PART I

Overview: RealWorld Evaluation
CHAPTER 1

RealWorld Evaluation and the Contexts in Which It Is Used

The chapter begins with an overview of the RealWorld Evaluation (RWE) approach, the contexts in which real-world evaluations are conducted, and the many different constraints, pressures, and influences under which evaluations are formulated, conducted, disseminated, and used. The RWE approach was developed to address four of the most common constraints evaluators face: budget, time, data, and political influences. The two most common RWE scenarios are reviewed. The first is when the evaluator is brought in at the start of the project but with constraints on the types of information that can be collected or the designs that can be used. The second, and probably the more common scenario, is when the evaluator is not called in until the project has been operating for some time and may even be almost completed. In most of these cases, no baseline data have been collected and usually no comparison (control) group has been identified. We compare the similarities and differences between approaches to program evaluation and the potential demand for RWE in developing and developed countries. The chapter concludes by identifying the main groups of users of program evaluation and discusses how and when they use evaluations.

Welcome to RealWorld Evaluation

Most evaluators are familiar with situations in which programs have been underway for some time or perhaps are almost completed before implementing or funding agencies begin to think seriously about evaluating the extent to which the programs are achieving their objectives and producing the intended impacts. Usually, the belated interest in evaluation is motivated by the need for solid evidence on which to base decisions about whether the program should be
continued or perhaps expanded. When the evaluations do finally get underway, many have to be conducted under budget and time constraints, often with limited access to baseline data and comparison groups. Consequently, it is difficult, if not impossible, to apply many of the methodologically most robust evaluation designs.

Although more resources are allocated to evaluation in developed countries, many evaluators in the United States, Canada, Europe, and Australasia report that they operate under similar, although sometimes less severe, constraints to those faced by their colleagues in developing countries. As if these problems were not enough, many evaluations in both developed and developing countries are often conducted in a political environment where funding agencies, clients, and key stakeholders have strongly held views on what are the “right” evaluation methods, what types and amounts of information should be collected, and which groups should and should not be asked to comment on (or even see) the findings. New evaluators soon discover that “technical” issues such as whether to use randomized control groups; the choice of qualitative, quantitative, or mixed-method designs; and who to interview and what questions to ask can provoke strong reactions from clients and stakeholders.

Despite the difficult circumstances under which many evaluations have to be conducted, there is a growing demand from funding agencies, governments, civil society, and intended beneficiaries for systematic program evaluations, including whether the program could and/or should be continued or expanded to other communities or locations. Consequently, there is a strong demand from many sides for evaluators to answer basic questions such as these:

- Did the project meet its objectives?
- Did it have an impact?
- Who benefited and who did not?
- Should the program continue?

There is also an increasing awareness that evaluation conclusions need to be supported by sound evidence and not just opinions—although there are often major disagreements as to what constitutes sound evidence.

The pressures of conducting evaluations under budget and time constraints have often resulted in inattention to sound research design or to identifying and addressing factors affecting the validity of the findings. The RWE approach presented in this book was developed in response to the demand for guidance on how to conduct evaluations within budget, time, data, and political constraints while at the same time ensuring maximum possible methodological rigor within the particular evaluation context. RWE is based on the following seven-step approach, summarized more specifically in Figure 1.1 and described in detail in Chapters 2 through 8:

- Step 1: Planning and scoping the evaluation. Understanding client information needs and the political context within which the evaluation is being conducted; defining a program theory that describes the objectives and intended method of operation of the program; identifying the budget, time, and data constraints and any political factors
affecting how the evaluation will be conducted, disseminated, and used; combining this information with preliminary analysis from Steps 2 to 5 to select the design that best addresses client needs within the **RWE constraints**

- **Step 2: Strategies for addressing budget constraints.** Reducing the costs through simplifying the evaluation design; reducing the amounts of data to be collected; making greater use of secondary data; revising the sample design and streamlining data collection and analysis. Many of the techniques described in Step 3 for using expensive consultants more efficiently can also help reduce costs.

- **Step 3: Strategies for addressing time constraints.** In addition to many of the approaches used in Step 2, strategies include: planning ahead to avoid delays and bottlenecks, particularly during the short periods when outside consultants are involved; building impact-related indicators into routine project monitoring data collection and using videoconferencing to reduce travel and to permit more frequent interactions between the evaluation team and agency staff

- **Step 4: Strategies for addressing data constraints.** Reconstructing baseline or comparison group data if either of these was not collected; making effective use of secondary data; making sure that data are collected from the right people; and time- and cost-effective methods for collecting data on sensitive topics and for locating and interviewing difficult-to-reach groups

- **Step 5: Understanding and coping with political factors influencing how the evaluation is designed, implemented, disseminated, or used.** Identifying the key actors and their political perspectives and understanding how these affect their orientation to the evaluation; developing strategies to address the political realities without compromising the evaluation

- **Step 6: Strengthening the evaluation design and the validity of conclusions.** Creating a checklist to identify threats to the validity and adequacy of the evaluation design and conclusions; creating a set of strategies to address the threats and improve the quality of the evaluation

- **Step 7: Helping clients use the evaluation.** Ensuring that clients are actively involved from the start and that they “buy into” the evaluation; maintaining contact with clients throughout the evaluation and ensuring that by the time the major reports are published, they do not contain any surprises for the client; adapting the presentation of findings to the preferred communication style of different stakeholders

---

**The RealWorld Evaluation Context**

As noted, the RWE approach was developed to assist the many evaluators in both developing and developed countries who must conduct evaluations with budget, time, data, and political constraints. In one common scenario, the client delays contracting an evaluator until late in the project when the funding agency (government, international development agency, foundation, etc.) is about to decide whether to continue to support a project or possibly launch a larger second phase. Such tardiness occurs even when evaluation was built into the original project agreement (see Box 1.1). With the decision point approaching, the funding agency may suddenly realize that it does not have solid information on which to base a decision about future funding of the project, or the project implementing agency may realize it does not have
the evidence needed to support its claim that the project is achieving its objectives. An evaluator called in at this point may be told it is essential to conduct the evaluation by a certain date and to produce “rigorous” findings regarding project impact, although unfortunately, very limited funds are available.
In other scenarios, the evaluator may be called in early but finds that for budget, political, or methodological reasons, it will not be possible to collect data on a comparison group for purposes of determining program impact by comparing participants with nonparticipants. In some cases, it may not even be possible to collect baseline data on the project population for purposes of analyzing progress or impact over time. Data constraints may also result from difficulties of collecting information on sensitive topics such as HIV/AIDS, domestic violence, postconflict reconstruction, or illegal economic activities (e.g., commercial sex workers, narcotics, or political corruption).

Determining the most appropriate evaluation design under these kinds of circumstances can be a complicated juggling act involving a trade-off between available resources and acceptable standards of evaluation practice. Often, the client’s concerns are more about budgets and deadlines, and basic principles of evaluation may receive a lower priority. Box 1.2 illustrates this difficult trade-off between budgets and deadlines on the one hand and desired standards of methodological rigor on the other. Failure to reach satisfactory resolution of these trade-offs may also contribute to a much lamented problem: low use of evaluation results (see Chelimsky 1994; Operations Evaluation Department 2005; Patton 1997). RWE is a response to the all-too-real difficulties in the practical world of evaluation.

### Box 1.1 A Familiar Evaluation Story

When a social development fund was launched in an African country a few years ago, it was suggested that a baseline study be conducted as the first phase of a longitudinal impact evaluation study. The project manager asked, “What is the point of spending money and time on a baseline study when we do not know if the project model will work in our country?” He also indicated that staff members were under pressure to launch the project and could not spend time on something that would not be useful until the project was completed. Three years later, when the possibility of a second project was being discussed, consultants were called in to conduct an impact evaluation study. It was agreed that it was unfortunate that no baseline data were available to permit a rigorous measurement of the changes produced by the project. The consultants had to try to reconstruct baseline data using methods described in Chapter 5.

In other scenarios, the evaluator may be called in early but finds that for budget, political, or methodological reasons, it will not be possible to collect data on a comparison group for purposes of determining program impact by comparing participants with nonparticipants. In some cases, it may not even be possible to collect baseline data on the project population for purposes of analyzing progress or impact over time. Data constraints may also result from difficulties of collecting information on sensitive topics such as HIV/AIDS, domestic violence, postconflict reconstruction, or illegal economic activities (e.g., commercial sex workers, narcotics, or political corruption).

Determining the most appropriate evaluation design under these kinds of circumstances can be a complicated juggling act involving a trade-off between available resources and acceptable standards of evaluation practice. Often, the client’s concerns are more about budgets and deadlines, and basic principles of evaluation may receive a lower priority. Box 1.2 illustrates this difficult trade-off between budgets and deadlines on the one hand and desired standards of methodological rigor on the other. Failure to reach satisfactory resolution of these trade-offs may also contribute to a much lamented problem: low use of evaluation results (see Chelimsky 1994; Operations Evaluation Department 2005; Patton 1997). RWE is a response to the all-too-real difficulties in the practical world of evaluation.

### The Four Types of Constraints Addressed by the RealWorld Approach

Table 1.1 illustrates the different ways in which RWE constraints are combined in the typical contexts in which evaluations are conducted. In some cases, the evaluator faces a single constraint. For example, the budget may be limited, but there is plenty of time. Or the evaluation may begin at the start of the project with no time constraint, but the evaluator is told that for political or ethical reasons, it will not be possible to collect data on a comparison group. Many unlucky evaluators find themselves simultaneously contending with several or all of these constraints!
Budget Constraints

Sometimes, funds for the evaluation were not included in the original project budget, and the evaluation must be conducted with a much smaller budget than would normally be allocated. As a result, it may not be possible to collect the desirable data or to reconstruct baseline or comparison group data. Lack of funds may create or exacerbate time constraints because researchers may not be able to spend as much time in the field as they consider necessary. Box 1.3 makes the point that it is important to understand whether the main constraint is budget or time (or both), because the best strategy will often be different in each case.

Time Constraints

The most common time constraint is when the evaluator is not called in until the project is already well advanced and the evaluation has to be conducted within a much shorter period of time than the evaluator considers necessary—in terms of a longitudinal perspective over the life of the project, the time allotted for conducting the end-of-project evaluation, or both. Time constraints often make it impossible to conduct a pretest-posttest evaluation design with a baseline study that can be repeated after the project has been implemented. The time available for planning stakeholder consultations, site visits and fieldwork, and data analysis may also have to be drastically reduced to meet the report deadline. These time pressures are
### Table 1.1  
RealWorld Evaluation Scenarios: Conducting Impact Evaluations with Time, Budget, Data, and Political Constraints

<table>
<thead>
<tr>
<th>Time</th>
<th>Budget</th>
<th>Data</th>
<th>Political</th>
<th>Typical Evaluation Scenarios</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td>The evaluator is called in late in the project and told that the evaluation must be completed by a certain date so that it can be used in a decision-making process or contribute to a report. The budget may be adequate, but it may be difficult to collect or analyze survey data within the time frame.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>The evaluation is allocated only a small budget, but there is not necessarily excessive time pressure. However, it will be difficult to collect sample survey data because of the limited budget.</td>
</tr>
<tr>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td>The evaluator is not called in until the project is well advanced. Consequently, no baseline survey has been conducted either on the project population or on a control group. The evaluation does have an adequate scope, either to analyze existing household survey data or to collect additional data. In some cases, the intended project impacts may also concern changes in sensitive areas such as domestic violence, community conflict, women’s empowerment, community leadership styles, or corruption on which it is difficult to collect reliable data—even when time and budget are not constraints.</td>
</tr>
<tr>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td>The funding agency or a government regulatory body has requirements concerning acceptable evaluation methods. For example: In the United States, the No Child Left Behind Act of 2001 includes funding preference for certain types of research designs. In other cases, a client or funding agency may specifically request qualitative data, tests of statistical significance regarding measured program effects, or both.</td>
</tr>
<tr>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td>There is overwhelming indication that the evaluation is being commissioned for political purposes. For example, an evaluation of the effects of conservation policy might be commissioned to stall its expansion.</td>
</tr>
<tr>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td>There is reason to suspect that the evaluation will be used for political purposes other than or contrary to those articulated in preliminary discussions. For example, an evaluator might suspect that an evaluation of charter schools might be used (and even misused) by a client with known advocacy for privatization of education.</td>
</tr>
<tr>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td>The evaluator has to operate under time pressure and with a limited budget. Secondary survey data may be available but there is little time or few resources to analyze it them.</td>
</tr>
<tr>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td>The evaluator has little time and no access to baseline data or a comparison group. Funds are available to collect additional data, but the survey design is constrained by the tight deadlines.</td>
</tr>
<tr>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td>The evaluator is called in late and has no access to baseline data or comparison groups. The budget is limited, but time is not a constraint.</td>
</tr>
<tr>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>The evaluator is called in late, has given a limited budget, and has no access to baseline survey data; and no comparison group has been identified.</td>
</tr>
</tbody>
</table>

NOTE: To simplify the table, the possible combinations of political constraints with the other three factors have not been included in the table.
particularly problematic for an evaluator who is not familiar with the area, or even the
country, and who does not have time for familiarization and for building confidence with the
communities and the agencies involved with the study. The combination of time and budget
constraints frequently means that foreign evaluators (and out-of-town U.S. evaluators) can
be in the country or the state for only a short period of time—often requiring them to use
shortcuts that they recognize as methodologically questionable.

Data Constraints

When the evaluation does not start until late in the project cycle, there is usually little or no
comparable baseline information available on the conditions of the target group before the
start of the project. Even if project records are available, they are often not organized in the
form needed for comparative before-and-after analysis. Project records and other documen-
tary data often suffer from reporting biases or poor record-keeping standards. Even when sec-
ondary data are available for a period close to the project starting date, they usually do not fully
match the project populations. For example, employment data may cover only larger compa-
nies, whereas many project families work in smaller firms in the informal sector, or school
records may cover public schools but not religious and other private schools. In developing
countries, survey data are often analyzed and presented only at the household level, even when
the data may have been collected on each household member.3 This is a particular problem for
gender analysis.

Most clients are interested in collecting data only on the groups or communities with which
they are working. They may also be concerned that collection of information on nonbeneficia-
ries might create expectations of financial compensation or other benefits (for which the pro-
ject has no budget), which further discourages the collection of data on a comparison group.
Even if funds are available, it is also often difficult to identify a comparison group, because
many project areas have unique characteristics. Where intended project impacts concern sen-
sitive topics such as women’s empowerment, contraceptive usage, or domestic violence, espe-
cially in paternalistic societies, information may be difficult to collect even when funds are
available (see Box 1.4). Similar data problems can arise when working with difficult-to-reach
groups such as drug addicts, criminals, ethnic minorities, migrants, or illegal residents.
We use the term political influences and constraints in a broad sense to refer not only to pressures from government agencies and politicians but also to include the requirements of funding or regulatory agencies, pressures from stakeholders, and differences of opinion within an evaluation team regarding evaluation approaches or methods.

Evaluations are frequently conducted in contexts where political and ethical issues affect design and use. All programs affect some portion of the public, and most programs consume public funds, always limited and often scarce. Decisions based on evaluation results may intensify competition for funding, expand or terminate programs needed by some and paid for by others, or advance the agenda of a politically oriented group. Box 1.5 gives an example of how political pressures often affect the evaluation design—in this case, forbidding the use of a comparison group.

### Box 1.4 Problems in Capturing Information from or about Women

- Many household surveys only interview the “household head,” who is often considered to be the male. He often does not have all the information on female household members or gives low priority to their concerns. Many men, for example, say their wives are happy to spend several hours per day walking to collect water or fuel because they “sing and chat with their friends as they walk.”
- Women are often interviewed in the presence of other household members where they may not feel free to express their views.
- Donor agencies often insist that women be invited to attend community meetings to discuss proposed projects. However, the women often do not feel free to speak in public, or they always say they agree with their husbands.
- In many parts of the world, sexual harassment is one of the main reasons women do not use public transport. However, it is culturally impossible for women to mention this to an outside interviewer, so this major problem is often not captured in surveys.

### Political Influences

We use the term political influences and constraints in a broad sense to refer not only to pressures from government agencies and politicians but also to include the requirements of funding or regulatory agencies, pressures from stakeholders, and differences of opinion within an evaluation team regarding evaluation approaches or methods.

Evaluations are frequently conducted in contexts where political and ethical issues affect design and use. All programs affect some portion of the public, and most programs consume public funds, always limited and often scarce. Decisions based on evaluation results may intensify competition for funding, expand or terminate programs needed by some and paid for by others, or advance the agenda of a politically oriented group. Box 1.5 gives an example of how political pressures often affect the evaluation design—in this case, forbidding the use of a comparison group.

### Box 1.5 Political Influence on the Evaluation of a Power Project in Asia

Consultants were asked to design an evaluation to assess the impacts of a hydroelectric power project in an Asian country that would involve the forced resettlement of a number of villages in the area where the dam was to be constructed. Families who had title to their land would receive compensation. The consultants proposed that the evaluation should include a comparison group of families who did not have land title. They were informed by the power authority that it would not be possible to do this because this would create expectations that these families would also receive compensation for being relocated, and funds for this were not included in the project budget.
While evaluators are always quick to spot the political or ideological biases of their clients and stakeholders, they are often less aware (or open) about their own ideological orientations. Many of the ongoing debates between quantitative and qualitative evaluators are fueled by the search for the “correct” or “best” research paradigm. Let us take, as an example, different approaches to a data set on income and employment in enterprises employing 10 or more people, a common cutoff point in employment surveys. A researcher in an economics department might decide to use econometric analysis to estimate the income and employment characteristics of people working in the smaller enterprises not covered by the data set, whereas a researcher in a sociology department might locate and interview a sample of workers in firms employing fewer than 10 workers.

The RealWorld Approach to Evaluation Challenges

Although RWE does not develop many new data collection or analysis methods, the approach makes several contributions to the conduct of evaluations under real-world budget, time, data, and political constraints. First, it presents ways to draw from a wide range of evaluation approaches and methods to address the four types of constraints described earlier. The systematic use of mixed methods is emphasized throughout. Using mixed-method approaches is considered critical for several reasons: (a) It permits the evaluator to draw on the widest possible range of evaluation methods and tools, (b) it increases the validity of conclusions by providing two or more independent estimates of key indicators (triangulation), (c) it permits a deeper and richer analysis and interpretation of the context in which a program operates, and (d) it offers ways to reduce the costs or time of data collection (see Chapters 3 and 4).

Second, RWE’s seven-step approach to quality assurance offers corrective measures that can be introduced in different phases of the evaluation process, some even after a draft evaluation report has been produced, helping to enhance the quality of the evaluation. Quality promotes credibility and utility of findings, which, in turn, help ensure that evaluation contributes to the public good.

Third, many quantitative evaluations rely on the pretest-posttest design to estimate the changes and impacts produced by a project or program. This approach, when used in isolation, has two serious limitations: (a) It does not take into account the different socioeconomic and political contexts affecting each project, and (b) it implicitly assumes that each project is implemented as planned and in exactly the same way in each location. One of the contributions of RWE is to look inside the “black box” of the project implementation process to examine what actually happens during implementation and how much variation there is between different project sites (see Box 1.6). It also focuses on quality of implementation. This is a critical contribution because in many real-world contexts, some project components are not implemented at all or the quality is so low that it is hardly surprising that the intended impacts were not achieved. In other cases, the intended impacts were achieved, but what went on within the project was quite different from what had been planned!
Comparing the RealWorld Evaluation Context and Issues in Developing and Developed Countries

The Evolution of Program Evaluation

Discussions regarding how to ensure competent evaluation practice have been underway for more than half a century in the United States, at least since evaluation began to be required on a massive scale with the passage of the Elementary and Secondary Education Act of 1965. In the United States, education continues to be one of the areas in which much of the cutting-edge work on evaluation methodology is conducted. The Great Society programs of the 1960s and 1970s (e.g., Head Start and urban housing programs) were also seen as an opportunity to study scientifically the causes and effectiveness of large-scale programs to eradicate poverty. These produced some of the largest and most carefully designed quantitative program evaluations ever conducted.

Evaluation codes of conduct, especially the Program Evaluation Standards (Joint Committee 1994) and the Guiding Principles for Evaluators (American Evaluation Association 1995), have been promulgated and periodically reviewed and revised. An increasing number of developing countries have adopted or adapted the Program Evaluation Standards (Joint Committee 1994), although there is a continuing debate concerning the applicability of these standards in other countries (Russon and Russon 2005). In the mid-1980s, the Organization for Economic Cooperation and Development (OECD 1986) developed similar guidelines to assist donor countries in the evaluation of their assistance programs to developing countries.

Professional evaluation associations have been active for many years in Western Europe and in Australasia, but an important new development has been the dramatic increase in the number of national and regional evaluation organizations in developing countries. There are now at least 40 national and regional evaluation organizations covering Africa, Latin America, South and East Asia, and the Newly Independent States (see Appendix 4 for a listing with Web sites). In 2004, the International Organization for Cooperation and Evaluation (IOCE)’ was established to provide a forum and voice for these associations in both developing and industrialized nations.

---

Box 1.6  Getting inside the “Black Box”

Many impact evaluations assume that projects are implemented exactly as planned and in exactly the same way in each location. In fact, there are often major differences in how each project is implemented depending on local cultural, economic, administrative, and political factors. In some cases, the pretest-posttest evaluation is faithfully conducted without realizing that some of the project components were never implemented at all. Women did not apply for loans because it was too far to travel to the bank in town, teachers did not come to school during the planting season, textbooks never reached many of the schools, and parents in some areas did not send their daughters to school.

Unless the evaluation looks inside the “black box” of the project’s implementation process, many of the findings of an impact evaluation can be very misleading and of little practical utility.
IDEAS (the International Development Evaluation Association) was also created recently to provide a forum for evaluators who work in developing countries.

**Special Evaluation Challenges in Developing Countries**

Although the program evaluation tradition is well established in a number of developing countries (e.g., in India, the Program Evaluation Organization has been operating under the Planning Commission for nearly 50 years), in many countries program evaluation systems are not yet well established. Where public resources are desperately scarce, evaluation is often accorded low priority. Key stakeholders may perceive evaluation as impractical, unnecessary, or even threatening and may object to externally imposed monitoring or evaluation systems designed to respond to the information needs of funding agencies rather than to those of the national stakeholders (Bamberger 2001; Horton and Mackay 1999). Yet in countries or regions where it is most difficult to reallocate resources from programming to evaluation, the need for effective and efficient programs and for evaluation to identify and sustain these programs may be greatest.

Lack of support for program evaluation should not be attributed exclusively to a lack of understanding or motivation. Even when local program personnel and other stakeholders are interested in evaluation, formidable challenges can still be encountered, including the following:

- Not enough attention given to evaluation until the program is well advanced, with the consequence that there are no preprogram data (e.g., baseline studies), no defined comparison group, and insufficiently defined program objectives
- Modest budget allocation for evaluation studies
- Pressure to complete the evaluation as quickly as possible
- Cultural and political complexities affecting the conduct and use of evaluations are not easily grasped
- A limited pool of national evaluation expertise

Evaluators from developed nations do not always come prepared for the methodological challenges of working in international development evaluation where data may be more elusive and experienced research collaborators harder to find. Of perhaps even more import, they are often even less prepared for national organizational, institutional, political, and cultural contexts that can complicate an evaluation process. Constraints typically include limited access to data and often the absence of a culture of evaluation among managers, policymakers, and sectors of civil society. As we will see later, these factors should not come as a surprise because they are also operating in the industrialized nations from which many of these evaluators come!

Despite these constraints, demand from policymakers, managers, and civil society for information regarding goal achievement and the impact of projects and policies is increasing. However, many government bodies and international funding agencies do not begin to focus on evaluation until late in the project cycle (Bamberger 2000a), intensifying demand for rapid, cost-effective methodologies for assessing project impact. Unfortunately, enthusiasm for information about impact is usually not matched by adequacy of evaluation resources. Consequently, international and national evaluators are frequently asked to produce methodologically robust
impact evaluations under circumstances in which it is impossible to comply fully with conventional evaluation standards.

**Similarities and Differences of RealWorld Constraints in Developing and Developed Nations**

It would be a mistake to assume that resource and data constraints are always more severe in developing countries and that evaluations in developed countries are always better planned and financed. Although some high-profile evaluations are well funded and the sheer volume of evaluation reports produced by the U.S. General Accounting Office or other federal agencies such as the Department of Education or Health and Human Services is overwhelming, many U.S. state and local agencies and voluntary organizations suffer from serious underfunding, which critically limits their ability to conduct evaluations.

In the United States, Canada, Europe, Australasia, and Japan, evaluation budgets are often just as constrained as in developing countries by low evaluation priority, suspicion, and last-minute requests. While a relatively small number of high-profile evaluations are well funded, many evaluators face budgetary problems similar to those encountered by their colleagues in developing countries, especially during periods in which economic downturns have forced severe financial cuts. Even when expensive and technically sound evaluations are conducted of publicly funded programs in the United States, use of results can be low (e.g., Patton 1997).

It is also important to recall that most of the examples of rigorous, large-scale evaluations are conducted only on government-funded programs or a few large foundation-supported programs. Large numbers of nongovernmental organizations (NGOs), community groups, and state and locally funded government programs must conduct evaluations with similar budget, time, and data constraints facing evaluators in developing countries. So while many evaluators in industrialized countries often have greater access to secondary data and longitudinal series from censuses, government surveys, and public records, it is clear that issues addressed by RWE are relevant to many evaluators in the United States, Canada, Europe, and Australasia just as much as to their colleagues in developing countries.

**Who Uses RealWorld Evaluation, for What Purposes, and When?**

There are two main users of RWE: First, evaluation practitioners will find it useful to use RWE for a number of reasons. For example:

- To identify ways to conduct adequately rigorous evaluations given limitations of time and financial resources
- To overcome data constraints, particularly the lack of baseline and comparison data
- To identify and address factors affecting the validity and adequacy of the findings of the evaluation

Second, government agencies, international development agencies, and foundations who commission evaluations and/or use evaluation findings will find the RWE approach useful for these reasons:
• To identify ways to reduce the costs and time of evaluations
• To be more fully aware of the various constraints under which an evaluation is to be conducted
• To understand the implications of different RWE strategies on the ability of the evaluation to respond to the purposes for which it was commissioned

Table 1.2 shows that RWE can be conducted at three different points in a project or program: at the start during the planning stage, when the project is already being implemented, or at the end. When the evaluation begins at the start of the project, RWE is used (a) to help identify different options for reducing costs or time of the evaluation, (b) for deciding how to make the best use of available data, (c) to understand client information needs and the political context within which the evaluation will be conducted, (d) for deciding what evaluation design would be appropriate, (e) for deciding what data needs to be collected by the monitoring system during the implementation of the project, and (f) to help identify different options for minimizing costs or time required for evaluation while still providing adequately valid information to meet stakeholders’ needs. When the evaluation does not begin until project implementation is already underway, RWE is used to identify and assess the different evaluation design options that can be used within the budget and time constraints and to consider ways to reconstruct baseline data. Attention will be given to assessing the strengths and weaknesses of monitoring and administrative data available from the project and the availability and quality of secondary data from other sources. The feasibility of constructing a comparison group may also be considered. When the evaluation does not begin until toward the end of the project (or when the project has already ended), RWE is used in a similar way to the previous situation except that the design options are more limited because it is no longer possible to directly observe the project implementation process. One of the innovative RWE approaches is to suggest measures that can be taken to strengthen the validity of the findings even up to the point when the draft final evaluation report is being reviewed.

Summary

• Many evaluations are affected by budget, time, and data constraints or by political influences that limit the design options available to the evaluator. We call these the RWE constraints.

• RealWorld evaluators most frequently face one of two main scenarios. The first is when the evaluator is called in at the start of the project but the choice of evaluation design is constrained by budget or time pressures, by technical and administrative difficulties in collecting certain kinds of data, or by pressures from clients and stakeholders.

• The second, and probably the most common, scenario is when the evaluator is not called in until the project has been underway some time or may even be nearing completion. Often, the evaluator is again subject to budget and time constraints and political pressures, but even when budget and time are adequate, it is usually the case that no systematic baseline data have been collected and usually no comparison group has been identified.

• Although evaluation constraints such as lack of evaluation expertise, more limited access to secondary data, and often, less of an evaluation culture may be more apparent in developing countries, many evaluations in developed countries also face similar problems. Consequently, the RealWorld approach is applicable to varying degrees in all countries.
Table 1.2  Who Uses RWE, for What Purposes, and When?

<table>
<thead>
<tr>
<th>When does the evaluation start?</th>
<th>Evaluation practitioners who design or implement the evaluation</th>
<th>Managers and funding agencies</th>
</tr>
</thead>
</table>
| At the beginning of a project (baseline) | • Identify a life-of-project evaluation design that will meet the needs of key stakeholders, given anticipated budget, time, and data constraints  
• Advise management how to reduce costs and time while achieving evaluation objectives  
• Negotiate with managers to relax some of the constraints to reduce some of the threats to validity and adequacy  
• Advise management on plans for a baseline study consistent with evaluation objectives  
• Identify ways to produce the best evaluation under budget, time, and data constraints | • Seek ways to minimize the costs and time required for the proposed evaluation design, including the baseline study  
• Assess the relevance, required level of rigor, and quality of the proposed life-of-project evaluation design |
| During project implementation | • Identify ways for relevant monitoring data to be collected and documented that inform implementers and are relevant for evaluation purposes, given budget, time, and data constraints  
• If there was no baseline, reconstruct baseline data  
• Ensure maximum quality under existing constraints | • Identify ways to strengthen the ongoing monitoring and evaluation (these measures may be directly implemented by project management or funding agencies or recommended to the agency conducting the evaluation)  
• Keep data collection minimized and prioritized on information that informs decision making and learning |
| At the end of the project | • Identify ways to meet evaluation objectives within limitations of budget, time, political considerations, and data availability  
• Use the RWE checklist to identify and deal with threats to validity and reliability  
• Reconstruct baseline data  
• Ensure maximum quality under existing constraints | • Be clear on the purpose of evaluation and the relevant degree of rigor required  
• Identify ways to correct weaknesses in the evaluation within the budget and time constraints and/or be willing to allocate more funds and time to achieve required credibility |

Further Reading


The evaluation guidelines approved by the American Evaluation Association.

An essential reference on evaluation standards.


Case studies of evaluations that had a demonstrable influence on clients and stakeholders and a discussion of the factors determining whether evaluations will be used.


One of the most cited texts on how to design evaluations that will be utilized.


Chapter 2 introduces the evaluator-stakeholder relationship, and Chapter 12 discusses the social context of evaluation and the ethical issues discussed in this chapter.


Discussion of the experiences and issues when other countries in different regions consider adopting and/or adapting U.S. evaluation standards.

**Notes**

1. **Bold** technical terms are defined in the glossary at the end of this book.

2. One of our colleagues who has worked with major U.S. foundations that support community-level initiatives stated that there is a huge unmet need in the United States for material on how to conduct evaluations when working with very limited financial and professional resources. He stated that his “and other foundations make lots of small grants. There is often not enough money in the grants to hire an external consultant. And the recipients of these small grants don't have the capacity to do internal evaluation. The evaluation work done by these nonprofits is usually pretty bad. I don't really know of any materials targeted to this group.”

3. Many surveys include a household roster in which information on age, sex, education, employment status, and the like is collected for each household member. However, even when these data have been collected, analysis is often only conducted at the household level so that many of the data on individual household members are not used.

4. For more information on the IOCE, see http://ioce.net