ONE

WHAT IS EVALUATION?

As promised in the preface, this book's approach is to give you a "bare-bones," nuts-and-bolts guide about how to do an **evaluation.**[1] Although we will not be spending a huge amount of time on evaluation theory, it is certainly a good idea to start with a clear notion of what it is we are getting ourselves into.

BASIC DEFINITIONS

In terms of the evolution of the human race, evaluation is possibly the most important activity that has allowed us to evolve, develop, improve things, and survive in an ever-changing environment. Every time we try something new—a farming method, a manufacturing **process,** a medical treatment, a social change **program,** a new management team, a **policy** or **strategy,** or a new information **system**—it is important to consider its value. Is it better than what we had before? Is it better than the other options we might have chosen? How else might it be improved to push it to the next level? What did we learn from trying it out?

> Professional evaluation is defined as the **systematic** determination of the **quality** or **value** of something (Scriven, 1991).

Things that we might (and should) evaluate systematically include the following[2]:

- **Projects,** programs, or organizations
- Personnel or performance

- Policies or strategies
- Products or services
- Processes or systems
- Proposals, contract bids, or job applications

There is a fundamental logic and methodology that ties together the evaluation of these different kinds of evaluands. For example, some of the key learnings from the evaluation of **products** and personnel often apply to the evaluation of programs and policies and vice versa. This *transdisciplinary* way of thinking about evaluation provides a constant source of innovative ideas for improving how we evaluate. For this reason, this book contains illustrative examples drawn from a variety of settings and evaluation tasks.

Evaluations are generally conducted for one or two main reasons: to find areas for improvement and/or to generate an **assessment** of *overall* quality or value (usually for reporting or decision-making purposes). Defining the nature of the evaluation question is key to choosing the right methodology.

Some other terms that appear regularly in this book are merit, worth, quality, and value. Scriven (1991) defines these as follows:

> **Merit** is the "intrinsic" value of something; the term is used interchangeably with *quality.*
>
> **Worth** is the value of something to an individual, an organization, an institution, or a collective; the term is used interchangeably with *value.*

This distinction might seem to be a fine one, but it can come in handy. For example, in the evaluation of products, **services,** and programs, it is important to critically consider the extent to which improvements in *quality* (e.g., adding more "bells and whistles") would actually provide enough incremental *value* for the individuals and/or organization concerned to justify their cost.

More often than not in evaluation, we are looking at whether something is "worth" buying, continuing to fund, enrolling in, or implementing on a broader scale. Accordingly, most "big picture" evaluation questions are questions of value (to **recipients/**users, funders/taxpayers, and other relevant parties) rather than of pure merit. There are exceptions, however, and that is why I have kept both considerations in play.

FITTING EVALUATION APPROACH TO PURPOSE

For any given evaluation, a range of possible approaches is available to the practitioner and the **client.** The option that is most often discussed in evaluation circles pertains to whether an evaluation should be conducted independently (i.e., by one or more outside contractors) or whether the program or product designers or staff should be heavily involved in the evaluation process.

If the primary purpose of the evaluation is for *accountability,* it is often important to have an **independent evaluation** conducted (i.e., nobody on the evaluation team should have a significant vested interest in whether the results are good or bad). This is not always a requirement (e.g., managers in all kinds of organizations frequently report on the performance of their own units, products, and/or people), but this credibility or independence issue is definitely one to consider when choosing how to handle an accountability-focused evaluation.

There are many cases where independence is not essential, but building **organizational learning capacity** is key; that is, a primary goal is to improve **organizational learning** (i.e., the organization's ability to learn from its successes and failures). In such cases, an evaluation can (and should) be conducted with a degree of **stakeholder** participation. Many high-quality professional evaluations are conducted collaboratively with organizational staff, internal human resources consultants, managers, **customers** or recipients, or a combination of these groups.

A **learning organization** is one that acquires, creates, evaluates, and disseminates knowledge—and uses that knowledge to improve itself—more effectively than do most organizations. The best learning organizations tend to use both independent and **participatory evaluations** to build learning capacity, gather multiple perspectives on how they are doing, and keep themselves honest (Davidson, 2003).

THE STEPS INVOLVED

Whether the evaluation is conducted independently or in a participatory mode, it is important to begin with a clear understanding of what evaluation is and what kinds of evaluation questions need to be answered in a particular case. Next, one needs to identify relevant "values," collect appropriate data, and systematically combine the values with the descriptive data to convey, in a useful and concise way, defensible answers to the key evaluation questions (see Exhibit 1.1).

Exhibit 1.1 Overview of the Book's Step-by-Step Approach to Evaluation

CHAPTER 1	Understanding the basics about evaluation
	↓
CHAPTER 2	Defining the main purposes of the evaluation and the "big picture" questions that need answers
	↓
CHAPTER 3	Identifying the evaluative **criteria** (using needs assessment and other techniques)
	↓
CHAPTER 4	Organizing the list of criteria and choosing sources of evidence (mixed method data)
	↓
CHAPTER 5	Dealing with the causation issue: how to tell the difference between **outcomes** or **effects** and coincidental changes not caused by the evaluand
	↓
CHAPTER 6	Values in evaluation: understanding which values should legitimately be applied in an evaluation and how to navigate the different kinds of "subjectivity"
	↓
CHAPTER 7	Importance weighting: figuring out which criteria are the most important
	↓
CHAPTER 8	Merit determination: figuring out how well your evaluand has done on the criteria (excellent? good? satisfactory? mediocre? unacceptable?)
	↓
CHAPTER 9	Synthesis methodology: systematic methods for condensing evaluative findings
	↓
CHAPTER 10	Putting it all together: fitting the pieces into the Key Evaluation Checklist framework
	↓
CHAPTER 11	Meta-evaluation: how to figure out whether your (or someone else's) evaluation is any good

THE INGREDIENTS OF A GOOD EVALUATION

The overarching framework used for planning and conducting an evaluation and presenting its results is Scriven's (2003) Key Evaluation Checklist (KEC) with a few modifications and simplifications. This is a guiding framework for the evaluation team members (be they organizational members, **external evaluators,** or a mix) to make sure that all important ingredients that will allow valid **evaluative conclusions** to be drawn are included.

The KEC should be thought of both as a checklist of necessary ingredients to include in a solid evaluation and as a framework to help guide evaluation planning and reporting. Because the KEC was designed primarily for application to program evaluation, some of the points might need reframing when the KEC is used for other **evaluands** or **evaluees** (the term used in personnel evaluation). In a posting to a listserv on November 16, 2002, Scriven describes how and why the KEC was developed:

> The Key Evaluation Checklist evolved out of the work of a committee set up by the U.S. Office of Education which was to hand out money to disseminate the best educational products to come out of the chain of Federal Labs and R&D Centers (some of which still exist). The submissions were supposed to have supporting evidence, but these documents struck me as frequently making a few similar mistakes (of omission, mostly). I started making a list of the recurring holes, i.e., the missing elements, and finished up with a list of what was needed in a good proof of merit, a list which we used and improved.

A brief overview of the KEC is shown in Exhibit 1.2. Each line of KEC checkpoints represents another layer in the evaluation. We begin with the Preliminaries (Checkpoints I–III), which give us some basic information about the evaluand and the evaluation. From there, we move to the Foundations (Checkpoints 1–5), which provide the basic ingredients we need, that is, descriptive information about the program, who it serves (or should serve), and the values we will apply to evaluate it. The third level, which Scriven called the Sub-evaluations (Checkpoints 6–10), includes all of the explicitly evaluative elements in an evaluation (i.e., where we apply values to descriptive facts to derive evaluative conclusions at the analytical level). Finally, we reach the Conclusions section (Checkpoints 11–15), which includes overall answers to the evaluation questions plus some follow-up elements.

Exhibit 1.2 The Key Evaluation Checklist (modified from Scriven's 2003 version)

Preliminaries	**I. Executive Summary** One- to two-page overview of the evaluand and findings	**II. Preface** Who asked for this evaluation and why? What are the main evaluation questions? Who are the main audiences?		**III. Methodology** What is the overall design of the evaluation (e.g., quasi-experimental, participatory, goal free) and (briefly) why?	
Foundations	**1. Background and Context** Why did this program or product come into existence in the first place?	**2. Descriptions and Definitions** Describe the evaluand in enough detail so that virtually anyone can understand what it is and what it does.	**3. Consumers** Who are the actual or potential recipients or impactees of the program (e.g., demographics)?	**4. Resources** What resources are (or were) available to create, maintain, and help the program or policy succeed?	**5. Values** On what basis will you determine whether the evaluand is of high quality, valuable, and so forth? Where will you get the criteria, and how will you determine "how good is good"?
Sub-Evaluations	**6. Process Evaluation** How good, valuable, or efficient is the evaluand's content (design) and implementation (delivery)?	**7. Outcome Evaluation** How good or valuable are the **impacts** (intended and unintended) on immediate recipients and other impactees?	**8 & 9. Comparative Cost-Effectiveness** How costly is this evaluand to consumers, funders, staff, and the like, compared with alternative uses of the available resources that might feasibly have achieved outcomes of similar or greater value? Are the costs excessive, quite high, just acceptable, or very reasonable?		**10. Exportability** What elements of the evaluand (e.g., innovative design or approach) might make it potentially valuable or a significant contribution or advance in another setting?

11. Overall Significance
Draw on all of the information in Checkpoints 6 through 10 to answer the main evaluation questions (e.g., What are the main areas where the evaluand is doing well, and where is it lacking? Is this the most cost-effective use of the available resources to address the identified needs without excessive adverse impact?).

Conclusions

12. Recommendations and Explanations [optional] A more in-depth analysis of why/how things went right/wrong, perhaps including recommendations for improvement	**13. Responsibilities** [optional] A more in-depth analysis of exactly who or what was responsible for good or bad results (Note: This is very tricky and is usually not the kind of territory you want to get into unless you are highly skilled.)	**14. Reporting and Follow-up** Who will receive copies of the evaluation report and in what form (e.g., written, oral, detailed versions, executive summary)?	**15. Meta-evaluation** A critical assessment of the strengths and weaknesses of the evaluation itself (e.g., How well were all of the Key Evaluation Checklist checkpoints covered?) and conclusions about its overall utility, accuracy or validity, feasibility, and propriety (see the Program Evaluation Standards for details)

SOURCE: Adapted and reprinted by permission by Michael Scriven.

Scriven (1991) asserts that evaluations should generally cover all of these checkpoints (except for Checkpoints 12 and 13, which are optional) to draw valid conclusions. Each point listed in the KEC is backed by a carefully thought-out rationale showing why omission of the particular point is likely to compromise the validity of conclusions. Although this should not be taken to mean that all checkpoints must *always* be included in all evaluations, it does mean that decisions to omit certain elements should be carefully justified. This is particularly important for Checkpoints 5 through 9 and 11, which form the core of the evaluation.

Obviously, there is a lot more to the KEC than one can fit on a one-page summary. Throughout this book, we work through many of the KEC checkpoints, paying particular attention to the truly evaluative ones (from Checkpoint 5 [Values] through Checkpoint 11 [Overall Significance]), which is where **evaluation-specific logic and methodology** come into play. Later, in Chapter 10, we return to the KEC and show how all of the information we have covered fits into the big picture.

It is important to note that the KEC can be applied to a participatory or **collaborative evaluation** just as easily as it fits into the conduct of an independent evaluation being done for accountability. Whether the evaluation is a facilitated collaborative effort or not, the evaluation team members (be they external or **internal evaluators**) still need some guidelines for figuring out what should go into an evaluation to make sure that it provides the most accurate answers to the most important questions.

IDENTIFYING THE EVALUAND, ITS BACKGROUND, AND ITS CONTEXT

Before we plunge into the nuts and bolts of evaluation design, it is a good idea to first clarify what it is you plan to evaluate (i.e., your evaluand). This might seem like an incredibly basic question, but it trips up a lot of people. For your first evaluation, it is important to choose something manageable to which you could reasonably expect to gain access.

A clear and accurate description of your evaluand should appear under Checkpoint 2 (Descriptions and Definitions) of the KEC and should also have a brief mention in your evaluation report's Executive Summary (Checkpoint I). Equally important is to gain a solid understanding of the evaluand's Background

and Context (Checkpoint 1). These three checkpoints are the focus of this chapter (Exhibit 1.3).

Exhibit 1.3 The Checkpoints Where the Evaluand, Its Background, and Its Context Are Described

I. Executive Summary One- to two-page overview of the evaluand and findings	
1. Background and Context Why did this program or product come into existence in the first place?	**2. Descriptions and Definitions** Describe the evaluand in enough detail so that virtually anyone can understand what it is and what it does.

When completing the Descriptions and Definitions checkpoint, the evaluation team should not just use brochures or Web sites to find out what the evaluand is *supposed* to be like; instead, the team should describe it as it *really* is. This usually involves, at a minimum, a firsthand visit and some interviewing of key stakeholders. The information presented under this checkpoint should be purely descriptive in nature; that is, you should not make comments here about the merits of the evaluand or its design.

At the same time, the evaluation team should conduct a preliminary investigation to find out what it was that led to the development of the evaluand in the first place and any underlying rationale for how or why it was intended to address the original need, problem, or issue. This information will go under the Background and Context checkpoint.

ADVICE FOR CHOOSING YOUR FIRST EVALUATION PROJECT

Whether you are attending an evaluation class or just trying to figure out for yourself how to put together an evaluation, a key part of the process will be working through an example of your own as you go through this book. This process can be made easier or harder depending on what you choose as your

first project. Here are a few tips for choosing a project that will allow you to get the most out of this book:

1. Make life easier for yourself by choosing an intervention, program, or the like that is designed to benefit *people* in some way. In this book, we talk a lot about assessing the needs of recipients and impactees, so it helps if these are (a) a clearly defined group and (b) human. Try to avoid abstract evaluands or very complex systems. You can get into these later.

2. For this exercise, the evaluand should be a "live" program or intervention that is currently in existence and that you can go and see with your own eyes. Inanimate objects, distant programs, things you have only seen on the Internet, and things that no longer exist are not good ideas for first projects because they make it harder to get access to "the clients" (an important part of getting a feel for evaluation).

3. Do not tackle something that could have political ramifications for you (or for your instructor if you are a student). Examples might include your boss's pet project, another professor's class, and university administrative systems (these are too complex anyway).

4. It is better if you can choose something of which you are *not* a current, recent, or future recipient or consumer (e.g., a graduate program in which you studied, a workshop you attended). Although the "inside perspective" might seem to be advantageous at first, people tend to get way too distracted with their own personal perspectives or agendas and end up missing a lot of important issues and not doing so well on their evaluation projects.

If you are already working in evaluation, you no doubt have plenty of evaluands from which to choose. For students who need to track down an evaluand, the following are some ideas for evaluands to consider as first-time evaluation projects:

- A community health program
- A workplace wellness initiative
- A school counseling service
- An internship program
- An AIDS prevention program
- A jail diversion program for first-time offenders

- A training program or workshop
- A summer camp
- A performance management and reward system
- A mentoring program
- A distance learning course
- A fast-track program for high-potential employees or students
- An organizational change intervention
- A distribution system for a particular product
- An executive recruitment service

After you have identified an appropriate evaluand, work through the exercises at the end of this chapter. These will yield a draft of Checkpoints 1 and 2 of the KEC.

NOTES

1. Definitions of the key terms used in this book may also be found in the Glossary.
2. This list is an elaboration of Scriven's (1991) list of the Big Six categories of evaluand, expanded here to be more inclusive of the various terminology used across different fields.

ADDITIONAL READINGS

Entries in Scriven's (1991) *Evaluation Thesaurus:*

- Consultant
- Contextually evaluative
- Cost-free evaluation
- Criterion
- Descriptive
- Evaluand
- Evaluate/Evaluation
- Introduction: The nature of evaluation

Davidson, E. J. (2002). The discipline of evaluation: A helicopter tour for I/O psychologists. *The Industrial–Organizational Psychologist, 40*(2), 31–35. Available online: http://siop.org/tip/tip.html

Fetterman, D. M. (2000). *Foundations of empowerment evaluation.* Thousand Oaks, CA: Sage.

Patton, M. Q. (1997). *Utilization-focused evaluation* (3rd ed.). Thousand Oaks, CA: Sage.

Rose, D. S., & Davidson, E. J. (2003). Overview of program evaluation. In J. E. Edwards, J. C. Scott, & N. S. Raju (Eds.), *The human resources program evaluation handbook* (pp. 3–26). Thousand Oaks, CA: Sage.

Scriven, M. (1993). *Hard-won lessons in program evaluation* (New Directions for Program Evaluation, No. 58). San Francisco, CA: Jossey-Bass.

EXERCISES

1. Clearly identify your evaluand, that is, what you plan to evaluate (or to just write an evaluation plan for). Explain, on half a page or less, what it is (be sure to include the points that follow).
   a. Is it a program, policy, product, service, system, or something else?
   b. Who exactly does it (or should it) serve (e.g., customers, **consumers,** recipients, people in need, **target market**)?
   c. Who is in charge of it?

2. Try explaining what you have written to a colleague to make sure that it makes sense. (Common mistakes here include prematurely specifying what criteria you plan to use [e.g., saying that you are going to evaluate something "in terms of X"], choosing something far too complex for your first evaluation [e.g., having two or more nested evaluands such as a project within a program within a system], and commenting on the merits of the evaluand [at this stage you should be purely descriptive].)

3. Interview key stakeholders to gain an understanding of your evaluand's background and context.
   a. Find out why your evaluand came into existence in the first place—to address what need or problem?
   b. What rationale can you find (from documentation, interviews, or other methods) that reveals how or why your evaluand was supposed to meet this need or address this problem?
   c. What other events were happening at the time (e.g., political environment, legislation, technological developments, cultural issues), and how did they lead to the development of your evaluand at that time?