factors occurring simultaneously?
- Do key program impacts vary across different groups of intended beneficiaries (males, females, indigenous people), regions, and over time? If so, what are the cultural, economic, and political factors that limit the full participation of women or other vulnerable groups in the program benefits?
- Are there any unintended effects, either positive or negative?
- How effective are key programs in comparison with alternative interventions?
- Are key programs worth the resources they cost?

The first step is to decide what policies and programs should be evaluated. Designing an impact evaluation then involves defining the expected outcomes and their timeframe, selecting an evaluation design and obtaining the data needed. As with the monitoring system, impact evaluations also require a well-established feedback mechanism into policymaking and a clearly defined institutional framework. These issues will be covered in section 3.5.

Box 3.3. The Core Welfare Indicators Questionnaire

A number of countries in Africa (for example, Ghana and Tanzania) have started using a new survey tool, the CWIQ, for monitoring outputs and outcomes in the context of poverty reduction strategies. The CWIQ is a household survey designed to provide very rapid feedback through the tracking of leading indicators and can show who is and who is not benefiting from programs and policies. It focuses on simple indicators of usage, access, and satisfaction.

The CWIQ is a ready-made survey package that national statistical offices can implement on an annual basis and can supplement, when necessary, with special modules. It is meant to complement other surveys. It is designed to be administered to large samples of households, so that results can be disaggregated to relatively low levels, and to be repeated annually, so that time-series can be quickly built up. The standard output tables and graphs present access, usage, and satisfaction indicators broken down by geographic and socioeconomic groupings.

The CWIQ does not collect information on consumption or income, which cannot be done accurately using a short questionnaire, but can collect information on indicators that are related to economic well-being, such as consumption of certain goods or ownership of assets. A recent multi-topic or budget survey is usually used to identify core indicators that are easy to monitor and correlated with consumption or income; if such a survey is not available, information from a participatory poverty assessment can be used, as was done for the first pilot in Ghana. The CWIQ can include up to 10 such indicators, and these can be used as proxy indicators to track changes in consumption/income and income poverty.

3.3.1 Deciding when to conduct an impact evaluation

Impact evaluations should be conducted only for a selected set of interventions (section 3.4 includes a brief discussion on the evaluation of overall poverty reduction strategies). Impact evaluations can be demanding activities in terms of analytical capacity and resources. Therefore, it is very important that they are conducted only when the characteristics of the intervention warrant an impact evaluation. There are other less rigorous and capacity-intensive evaluation methodologies that should be considered when measuring the magnitude of program effects, and assign causation is not a first priority. The selection of programs and policies for an impact evaluation should be done so as to maximize the learning from current poverty reduction efforts and inform program and policy choices. Since donors are often interested in supporting impact evaluations, countries should carefully explore the possibility of getting and coordinating technical and financial support. Three questions can help guide the decision of when to conduct an impact evaluation.

First, is the policy or program considered to be of strategic relevance for poverty reduction? Policies and programs expected to have the highest poverty impacts may be evaluated to ensure that the poverty reduction strategy is on the right track and allow for any necessary corrections. For example, in a poor agrarian economy, expansion of agricultural technology and improvement of grain production may be critical for household and food security as well as for poverty reduction. An evaluation of policies or programs to expand food production and productivity would then become a high-priority task. Likewise, an evaluation of active labor market programs and public works may be critical for a country that has high unemployment and is emerging from a serious financial crisis.

Second, will the evaluation of a particular policy or program contribute to filling in knowledge gaps of what works and what does not in poverty reduction? If knowledge gaps exist about what works best to reduce poverty, an impact evaluation is well justified. For example, despite a widespread belief in the importance of rural roads in alleviating poverty, little hard evidence exists on the nature and magnitude of their impact. This knowledge gap has prompted an evaluation of a World Bank-financed rural transport project in Vietnam.

Third, is the policy or program testing an innovative approach to poverty reduction? Impact evaluations can help to test pioneering approaches and decide whether they should be expanded and pursued on a larger scale. Hence, the innovative character of policies or programs also provides a strong reason to evaluate. For example, Morocco is evaluating the impact of an innovative nonformal school program to see whether nonformal schools are suitable alternatives to other basic educational services. One important caveat, however, is that fruitful evaluations require sufficiently mature programs. Although programs may be testing innovative approaches, they need clearly defined objectives and well-delineated activities, as well as a stable institutional framework for implementation.

3.3.2 Measuring the impacts of policies and programs

To evaluate a program or policy, it is first necessary to understand the nature of the welfare benefits that it is expected to generate. This, of course, depends on the type of intervention and its objectives. Some interventions may have set multiple objectives. In this case, it is best to focus the evaluation on a few key objectives. Equally important is the need to be clear about the time within which welfare changes are to be expected. Some policies or programs may only realize their full effects in the longer term. In such instances, indicators of shorter term outcomes may be needed to form a judgment on the direction and speed of realization of the intervention's objective. For example, it may take several years to observe changes in the cognitive development of young children resulting from early childhood development programs. Hence, in the shorter term, the evaluation may focus on measuring the effect of the program on child-rearing practices of caregivers rather than on cognitive development. Additional examples of interventions follow:

Intervention	Impacts	Timeframe	Shorter term outcomes
Public works program	Consumption gains	Immediate	–
Nutrition intervention	Improved nutritional status of children (weight-for-age)	Medium term	Improved caloric intake
Early childhood development	Improved health, nutrition, and cognitive development of young children	Medium and long term	Improved child-rearing practices

Choosing an appropriate evaluation design

Evaluating the impact of a policy or program hinges on asking the fundamental question: What would the situation have been if the intervention had not taken place? Although one obviously cannot observe such a situation, it is possible to approximate it by constructing an appropriate counterfactual, which is a hypothetical situation that tries to depict the welfare levels of individuals in the absence of a policy or program. How a counterfactual is constructed or visualized depends on a number of factors, including program coverage.

For partial-coverage programs, counterfactuals are simulated by comparing program participants (the treatment group) with a control or comparison group. The control or comparison group is made up of individuals (or other unit of analysis, such as households, schools, organizations) that have the same characteristics as program beneficiaries, especially with respect to those characteristics that are relevant to program participation and program outcomes, but do not participate in the program being evaluated.

The key issue when evaluating the impact of partial-coverage programs is how to select or identify nonparticipants. The group can either be selected randomly through a process similar to a lottery or be constructed using special statistical techniques. The nonparticipant group is called a control group when its members are randomly selected; otherwise, it is called a comparison group. The choice of method to identify the group of nonparticipants determines the evaluation design, which can be broadly classified into three categories: experimental, quasi-experimental, and nonexperimental. These evaluation designs vary in feasibility, cost, and the degree of clarity and validity of results. Technical note C.2 describes them in greater detail and discusses their advantages and limitations.

In some situations it is not possible to have a group of individuals from which the intervention is withheld. For example, there is no scope for control or comparison groups in a nationwide school lunch program. For this type of intervention (full-coverage interventions), the same evaluation question applies—what would the situation be without the policy or program?—but the methodology to answer it is different.

Evaluations of full-coverage interventions rely mostly on comparing the situation of the relevant population group before and after the program. This is a quasi-experimental methodology called reflexive comparison (see technical note C.2). Additional methods to evaluate full-coverage interventions include simulations using computable general equilibrium (CGE) models, comparisons of countries with and without the program, and statistical controls. These methods are further discussed in technical note C.3.

3.3.3 Determining data requirements

Household data are probably the most widely used in impact evaluation. In some instances, data at other levels of disaggregation are desirable. To assess the impact of an intervention on particular members of the household (for example, women and children), it is necessary to collect data at the individual level.

Ideally, data for impact evaluation would be collected from the same set of households at least two times, before and after the intervention.[11] Nonetheless, it is important to distinguish between desirability and feasibility. The existing information base and time and resource constraints are key factors to be considered when deciding which data sources to use. If only postintervention data are available, it is still possible to conduct a sound evaluation by choosing an appropriate evaluation design. Technical note C.4 describes different types of data sources for impact evaluation, their advantages, and their shortcomings.

Quantitative and qualitative methods for data collection

The validity of evaluation results depends in large part on the adequacy and reliability of the data. Hence, it is important to use different sources of data collected through quantitative as well as qualitative methods. In general, qualitative methods are aimed at studying selected issues, cases, or events in depth by gathering information on people's attitudes, preferences, and perceptions; data collection is not constrained by predetermined standardized formats or categories of analysis. By contrast, quantitative methods typically rely on random sampling and structured data collection instruments that fit diverse experiences into predetermined response categories (for example, Living Standards Measurement Surveys (LSMS)-type surveys). Although the two approaches differ substantially in their objectives and characteristics (see table 3.4), they are highly complementary. Quantitative methods produce results that are easy to summarize, compare, and generalize, while the qualitative approach provides in-depth and detailed data that can be useful in understanding the processes behind observed results and assessing changes in people's perceptions of their well-being. Examples of evaluations using a combined quantitative and qualitative approach can be found in case study C.6.

Gender analysis is one of the areas where a combination of quantitative and qualitative methods will frequently be required. In many cultures, it is more difficult to obtain reliable information from or about women using conventional quantitative survey methods, and it will often be necessary to use qualitative data collection methods such as focus groups, participant observation, use of drawings, or pictures to describe how women spend their time, and so on. For a detailed discussion of qualitative methods and how they can be used in gender analysis, see chapter 10, "Gender."

Table 3.4. Comparison of Quantitative and Qualitative Approaches for Evaluation

Aspect	*Quantitative approach*	*Qualitative approach*
Objectives	• To assess causality and reach conclusions that can be generalized	• To understand processes, behaviors, and conditions as perceived by the groups or individuals being studied
Data collection instrument	• Structured, formal, predesigned questionnaires	• In-depth, open-ended interviews • Direct observation • Written documents (for example, open-ended written items on questionnaires, personal diaries, program records)
Sampling	• Probability sampling	• Purposive sampling
Methodology for analysis	• Predominantly statistical analysis	• Triangulation (simultaneous use of several different sources and means of gathering information) • Systematic content analysis • Gradual aggregation of data based on selected themes

Source: Adapted from Carvalho and White (1997) and Baker (2000).

Table 3.5. Evaluation Methods and Data Requirements

Evaluation design	Data requirement		Use of qualitative approach
	Minimal	Ideal	
Experimental	Single cross-section data of treatment and control group	Panel data on both treatment and control group	• Inform design of survey instrument, sampling • Identify indicators • Data collection and recording using textual data, informal or semi-structured interviews, focus groups or community meetings, direct observation, participatory methods, photographs • Triangulation • Data analysis
Quasi-experimental			
Matching comparison	National cross-section (census, national budget or LSMS-type survey) and oversampling of program participants	National survey and smaller project-based household survey, both with two points in time	
Reflexive comparison	Baseline and follow-up data on program participants	Time series or panel studies that collect data for several years before and after the program	
Nonexperimental	Cross-section data representative of the whole population with corresponding instrumental variables	Cross-section and time series representative of both the beneficiary and nonbeneficiary population with corresponding instruments	

Source: Adapted from Baker 2000.

Linking data requirements to evaluation methods

Data needs depend on the kinds of outcomes to be measured and the type of evaluation design that will be implemented. Since programs selected for evaluation will look at a range of indicators and will require different evaluation designs, data requirements will also differ.

On the one hand, data needs depend on evaluation design (see table 3.5). On the other hand, the choice of evaluation methodology is determined by the type of intervention to be evaluated (full or partial coverage); the desired level of reliability of results; time and resource constraints; and data availability.

Conducting an impact evaluation may seem a daunting task given the informational and analytical requirements. However, it is important to emphasize that the choice of evaluation design can accommodate time and resource constraints, and that the evaluation strategy should be tailored to in-country capacity. If in-country capacity is limited, the number and frequency of evaluations can be gradually scaled up as capacity constraints are eased.

3.3.4 Obtaining data

Data collection can be both expensive and time consuming. Thus the main challenge is how to take advantage of existing data sources and how to plan additional data collection to maximize its use for both impact evaluation and outcome monitoring.

Impact evaluations can draw on a variety of data sources, including surveys, administrative records, and management information systems (see box 3.4 and chapter 1, "Poverty Measurement and Analysis," and chapter 5, "Strengthening Statistical Systems"). Hence, one of the early steps in designing an evaluation strategy is to take stock of different types and quality of data already available. Some of the data used for poverty monitoring and analysis are likely to be useful for impact evaluation.

If the existing data are insufficient, the next step is to find out whether there are any planned or ongoing data collection efforts. Surveys or other data collection instruments that are at a planning or early implementation stage can be adapted to provide information for evaluation by oversampling in the program areas or by introducing additional modules on issues related to the evaluation. Oversampling involves increasing the sample of the population surveyed to include enough individuals (or other unit of analysis) with a particular characteristic, such as being a program participant. For example, the evaluation of the Trabajar program in Argentina piggybacked on a national survey that was already in progress by oversampling program participants (see case study C.4). The use of this alternative, however,

Box 3.4. Examples of Sources of Data for Evaluation

- **Household income and expenditure surveys**
- **Living Standards Measurement Surveys**
- **Demographic and Health Surveys (DHS)**
- **National census**
- **Labor market surveys**
- **Records of cooperatives, credit unions, and other financial institutions**
- **Administrative records (for example, school records on attendance, repetition, examination performance; or public health records on incidence of infectious diseases, number of women seeking advice on contraception)**
- **Specialized surveys conducted by universities, nongovernmental organizations (NGOs), consulting groups**
- **Monitoring data from program administrators**
- **Project case studies**

Source: Adapted from Baker 2000.

may be limited by the timing of the existing data collection and the degree of flexibility in the design of the data collection instrument.

Some evaluations will require the collection of new data. If this is the case, it is important to be aware of the additional institutional capacity and other resources demanded by the data collection task. Where data needs are paramount and institutional capacity is weak, it is important to coordinate efforts across institutions, both public and nonpublic, to design instruments that collect information that is useful for as many purposes as possible. One example of this is the Panel Data Initiative in Africa (see box 3.5). Section 3.5 further discusses the issue of institutional capacity for evaluation.

In conclusion, figure 3.3 summarizes the steps to be taken in designing an evaluation system.

3.4 Challenges Ahead for Monitoring and Evaluation

3.4.1 Assessing the process of formulation and implementation of poverty reduction strategies

The main objective of a poverty reduction strategy is to reduce poverty, and this chapter has focused on monitoring progress in achieving poverty reduction goals and evaluating the poverty impact of interventions that are part of the strategy. But the *process* of formulating and implementing a poverty reduction strategy also seeks to achieve several objectives: increase country ownership; foster through deeper participation the partnership between the government and civil society, on one hand, and between the government and donors on the other hand; take a long-term, comprehensive approach to poverty reduction. It would be important to monitor these objectives and assess whether they are met.

The steps described in section 3.2 to set up a poverty monitoring system apply equally to setting up a system to monitor progress towards process objectives. Agreement is needed on the objectives to achieve, and on the indicators to be used. Objectives and indicators should be selected in a participatory manner. Indicators could refer to inputs and outputs of the process as well as to outcomes; for example, the following indicators have been suggested to monitor participation in the preparation of a Poverty Reduction Strategy Paper (PRSP):[12]

- ***Input.*** Public resources used to increase quality and scope of participation.

Box 3.5. Impact Evaluation in the Africa Region: A Cross-Sectoral Initiative

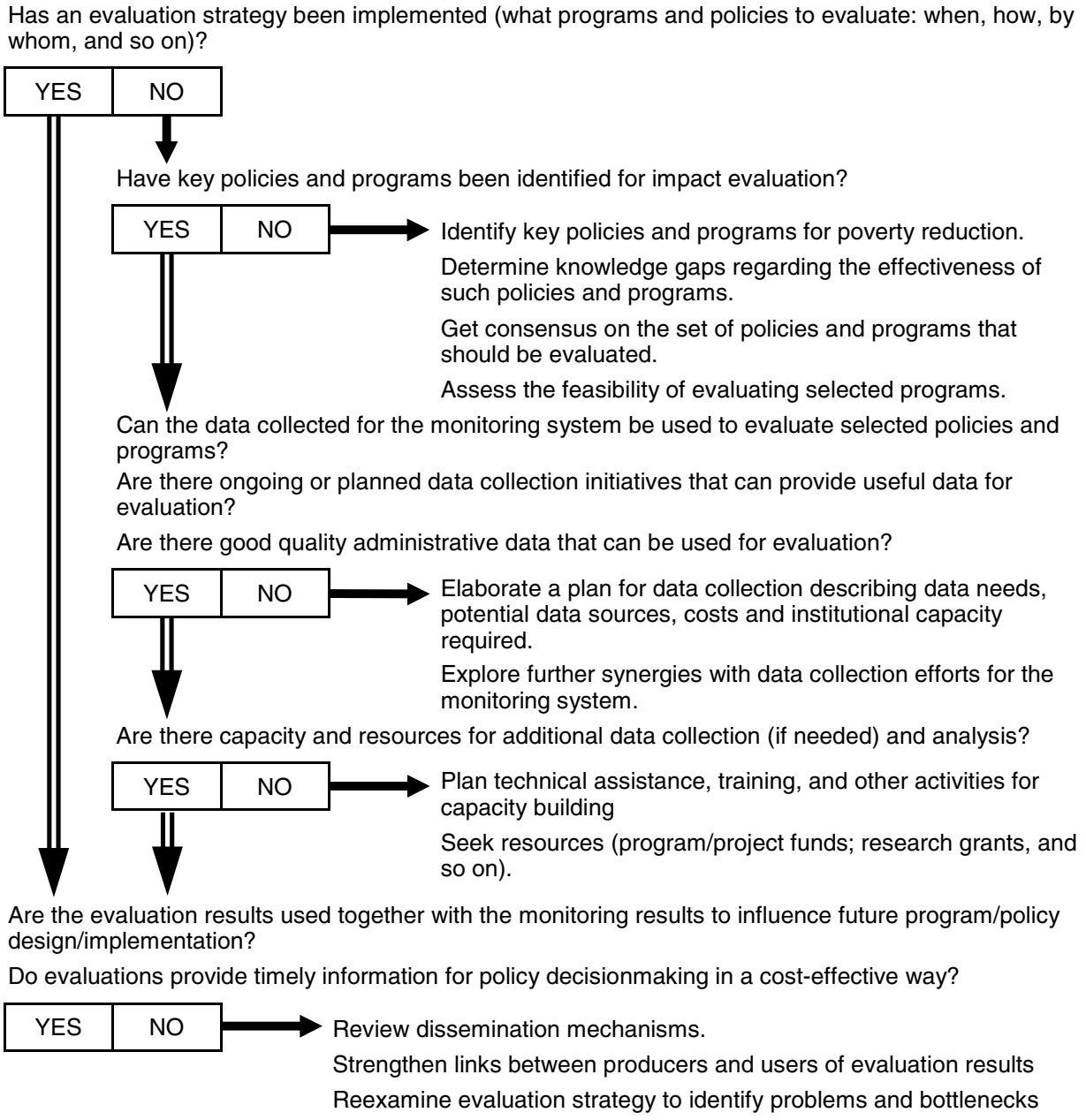
The Panel Data Initiative aims at improving data collection and analysis in several African countries by creating sustained partnerships with African research centers and building capacity as well as consensus on the importance of program evaluation. Given the desirability of panel data for impact evaluation, this initiative will use existing quality household surveys as baselines and develop panel data sets that will be available to researchers.

Data obtained through this initiative will be used to evaluate the impact of policy changes (structural adjustment and sectoral policies), investment programs (national, regional, and community based), as well as exogenous shocks (drought, AIDS, civil strife, and commodity price cycles) on household welfare. In particular, this initiative will provide information on variables such as nutritional status, income levels, and productivity. Quantitative survey data will be complemented with qualitative data for a subset of samples.

- ***Output.*** Measures of the extent to which meaningful participatory arenas (that include all stakeholders who want to participate) have been opened across the country to discuss the design, implementation, and monitoring and evaluation of a PRSP.
- ***Outcome***. Measures of the extent to which the PRSP takes into account the needs and priorities of key stakeholders, including poor people; civil society and government have a higher capacity to decide on the country's poverty reduction strategy and more opportunities to negotiate with donors and creditors over it.

Where appropriate, indicators should be disaggregated by gender, geographic area, social group, and so forth (for example, the number of participatory meetings held could be disaggregated by area; participation of women could be tracked separately) and, whenever possible, should be specified precisely.

Figure 3.3. Strengthening Impact Evaluation



Source: Authors.

Where data exist on those indicators that can be quantified, it may be useful to identify initial (baseline) values and define targets. For example, baseline values for participation indicators could reflect the situation before the PRSP process is initiated. Where data do not exist, as will often be the case with process indicators, a system to collect and analyze the needed data would have to be set in place. As for indicators in general, what is desirable may not be feasible or affordable, so the final decision on what to monitor, with what instruments and what frequency, will be influenced by available resources. Moreover, in many cases process indicators may be qualitative in nature and not quantifiable.

The process of selecting indicators and monitoring the process of formulating a poverty reduction strategy offers a real opportunity to foster partnership between the government, civil society organizations, and donors. It also is a learning opportunity, as most of the experience so far in assessing process objectives has been gained at the microeconomic level (projects and programs) rather than at the macroeconomic level (strategy).

3.4.2 Evaluating the overall poverty impact of poverty reduction strategies

After a few years of implementation of a poverty reduction strategy, the question of whether the strategy as a whole (rather than specific interventions within it) has been effective in reducing poverty may arise. Evaluating the poverty impact of the entire strategy poses a tremendous challenge, since it requires an evaluation framework that considers a large number of economic and institutional changes occurring simultaneously and can sort out the causal relationships between actions. One possible approach is to use methodologies similar to those for evaluating the poverty impact of countrywide, or full-coverage interventions: comparing the situation before and after implementation of the strategy using time series (see reflexive comparison in technical note C.2); simulating the situation without the strategy using CGE models; and comparing countries with different strategies through regression analysis and other methods (see technical note C.3). For indicators of poverty that capture empowerment and security dimensions, participatory methods may be more appropriate. Experience is limited and much remains to be learned.

Because of the complexity of such overall evaluation exercises and the capacity and resources they require, countries are not expected to carry them out. Moreover, given that the poverty impacts of a strategy may only be observed several years after the start of implementation—as noted, it takes time for policies and programs to affect well-being—it is not advisable to evaluate the overall poverty impact of a poverty reduction strategy within the three-year time frame of a PRSP. Within this timeframe, it is possible to assess the *process* of formulating and implementing the strategy (as discussed in the previous section), monitor outcomes, and carry out other types of evaluation, including qualitative and participatory assessments that examine the links between the inputs and processes of the strategy and any outcomes observable within the three-year time frame (see technical note C.1). What is most important in the short and medium term is to set up a solid monitoring system: without the basic information collected through the monitoring system, no evaluation exercise can be carried out.

3.5 Strengthening Monitoring and Evaluation Capacity and Feedback Mechanisms

3.5.1 Strengthening capacity

Poverty monitoring and impact evaluation activities involve the participation of several agencies both inside and outside the government, each with their own role. Within the government, central ministries such as finance and planning usually have a large role in designing the overall monitoring and evaluation strategy, monitoring its implementation, and using the results, as well as providing key data on expenditures; sectoral ministries usually provide data on outputs; the central statistical agency is usually responsible for the collection of data from households and individuals. Agencies and institutions outside the government, such as research centers, universities, and NGOs, often also collect and analyze information. Donors can provide technical assistance to strengthen capacity. Box 3.6 summarizes these roles.

Strong country demand at all levels is generally the main precondition for the development of a national M&E system. Sustainable capacity is usually built up if governments and civil society are truly committed to measuring the outcomes and impact of public action and to using this information to achieve better results. Thus the participatory processes followed in designing poverty reduction strategies can be critical in creating a strong demand for monitoring and evaluation.

Donors can contribute to create demand for M&E activities through the requirements of their assistance. For example, the International Monetary Fund (IMF) and the World Bank, under the PRSP approach, require as one of the conditions associated with the provision of concessional assistance and debt relief that governments prepare an annual progress report on the implementation of the poverty reduction strategy. This annual report would discuss actions taken and changes in those indicators that are tracked annually; if annual targets were set, the report would discuss whether they were attained and indicate the reasons for any differences between actual values and targets.[13] While such donor requirements do create demand for monitoring and evaluation, sustainable capacity will be built only if there is strong in-country demand.

Once there is a strong country demand for monitoring and evaluation, feasible options to build capacity vary across countries depending on local circumstances and opportunities, the actors involved, the institutional framework, and the distribution of existing capacity across agencies.[14] An important consideration is that it may be appropriate to gradually scale up monitoring and evaluation activities. Experience suggests that it may be better to put in place a few mechanisms that can be implemented immediately rather than start with the design and development of a comprehensive or very sophisticated setup. A first step can be to take stock of existing M&E capabilities and activities among central and line ministries, local governments, national statistical agencies, and other organizations such as universities and NGOs. On the basis of this assessment, various alternatives can be implemented to ease capacity constraints and develop local skills, including the following:

- Establish partnerships to collect and analyze data and provide training on skills relevant to monitoring and evaluation. Potential partners are universities, research institutions, NGOs, consulting firms, and development agencies. Collaboration with these institutions can take several forms, including carrying out joint evaluations, providing grants for the professional development of monitoring and evaluation specialists, and contracting out survey implementation.
- Disseminate national and international lessons about experience in monitoring and evaluation. Identify good-practice examples within the country and in similar countries and create a database. Selected cases from this database can be presented at workshops for key central and local government officials.
- Build a network to facilitate exchange among practitioners, academics, and civil servants in charge of M&E activities. Network activities can include knowledge dissemination and training. At the international level, the International Development Evaluation Association provides a forum to exchange information on good practices and methodologies.

As decentralization of administrative functions and service provision takes place in a country, it is important to build up M&E capacity at the subnational level. Regional and provincial administrations, and citizens, will need to assess the effectiveness of the strategy pursued at the local level. Central

Box 3.6. Roles of Various Agencies in Monitoring and Evaluation

Central ministries **such as planning and finance are usually in a good position to coordinate the design, monitoring, and support for M&E activities. The finance ministry also provides key data on public expenditures.**

Line ministries **are usually in charge of sectoral program coordination and supervision. Thus they play an important role in supervising the implementation of M&E activities at the sectoral level, and they are the key source of administrative records and data from management information systems.**

Project implementation agencies **are in charge of project and program management. They are responsible for the timely and appropriate implementation of program monitoring and evaluation.**

Central statistical offices **are key providers of data as well as expertise in data collection and analysis.**

Universities, research centers, and consulting firms **are potential suppliers of analysis and evaluation skills and also can offer training in a range of skills.**

Development assistance agencies **can help develop M&E capacity by providing technical assistance.**

statistical agencies are reluctant at times to build decentralized capacity, but this reluctance can be overcome if central and local M&E systems are seen as complementary. National agencies can continue to have responsibility for the conduct of data collection and analysis exercises at the national level; local agencies can develop the capacity to analyze subsets of the national data as well as collect and analyze data to assess the impact of local policies and programs.

Chapter 5, "Strengthening Statistical Systems," discusses in more detail assessing capacity and developing short- and long-term plans to strengthen capacity for quantitative data collection, while section 3.6 discusses the role of nongovernmental actors.[15]

3.5.2 Strengthening feedback mechanisms

Monitoring and impact evaluation should not be stand-alone, technical activities. They should be closely linked to decisionmaking processes at all levels and provide feedback to project managers, policymakers, and civil society on, among other things, the performance of existing policies and programs. Thus a crucial element of the M&E system is the existence of a *feedback process*.

A feedback process is a mechanism by which monitoring and evaluation results are disseminated and used to decide on future courses of action. Results should be disseminated broadly. M&E systems that provide results to only a select group of users (central ministries, for example) risk being underused and losing financial and political support. Wide dissemination of results reinforces the system by strengthening an outcome-based culture.

The dissemination strategy should accommodate the diverse information needs of different groups, including policymakers, program managers, program beneficiaries, the general public, the media, and academics. For example, reports that include main findings and emphasize implications for policy and program design can be distributed among government officials in central and line ministries as well as local administrations. Detailed reports can be produced for program administrators and researchers. Press releases can be used to reach the media. Workshops and seminars can be used to disseminate results among the general public and civil organizations. Posting of information on the Web, if possible, makes it available to interested audiences within and outside the country.

It is important that findings and recommendations be accessible to community councils, local women's organizations, and ethnic, religious, environmental, and other groups representing communities to whom programs are targeted. Most of these groups may not have access to information technology and conventional dissemination mechanisms. In these cases, alternative dissemination methods, such as meetings, pamphlets, posters, and so on, may be required. Dissemination materials prepared in more than one language and separate meetings with different groups (for example, men and women) may also be required. Active participation of NGOs and other local organizations may be crucial to ensure that all sectors of the community are reached.

In addition to results, the actual data and careful documentation of methods of analysis should also be made available to the public. Reluctance in releasing unit record data can give rise to suspicion, while open access and discussion over data, methods, and results foster transparency and broad acceptance of the findings. Open access to unit record data also enables NGOs to carry out independent analysis and increases demand for data, which helps ensure the sustainability of the M&E system. In some countries there are legal impediments to the dissemination of raw data related to the protection of privacy; these can be overcome with technical solutions that make it very hard to identify respondents and changes in the legal framework; many countries now grant open data access, and lessons have been learned from their experience.

Beyond broad dissemination, a well-established process to feed M&E results back to policymakers is crucial if results are to be used in formulating policy. Since key policy decisions are made at the time of budget formulation, key results should be available then. This particularly means that data for the first six months of the fiscal year should be available not just on expenditures but also on outputs. Any data on other intermediate and final indicators tracked annually should also be made available at the time of budget formulation.

In some countries, poverty monitoring units have been established with the explicit purpose of providing policymakers with information on which to base decisions. These units have been most successful when they have been located close to decisionmaking centers (such as the Prime Minister's Office) and when they have acquired adequate capacity to provide competent and timely information. In other cases, independent agencies have been set up (such as the *observatoires* in some West African countries).

3.6 Promoting Participation in Monitoring and Evaluation

Nongovernmental actors, from researchers and community organizers to representatives of the poor, have an important role to play in monitoring and evaluation: they can contribute their knowledge and expertise to the design of the M&E system, carry out M&E activities directly, and use the results to keep governments honest.

Broad consultations during the design of the M&E system are important to build consensus on what to monitor and what to evaluate—the selection of indicators and targets—and generate a sense of ownership among different groups in society, thus increasing the acceptance and use of findings. Consultations help to identify adequate indicators of people's perception of well-being and bring into the process the expertise of NGOs.

In addition to providing their views, expertise, and knowledge during the design of the system, civil society organizations can contribute directly to implementing M&E activities, either independently or under contracts from the public sector. Research organizations and universities often have the capacity and expertise to carry out surveys and participatory work and analyze the results, while interest groups and community-based groups can take advantage of easy access to their members to get their views and opinions. Also, civil society organizations are sometimes more experienced than government agencies in the use of participatory methods of data collection and analysis.

Finally, civil society organizations have a crucial role to play as users of M&E results. Wide dissemination of results encourages participation. By accessing M&E findings, civil society organizations can generate a participatory review process of poverty reduction efforts that increases accountability and transparency of public resource allocation and public actions. Chapter 7, "Participation," expands on these issues and discusses alternative strategies to promote participation depending on country circumstances. For information on promoting women's participation, see chapter 10, "Gender."

Notes

1. This chapter takes the goals as given. See chapter 7, "Participation," for a discussion of participatory goal setting.
2. In this respect a poverty monitoring system combines implementation monitoring and performance- or results-based monitoring (sometimes the term "poverty monitoring system" is also used to refer to outcome/impact monitoring only).
3. For a discussion of how health targets can be reached with different degrees of improvement for the poorest and richest, see Gwatkin 2000a, 2000b.
4. For example, existing household survey data may be too old, or the sampling methodology may not ensure representativeness.
5. Guidance on the frequency of collection of gender-based indicators can be found in chapter 10, "Gender."
6. For more discussion of systems to improve the tracking of public expenditures, see chapter 6, "Public Spending." See also the assessment of expenditure tracking systems done by the World Bank for the Highly Indebted Poor Countries initiative: http://www.worldbank.org/hipc/tracking.pdf.
7. For more information on costing programs, see chapter 4, "Development Targets and Costs"

8. For more detail on the methodology and findings, see Abdo and Reinikka (1998) and Republic of Uganda (1998). Survey instruments can be found at http://www.worldbank.org/research/projects/publicspending/ tools/tools.htm.
9. For more information on ways to improve the timeliness of household survey data, see Grosh and Munoz 1996.
10. For more information on the Core Welfare Indicators Questionnaire, see www4.worldbank.org/afr/stats/cwiq.cfm; copies of the brochure, questionnaire, handbook, and various other documents about the CWIQ can be downloaded from the site. For more information on user surveys in Uganda, see http://www.worldbank.org/research/projects/publicspending/tools/tools. htm.
11. Where migration is an important issue, a new group of immigrant households can be incorporated into the sample at different points in time.
12. Adapted from a presentation by Rosemary McGee and John Gaventa of the Institute for Development Studies.
13. The annual progress report would also discuss any modifications in the strategy or its implementation that may be necessary given the findings of monitoring and evaluation activities. See IMF and World Bank, [December] 1999, "PRSPs—Operational Issues," IMF and World Bank, Washington, D.C.
14. See, for example, Blank and Grosh (1999) on how to use household surveys to build analytical capacity.
15. See also http://www.worldbank.org/html/oed/evaluation/html/monitoring_and_evaluation_capa.html for additional information on assessment tools and lessons learned in building institutional capacity for monitoring and evaluation.

Guide to Web Resources

Baker, Judy. 2000. "Evaluating the Poverty Impact of Projects: A Handbook for Practitioners." Directions in Development. World Bank, Washington, D.C. This handbook seeks to provide project managers and policy analysts with the tools needed for evaluating the impact of interventions. It includes a discussion of evaluation methodologies and implementation issues and presents several case studies, some of them also included in this chapter.

Available at http://www.worldbank.org/poverty/library/impact.htm.

MacKay, Keith. 1999. "Evaluation Capacity Development: A Diagnostic Guide and Action Framework." ECD Working Paper Series 6. World Bank, Operations Evaluation Department, Washington, D.C. This guide provides a detailed checklist of issues to be considered in developing a country's evaluation capacity.

Available at http://www.worldbank.org/html/oed/evaluation/html/ecd_doc.html.

World Bank. 1999. "CWIQ (Core Welfare Indicators Questionnaire) Handbook and CD-ROM." Africa Operational Quality and Knowledge Services. World Bank, Washington, D.C. This handbook provides guidance on the use and implementation of the CWIQ.

Available at http://www4.worldbank.org/afr/stats/cwiq.cfm.

Web Sites

Monitoring and Evaluation Capacity Development (http://www.worldbank.org/evaluation/me/). Contains assessment tools and lessons learned in building institutional capacity for monitoring and evaluation.

PovertyNet (http://www.worldbank.org/poverty/). Provides a number of resources for poverty monitoring, including links to the poverty monitoring database, LSMS site, Poverty in Africa site, Africa Household Survey databank, and impact evaluation site.

Poverty Reduction Strategy Papers (http://www.worldbank.org/poverty/strategies/index.htm). Includes interim and final PRSPs prepared by countries.

Bibliography and References

Ablo, Emmanuel, and Ritva Reinikka. 1998. "Do Budgets Really Matter? Evidence from Public Spending on Education and Health in Uganda." Policy Research Working Paper 1926. World Bank, Africa Region. Washington, D.C.

Baker, Judy. 2000. *Evaluating the Impact of Development Projects on Poverty: A Handbook for Practitioners.* Directions in Development. Washington, D.C.: World Bank.

Blank, J., and Margaret Grosh. 1999. "Using Household Surveys to Build Analytic Capacity." *World Bank Research Observer* 14(2):209–27.

Carvalho, Soniya, and Howard White. 1997. "Combining the Quantitative and Qualitative Approaches to Poverty Measurement and Analysis: The Practice and the Potential." Technical Paper 366. World Bank, Washington, D.C.

Fuller, Bruce, and Magdalena Rivarola. 1998. "Nicaragua's Experiment to Decentralize Schools: Views of Parents, Teachers, and Directors." Working Paper Series on Impact Evaluation of Education Reforms, No. 5. World Bank, Development Economics Research Group. Washington, D.C.

Grosh, Margaret, and Juan Muñoz. 1996. "A Manual for Planning and Implementing the LSMS Survey." Living Standards Measurement Survey Working Papers Series 126 (also available in Russian and Spanish). World Bank, Washington, D.C.

Grossman, Jean Baldwin. 1994. "Evaluating Social Policies: Principles and U.S. Experience." *World Bank Research Observer* 9(2):159–80.

Gwatkin, Davidson R. 2000a. "Health Inequalities and the Health of the Poor: What Do We Know? What Can We Do?" *Bulletin of the World Health Organization* 78(1):3-18.

———. 2000b. "Meeting the 2015 International Development Target for Infant Mortality: How Much Would the Poor Benefit?" Human Development Network, World Bank, Washington, D.C.

Hentschel, Jesko. 1998. "Distinguishing between Types of Data and Methods of Collecting Them." Policy Research Working Paper 1914. World Bank, Poverty Reduction and Economic Management Network, Poverty Division. Washington, D.C.

International Monetary Fund, World Bank. 1999, December. "PRSP—Operational Issues." Paper presented to the Boards of Directors. Washington, D.C.

King, Elizabeth, and Berk Ozler. 1998. "What's Decentralization Got to Do with Learning? The Case of Nicaragua's School Autonomy Reform." Working Paper Series on Impact Evaluation of Education Reforms No. 9. World Bank, Development Economics Research Group. Washington, D.C.

Kozel, Valerie, and Barbara Parker. 1998. "Poverty in Rural India: The Contribution of Qualitative Research in Poverty Analysis." World Bank, Poverty Reduction and Economic Management Sector Unit. Washington, D.C.

MacKay, Keith. 1999. "Evaluation Capacity Development: A Diagnostic Guide and Action Framework." ECD Working Paper Series 6. World Bank, Operations Evaluation Department. Washington, D.C.

MacKay, Keith, and Sulley Gariba, eds. 2000. "The Role of Civil Society in Assessing Public Sector Performance in Ghana: Proceedings of a Workshop." World Bank, Evaluation Capacity Development, Operations Evaluation Department. Washington, D.C.

Narayan, Deepa. 1993. "Participatory Evaluation: Tools for Managing Change in Water and Sanitation." Technical Paper 207. World Bank, Washington, D.C.

Nicaragua Reform Evaluation Team. 1996. "Nicaragua's School Autonomy Reform: A First Look." Working Paper Series on Impact Evaluation of Education Reforms 1. World Bank, Poverty and Human Resources Division, Policy Research Department. Washington, D.C.

Patton, Michael Q. 1987. *How to Use Qualitative Methods in Evaluation*. Newbury Park, Calif.: Sage Publications.

United Republic of Tanzania, The. 2000. "Poverty Reduction Strategy Paper." Dar Es Salaam.

Republic of Uganda. 1999a. "Uganda Poverty Status Report, 1999." Ministry of Finance, Planning and Economic Development. Kampala.

———. 1999b. "Five Year Strategy for Poverty Monitoring and Policy Analysis." Planning and Poverty Eradication Section, Ministry of Finance, Planning and Economic Development. Kampala.

———. 1998. "Monitoring and Evaluating Accountability and Transparency of Schools and Districts for UPE Funds." Ministry of Education and Sports. Kampala.

———. 1997. "Poverty Eradication Action Plan: A National Challenge for Uganda." Ministry of Finance, Planning and Economic Development. Kampala.

Rietbergen-McCracken, Jennifer, and Deepa Narayan, eds. 1998. *Participation and Social Assessment: Tools and Techniques*. Washington, D.C.: World Bank.

Rossi, Peter H., and Howard E. Freeman. 1982. *Evaluation: A Systematic Approach.* 2d ed. Beverly Hills, Calif.: Sage Publications.

Valadez, Joseph, and Michael Bamberger, eds. 1994. *Monitoring and Evaluating Social Programs in Developing Countries: A Handbook for Policymakers, Managers, and Researchers*. EDI Development Studies. Washington, D.C.: World Bank.

van de Walle, Dominique. 1999. "Assessing the Poverty Impact of Rural Road Projects." World Bank, Development Research Group. Washington, D.C.

Weiss, Carol H. 1998. *Evaluation*. 2d ed. Upper Saddle River, N.J.: Prentice-Hall.

World Bank. 1994. "Building Evaluation Capacity." Lessons and Practices No. 4. Operations Evaluation Department. Washington, D.C.