

CHAPTER 1

Overview: 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 RealWorld evaluations are conducted, and the many different constraints, pressures, and influences under which evaluations are formulated, conducted, disseminated, and used. The RWE approach was originally developed to address four of the most common constraints evaluators face: budget, time, and data constraints, and political influences. Subsequently, issues concerning organizational structures and management and administrative arrangements were added as additional challenges. 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.

1. 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**¹ 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

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

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, Japan, and Australasia report that they operate under similar constraints to those faced by their colleagues in developing countries.² As if these problems were not enough, many evaluations in both developed and developing countries are often conducted in *political* environments in which funding agencies, clients, and key stakeholders have strongly held views on what the “right” evaluation methods should be, 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 selection of project and control groups; the choice of qualitative, quantitative, or mixed-method designs; and whom 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 impact 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 or be replicated elsewhere?

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 credible evidence.³

The pressures of conducting evaluations under budget and time constraints, missing baseline data, and political pressures 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 when faced by these kinds of constraints, accommodating organizational structures and administrative procedures, while at the same time ensuring maximum possible methodological rigor within the particular evaluation context.

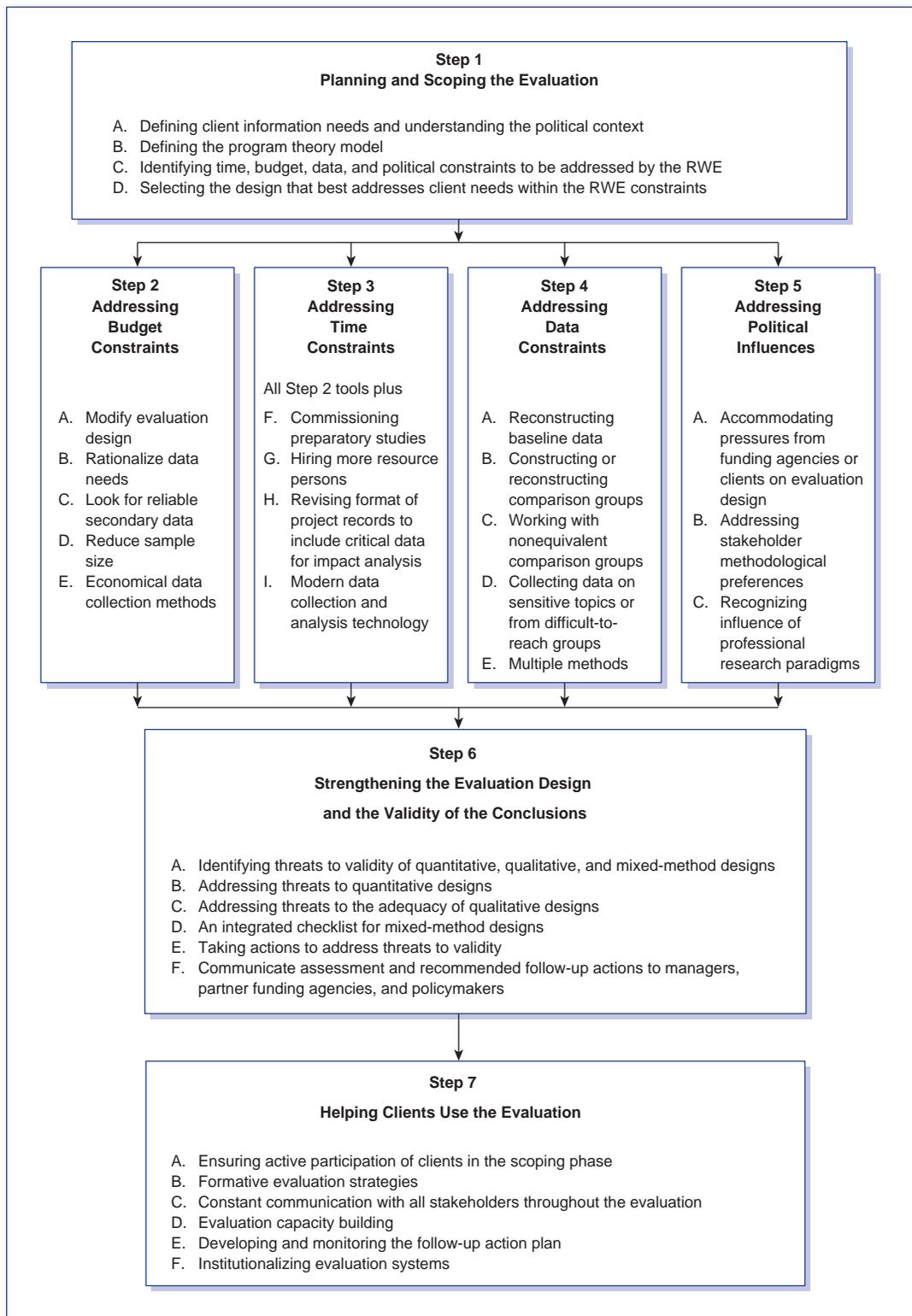
²One of our colleagues who has worked with major U.S. foundations that support community-level initiatives stated that there is a huge unmet need in the United States for material on how to conduct evaluations when working with very limited financial and professional resources. He stated that his “and 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.”

³For two recent publications on the question of credible evidence see: Donaldson, S., C. Christie and M. Mark. 2009. *What Counts as Credible Evidence in Applied Research and Evaluation Practice?* Thousand Oaks: Sage; and Rieper, O., F. Leeuw, and T. Ling. 2010. *The Evidence Book: Concepts, Generation and Use of Evidence*. New Brunswick, NJ, and London: Transaction Publishers.

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*. Before thinking about the evaluation design, it is important to fully understand the purpose of the evaluation, the information needs and expectations of the clients and stakeholders, and the constraints and pressures under which they are working. What is the client's bottom line? What do different stakeholders really want from the evaluation, and how will the results be used? Clarifying these questions is particularly critical for RWE because difficult choices will often have to be made to accommodate budget and time constraints or to recognize the limitations of the available data. This step also includes getting agreement on the articulation of the **program theory or logic model** (discussed in more detail in Chapter 10, which in addition to clarifying the underlying model on which the program is based, can also help identify the critical hypothesis or the linkages in the program implementation model on which the limited evaluation resources should focus). The scoping phase also involves identifying some different basic evaluation design strategies available for addressing the cost, time, and data constraints that a particular evaluation will face and then assessing the strengths and weaknesses (threats to validity and adequacy) of each option. The different design options are then discussed with clients, emphasizing the trade-offs involved in each option, and an agreement is then reached on which design would be most feasible and acceptable to the client. (We get into more detailed coverage of evaluation designs in Chapter 11.)
- Step 2: *Strategies for addressing budget constraints*. How many evaluators have been told by the client, "We really need a rigorous and professional evaluation as it is important to assess impacts, but . . . unfortunately our budget has been cut" Step 2 describes options for reducing costs. These include: simplifying the evaluation design; reducing the amount of data to be collected; making greater use of secondary data; revising the sample design and size, and streamlining data collection and analysis.
- 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*: These include ways to address common problems concerning the lack of important data or data quality and reliability when the evaluation is not commissioned until late in the project cycle. Thus one needs to consider a number of approaches for *reconstructing* baseline data. These include using secondary data sources, recall, key informants, focus groups, construct mapping and participatory group techniques such as PRA. The chapter covers techniques for collecting information on sensitive topics and on difficult-to-reach groups. Many of these groups are the poorest and most vulnerable sectors of the community, and reaching them is often more costly and time-consuming; consequently, there will often be pressures to ignore these difficult questions and inaccessible groups.
- Step 5: *Understanding and coping with political factors influencing how the evaluation is designed, implemented, disseminated or used*: Identify the key actors and their political perspectives and understanding how these affect their orientation to the evaluation and developing strategies to address the political realities without compromising the evaluation. We identify common political issues arising at the outset of an evaluation, during

Figure 1.1 The RealWorld Evaluation Approach



implementation, and during reporting and use of the evaluation, and we propose strategies for addressing all these issues. We also discuss some important professional and ethical issues concerning who should be given information on the evaluation and when. Often, the client would like to limit who sees and is invited to comment on the evaluation draft, whereas the evaluator may feel that the report should be given to the mass media and to the different stakeholder groups potentially affected by the project. We will return to these ethical issues throughout the book.

- Step 6: *Strengthening the Evaluation Design and the Validity of the Conclusions* discusses the analysis of threats to validity affecting the findings and recommendations of the evaluation and how these can be addressed once they have been identified. We present worksheets that have been developed for assessing the validity of QUANT, QUAL and mixed method designs respectively, and for communicating the assessment and the recommended follow-up to managers and policy-makers.
- Step 7: *Helping clients use the evaluation*. Ensure that clients and other key stakeholders 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; and adapting the presentation of findings to the preferred communication style of different stakeholders. On a broader level this also involves helping institutionalize evaluation systems at the sector and national level (see Chapters 17 and 18).

2. The RealWorld Evaluation Context

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

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.

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, no comparable baseline data are available.

In other scenarios, the evaluator may be called in early to help develop the monitoring and evaluation (M&E) plan but may find 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 (a **counterfactual**). 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 trade-offs 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 1.2 REALWORLD EVALUATION CONSTRAINTS IN THE EVALUATION OF AN EDUCATION PROJECT IN EGYPT

This excerpt from a meta-evaluation (review of evaluation methodologies and practices) by CARE International illustrates the many RealWorld constraints facing evaluation in the field.

The evaluators mentioned that the methodology they employed had to be more subjective and anecdotal than they would have desired. The decision not to use their preferred (more rigorous) quantitative design was made due to the limited time available (2 weeks for all data gathering) and the geographic spread, size, and diversity of the target population. Although they felt that a random or weighted sampling would lead to statistically significant or statistically representative findings, it was not realistic given time and other limitations. Instead, the team employed other techniques to try to ensure that the information gathered was as comprehensive and thorough and followed, as closely as possible, accepted approaches to classroom and teacher observation within the very severe time constraints.

The meta-evaluation, recognizing that it is not possible to achieve acceptable standards of evaluation rigor within these time and budget constraints, suggested that CARE may need to re-examine some of its evaluation policies to determine if the desire (or limitations imposed by the donors) to economize on evaluation costs and duration is working at cross-purposes with the level of rigor it hopes to achieve.

SOURCE: Russon (2005:12–13).

3. 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!

Table 1.1 RealWorld Evaluation Scenarios: Conducting Impact Evaluations with Time, Budget, Data, and Political Constraints

<i>The constraints under which the evaluation must be conducted</i>				
Time	Budget	Data	Political	Typical Evaluation Scenarios
X				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.
	X			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.
		X		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 comparison 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.
			X	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.
			X	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.
			X	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.

The constraints under which the evaluation must be conducted

Time	Budget	Data	Political	Typical Evaluation Scenarios
X	X			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.
X		X		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.
	X	X		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.
X	X	X		The evaluator is called in late, is given a limited budget, and has no access to baseline survey data; and no comparison group has been identified.

NOTE: To simplify the table, the possible combinations of political constraints with the other three factors have not been included in the table.

3.1 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 evaluators 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.

BOX 1.3

BUDGET AND TIME CONSTRAINTS HAVE DIFFERENT IMPLICATIONS FOR THE EVALUATION DESIGN

While budget and time constraints often have similar consequences for the evaluation design, in other cases they can require very different approaches. For example, if an evaluation must be completed by a certain date, the process of data collection can often be speeded up by bringing in consultants, hiring more experienced researchers, or increasing the number of **interviewers**. All these measures may require significant budget increases. If, on the other hand, budget is the main constraint, the decision might be made to contract with a local university that would use cheaper though less experienced graduate students who might require more time for data collection because they cannot work fulltime.

3.2 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 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.

3.3 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, or they measure activities and outputs but not outcomes. Project records and other documentary data often suffer from reporting biases or poor record-keeping standards. Even when secondary 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 companies, 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.

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 nonbeneficiaries might create expectations of financial compensation or other benefits (for which the project 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 sensitive topics such as women's empowerment, contraceptive usage, or domestic violence, especially in paternalistic societies, information may be difficult to collect even when funds are available (see Box 1.4). Similar data problems can arise

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.

when working with difficult-to-reach groups such as drug addicts, criminals, ethnic minorities, migrants, or illegal residents.

3.4 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 in which 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**.

4. Additional Organizational and Administrative Challenges

In addition to budget, time, data, and political constraints, all evaluations must conform to the organizational arrangements under which they are commissioned and the administrative procedures of the different agencies involved in commissioning, financing, managing, and using the evaluations. Often there will be a number of different agencies involved in the evaluation, and it is not uncommon for them to have different goals for the evaluation. These may involve the kinds of information to be obtained; the preferred methodology; preferences with respect to the stakeholders that should be involved and who is asked to comment on or approve the evaluation reports; the extent and form in which target populations are or are not involved; and how and to which audiences the evaluation will be disseminated and used. When different international agencies are involved, the basic logistics of arranging joint missions to the country where the evaluation is being conducted can be a major challenge, sometimes delaying implementation for significant periods of time.

Balancing the preferences and operating styles of different agencies can be a major challenge for the evaluation team, particularly in cases where there may be differences of opinion among stakeholders or lack of definition of their respective roles.

Even when only a single agency is involved, their administrative and operating procedures may provide further constraints and challenges for the evaluator. For example, when local counterparts have to be contracted, the procurement procedures of the funding agency or the host government may produce long delays or require the use of contractual procedures that do not work well for a particular evaluation. In other cases, the requirement to prepare an inception report and to delay the start of fieldwork until different departments have commented on the report can cause significant delays in the start of the evaluation. In some cases, the date for the completion of the evaluation report is not changed, despite time lost waiting for feedback, so that the effective time for consultants to work on the evaluation may be significantly reduced. Another common problem is that a fixed amount of time is allowed for fieldwork in every country, despite the fact that in some cases it is well known that considerable numbers of days are likely to be lost while arranging travel to difficult-to-reach parts of the country or while waiting for government clearance for travel. Often, when the evaluation consultants bring up these logistical problems, the evaluation manager will respond, “I entirely agree with you, but unfortunately, this is our administrative policy, so you will just have to do the best you can.”

5. 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 RealWorld 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, 4, and 14).

Second, RWE's seven-step approach 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 a pretest–posttest with statistical counterfactual 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 RealWorld 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!

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.

6. 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—or at least be aware of what an adequate budget and time frame would be required to conduct the kind of evaluation they may have in mind
- To be more fully aware of the various constraints under which an evaluation is to be conducted, and what can be done to address those constraints
- 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 Who Uses RWE, for What Purposes, and When?

When does the evaluation start?	Evaluation practitioners who design or implement the evaluation	Managers and funding agencies
At the beginning of a project (baseline)	<ul style="list-style-type: none"> • 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 (e.g., provide adequate budget and time) to reduce some of the threats to validity and adequacy • Advise management on plans for a baseline study consistent with evaluation objectives 	<ul style="list-style-type: none"> • Be realistic in estimating the budget 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	<ul style="list-style-type: none"> • Identify ways to produce the best evaluation under budget, time, and data constraints • Identify ways for relevant monitoring data to be collected and documented that inform implementers and are relevant for evaluation purposes • If there was no baseline, reconstruct baseline data • Ensure maximum quality under existing constraints 	<ul style="list-style-type: none"> • 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

When does the evaluation start?	Evaluation practitioners who design or implement the evaluation	Managers and funding agencies
At the end of the project	<ul style="list-style-type: none"> • 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 	<ul style="list-style-type: none"> • 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

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 understand client information needs and the political context within which the evaluation will be conducted, (b) to help identify different options for minimizing costs or time required for evaluation while still providing adequately valid information to meet stakeholders’ needs, (c) for deciding what evaluation design would be appropriate, (d) for deciding what data need to be collected by the monitoring system during the implementation of the project, and (e) for deciding how to make the best use of available data.

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 and data collection methodologies 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.
- We have found that the RealWorld Evaluation approach is applicable to varying degrees in all countries.

FURTHER READING

American Evaluation Association. 2004. *Guiding Principles for Evaluators*. Fairhaven, MA: Author. Retrieved from www.eval.org/Publications/GuidingPrinciples.asp

The evaluation guidelines approved by the American Evaluation Association.

Morra-Imas, L. and R. Rist. 2009. *The Road to Results: Designing and Conducting Effective Development Evaluations*. Washington, DC: World Bank.

Covers many of the topics addressed by RWE.

Operations Evaluation Department. 2004. *Influential Evaluations: Evaluations That Improved Performance and Impacts of Development Programs*. Washington, DC: World Bank.

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.

Patton, M. Q. 2008. *Utilization-Focused Evaluation*. 4th ed. Thousand Oaks, CA: Sage.

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

Patton, M.Q. 2011. *Developmental Evaluation: Applying Complexity Concepts to Enhance Innovation and Use*. New York: Guilford.

Discussion of the limitations of conventional impact evaluation designs, particularly for the evaluation of complex and evolving programs.

Rossi, P., M. Lipsey, and H. Freeman. 2004. *Evaluation: A Systematic Approach*. 7th ed. Thousand Oaks, CA: Sage.

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

Russon, C. and G. Russon, eds. 2005. *International Perspectives on Evaluation Standards*. New Directions for Evaluation, No. 104. San Francisco: Jossey-Bass.

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