SIX

"VALUES" IN EVALUATION

Right at the beginning of this book, it was noted that the special thing about evaluation—the part that makes it different from (and harder than) descriptive research—is that it involves more than simply collecting data and presenting results in "value-neutral" (i.e., purely descriptive) terms. E-*valu*-ation involves applying **values** to descriptive data so as to say something explicit about the quality or value of the evaluand in a particular context. Our goal is to do this with a level of certainty that is appropriate for those who might potentially make decisions on the basis of our findings.

The content of this chapter is relevant not only to the Values checkpoint but also to all of the Sub-evaluation checkpoints (6–10) and to the Overall Significance checkpoint of the Key Evaluation Checklist (KEC) (Exhibit 6.1).[1] Under the Values checkpoint, the evaluation team needs to identify broadly the sources of value that were used to determine what should be considered "good," "valuable," or "worthwhile" for this particular evaluand. Then those values are applied to the descriptive data collected about process, outcomes, costs, comparisons, and exportability to draw explicitly evaluative conclusions within each of those checkpoints. Finally, the Overall Significance checkpoint is where the evaluation team needs to weigh all of the strengths and weaknesses and to draw overall conclusions about the evaluand—another task that requires the application of values.

To arrive at explicitly evaluative conclusions, at least three important methodological tasks are required that are not found in purely descriptive scientific research: (a) importance weighting, (b) merit determination, and (c) synthesis. These are the methodology topics we cover in the next few chapters.

Exhibit 6.1 The KEC Checkpoints Where Values Are Most Relevant

5. Values
On what basis will you determine whether the evaluand is of high quality or value? Where will you get the criteria, and how will you determine "how good is good"?

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 (both intended and unintended) on immediate recipients and other impactees?

8 & 9. Comparative Cost-Effectiveness
How costly is this evaluand to consumers, funders, staff, and so forth, 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, including the following. 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?

But before moving further into the nuts and bolts of evaluation-specific methodology, it is important to get a solid understanding of the following:

- The nature of the controversy surrounding this part of evaluation that centers on the question, "Aren't values all just subjective?"
- An important source of the disagreement on this issue, that is, the failure to clearly distinguish among three distinct kinds of subjectivity

- The tensions between taking a hard subjectivist line on this issue and the common sense of people (clients) trying to make informed choices and decisions in the real world
- Where the "values" in a solid evaluation really come from

At the end of this chapter, a pragmatic stance on this controversial issue is presented. The main purpose here is to find a way for most of us to wade out of the "values quagmire" and get on with the evaluative tasks outlined in the subsequent chapters.

THE CONTROVERSY

This is the part of the book where we really start getting into controversial territory. There are both professional evaluators and others who strongly believe that the steps from descriptive data to explicitly evaluative conclusions should not be tackled at all by the evaluator. For example, here is a fairly typical line of argument from some prominent applied researchers who subscribe to what Guba and Lincoln (1989) call the "scientific paradigm":

> This book [*Assessing Organizational Change*] is largely silent on the issue of combining outcomes from different domains in order to reach an overall conclusion about the effectiveness of a change effort. This is by design. The decision was made early on to simply report how the organization had changed on a wide array of outcome measures. No common metric was developed, nor was a weighting system developed that argued that gains in some measures are more important than gains in others. The rationale for not doing this is simple and to us persuasive. It is that different constituents value outcomes differently, and thus it is best to let interested parties reach their own overall conclusions. There are also practical problems in trying to translate diverse outcomes to a common metric. (Lawler, Seashore, & Mirvis, 1983, p. 542)

As we can see, Lawler and colleagues (1983) believe that evaluators should steer clear of assigning value or importance to various findings and instead should let stakeholders make their own determinations. In addition, they appear to be asserting that there is simply *no available methodology* that will yield valid and defensible findings, as evidenced by their closing sentence: "There are also practical problems in trying to translate diverse outcomes to a common metric."

Evaluators who subscribe to the **constructivist/interpretivist paradigm** (roughly the opposite of the scientific paradigm) take a somewhat different position. Although they agree with Lawler and colleagues (1983) that evaluative conclusions can never amount to anything more than the application of personal values in the interpretation of data, they tend to embrace this idea and build it into the design rather than leaving the findings in value-neutral terms (Guba & Lincoln, 1989). This view of evaluation sees the drawing of evaluative conclusions as a sensemaking process in which multiple stakeholders participate.

Despite their very different worldviews, evaluators who subscribe to the constructivist/interpretivist and scientific paradigms seem to share two underlying assumptions:

- All evaluative claims (about the value of certain outcomes or attributes, their relative importance, and what mixed results indicate about overall value) are arrived at "subjectively"; more specifically, the values that are applied to descriptive facts (data) to arrive at evaluative conclusions are *personal values.*
- Allowing stakeholders to make up their own minds, either individually or collectively, is the only valid way in which evaluative conclusions can be drawn at all.

In the remainder of this chapter, we critically examine these assumptions and consider some alternative lines of thought. The first task is to clarify a very common source of confusion about the different meanings of the term ***subjective.***

THE THREE TYPES OF SUBJECTIVITY

Evaluation is an intensely political activity that is viewed by many as a threat. As a result, evaluators very often run into situations where people are challenging or attacking their work. One of the most common attacks goes directly to the heart of the subjectivity issue and includes statements such as the following:

- "Well, that's just *your* opinion about the program."
- "Yes, but *who defines* 'acceptable performance'?"

- "Who are you to impose *your values* on our program?"
- "Evaluation is just so subjective!"

One important factor here is being clear about the different types of subjectivity that may exist in an evaluation and being sure not to confuse them. Scriven (1991) makes a very useful distinction among three different kinds of subjectivity, two of which might legitimately appear in an evaluation and one of which should not:

- *Subjective 1:* Arbitrary, idiosyncratic, unreliable, and/or highly personal (i.e., based purely on personal preferences or inappropriate cultural biases); the kind of subjectivity that has no place in serious evaluation
- *Subjective 2:* Assessment or interpretation by a person, rather than by a machine or measurement device, of something external to the person (e.g., expert judgment of a trainee's skills in the effective facilitation of discussions)
- *Subjective 3:* About a person's inner life or experiences (e.g., headaches, fears, beliefs, emotions, stress levels, aspirations), all absolutely real but not usually independently verifiable

Subjective 1: Inappropriate Application of Personal or Cultural Preferences/Biases

When people complain about an evaluation being "so subjective," the insinuation is usually that the conclusions were of the Subjective 1 type, that is, arbitrary or idiosyncratic. And sometimes that is true. Some of the most important examples of the inappropriate application of personal values have appeared in cross-cultural and gender-related evaluation where those doing the evaluations used preconceived frameworks or biases that failed to capture what was really important and/or seriously disadvantaged certain individuals or groups. Some examples include the following:

- Women managers generally are evaluated more negatively when they use an autocratic management style (which is generally viewed as more "masculine") than are male counterparts who exhibit the same behavior.

- Medical researchers evaluating a traditional treatment being used by witch doctors in Africa concluded, based on their observations of patient outcomes, that the treatment was indeed effective. They dutifully recorded the ingredients in the supplied medicine, which they brought back for use in the West. Unfortunately, the exact same concoction was found to be ineffective when given to Western patients. In this case, the researchers' cultural lens had led them to incorrectly identify the evaluand as being only the medication. What they had failed to understand was that the entire treatment also involved elaborate rituals performed by the witch doctors and required belief in both the rituals and the medication on the part of the patients.

In evaluation in particular, there is a grave danger that the application of inappropriate personal or cultural preferences or biases may lead to faulty conclusions and, therefore, misguided decisions. As evaluators, we need to be open to this possibility. If inappropriate values have crept in, we need to track them down and weed them out. Conversely, if important and relevant values (e.g., those that are relevant to the cultural context) have been excluded, we need to identify them and bring them to bear in the evaluation. If we are unfamiliar with the context to the extent that it would limit our ability to clearly identify or understand those values, it is our responsibility to bring onto the evaluation team people who can help us with that. With careful attention to these issues, the result should be an **objective** evaluation, that is, one that is free of inappropriate personal or cultural biases.

Subjective 2: Informed Judgment

In contrast to the first example, evaluations are often accused of being subjective (again, with the insinuation being Subjective 1, i.e., arbitrary or idiosyncratic) simply because there has been some use of human judgment, assessment, or interpretation (i.e., Subjective 2). For example, informed or expert judgment is used when an experienced facilitator rates the performance of trainees learning the art of facilitation on videotaped role-plays.

The key response in this case is to clarify the distinction between the valid use of expert judgment (Subjective 2) and the inappropriate use of personal preferences and biases in an evaluation (Subjective 1). Again, it is possible for expert judgment to be sloppy, and the evaluator should be extremely vigilant about this possibility.

In addition, even well-founded expert judgments need triangulation (i.e., verification of the findings from another source of data and/or from another informant) to ensure that they are **robust.** If the judgments are supported with solid explanations, if other evidence also points independently to the same conclusion, and if the details of any accusations regarding the inappropriate intrusion of personal preferences are thoroughly investigated and found to be baseless, the accusation of subjectivity (in the Subjective 1 sense) can be refuted.

Subjective 3: "About My Life"

There is a third sense in which the term *subjective* is used, and that is in reference to people's inner lives or experiences (e.g., headaches, fears, beliefs, emotions, stress levels, aspirations). These are the kinds of things that are usually not independently verifiable. Nevertheless, one seldom hears accusations that reports of headaches are "just subjective."

This type of subjective data quite often has a legitimate place in evaluation. The most common case is where the outcomes themselves are internal states such as confidence, stress, anxiety, and sense of cultural identity. Thus, subjective measures are essential for the evaluation of any program that produces outcomes such as these.

An extension of Scriven's (1991) third category of subjectivity is the intersubjective experience of a community or group (e.g., culture, sensemaking). These are aspects of group, community, or organizational life that cannot exist independently of people's shared perceptions and intersubjective sensemaking.

The Red Herring: Subjective Measures Versus Objective Measures

An additional source of confusion with the term *subjective* arises when an evaluation is accused of having too many "soft" or subjective measures and not enough "hard" data. Many of the social sciences use the term *objective measures* to refer to quantitative data relating to independently verifiable phenomena, whereas the term *subjective measures* refers to quantitative data relating to Subjective 2 judgments, Subjective 3 experiences, or qualitative data about anything.

This terminology is unfortunate because it leads people to believe that only quantitative measures of independently verifiable phenomena are really rigorous enough to use in an evaluation. However, it is usually essential

to incorporate a far wider range of data into good evaluations, including expert judgment, perceptions, subjective experiences, and any number of aspects of the evaluand that are best assessed using qualitative methods. The terms *subjective* and *objective* are not used in the hard/soft data sense in this book.

THE TENSIONS BETWEEN SUBJECTIVISM AND COMMON SENSE

Accusations about the subjectivity of evaluation assert that all evaluative claims are based on personal values and preferences. These accusations are reflective of a philosophical doctrine called *subjectivism.* According to this doctrine, all evaluative statements (e.g., saying that something is excellent, a waste of money, or better than the alternative) can be neither true nor false; rather, they are simply statements about the feelings of the people making the statements.[2]

For example, subjectivists believe that the statement "This is an excellent school" is not—and can never be—a statement of fact about the school, no matter what evidence is brought forward to support that claim. It can only be a statement of opinion. It is equivalent to the evaluator saying "I like this school" or reporting that lots of people say that they like it.

Although subjectivism is fashionable in academic circles with both quantitative and qualitative researchers right across the social sciences, it has never really caught on in the real world. People moving into new towns frequently ask around about good schools, honest mechanics, competent doctors, and so forth. These questions are not simply about which schools, mechanics, and doctors their new friends, colleagues, and neighbors happen to like personally; they are about which ones really are known to be the best in terms of their performance. In real life, people know for a fact that there is such a thing as good or poor quality or value, and they often seek out this information to help them make good decisions.

As evaluators working in the real world, we have clients and other "right-to-know" audiences who need good answers to their legitimate questions about quality. A community activist working on a shoestring will need to know whether there is any better way in which to channel limited resources to

achieve a more powerful and positive impact in the community. A manager might need to know whether the new performance bonus system is worth the money being spent on it.

When people ask for—and pay for—answers to these questions, they typically are not interested in hearing about how subjective evaluation is. Nor are they interested in being told that they will need to wade through statistics and/or narratives, or will need to personally participate in a lengthy evaluation process, to find their own answers. Clients often expect, quite reasonably, that the evaluation team (which may or may not include organizational members) not only should present them with descriptive data but also should finish the job by providing some defensible conclusions about the quality or value of the evaluand (or areas of strength and weakness).[3]

This is not to say that all evaluation tasks are straightforward enough to produce single defensible conclusions. In some cases, it is quite debatable whether and how certain values should be applied. For example, is it more valuable to successfully place into employment 10% of a group of long-term unemployed individuals with poor educational qualifications or to successfully place 50% of a similar-sized group whose members were better qualified and only briefly out of work? To answer such a question requires considerable thought and additional analysis (e.g., long-term impact on recipients, their families, and society; the crime rate). In the end, it might not be possible to find a clear answer based on an analysis of potential downstream impacts. Part of the answer might well rest on what the particular community or society values personally or as a collective. Or the evaluation team might need to report on the issue by raising it as a point for discussion rather than making a call on it. In such cases, it is quite legitimate to leave some shades of gray for the readers to sort out by themselves. But that option should not be abused to the point where the evaluator no longer does the groundwork on relevant values.

It is important to stress that there are many cases where we *can* draw well-supported, defensible conclusions about the quality or value of evaluands to a level of certainty that is appropriate for the particular decision-making context. The most important path to being able to achieve this is the very careful identification and application of relevant values. In particular, we need to make sure that this part of the evaluation goes a long way beyond collecting other people's opinions or asking people what their values are.

WHERE DO THE "VALUES" IN AN EVALUATION COME FROM?

When most people hear the word *values,* they think of the *personal* values that each of us holds individually. But as we have seen in earlier chapters, there are several other sources of values that are not based simply on personal preferences but that help us to figure out the extent to which a certain set of attributes or outcomes is *valuable* in this particular context.

Let's recap from earlier chapters what outcomes and attributes make an evaluand meritorious or valuable:

- The most important recipient, consumer, or user needs have been better met by the evaluand.
- There has been a noticeable positive impact by the evaluand on siblings, families, the community, the organization, and so forth.
- The content or design of the evaluand was scientifically sound and matched to consumer needs.
- The implementation or delivery of the evaluand was in compliance with all legal, ethical, and professional standards.
- There was a minimum of wastage or inefficiency in the time, money, and other resources spent on the evaluand.
- The evaluand was substantially more cost-effective than anything else that could feasibly have been produced or delivered with the available resources.
- The evaluand had other features or attributes that enhanced the experience of the consumers and others.

Now recall the subjectivist assumptions listed earlier in this chapter:

- All evaluative claims are arrived at subjectively; that is, the values that are applied to descriptive facts (data) to arrive at evaluative conclusions are personal values.
- Evaluative conclusions can be drawn only by allowing stakeholders to make up their own minds, either individually or collectively.

It is certainly true that some assessment or interpretation by a human(s) is likely to form part of the evaluation. However, in a good evaluation, this is limited to judgment by well-qualified experts.

The main accusation of subjectivity came from the second part of the first point, which implies that all evaluative conclusions are based on personal values. In fact, none of the sources of value listed previously represents arbitrary or idiosyncratic personal preferences, especially not on the part of the evaluation team. All of the sources of value came from valid sources—actual needs, ethics, professional standards, and so forth.

For those readers who still see themselves as dyed-in-the-wool value relativists, one could also say that the values on which a solid evaluation is based are defensible insofar as there is sufficiently widespread agreement within the relevant context about those values that they can reasonably be treated as givens. But what such a line implies, for example, is that if we were evaluating an orphanage, whether or not children are fed nutritious food, kept warm, treated for illness, and not sexually abused are only defined as "good" things because *enough people agree* that these qualities are good, not because they actually *are* good. However, common sense tells us unequivocally that these are demonstrably part of what it means to be a good orphanage.

Once we realize that a very substantial proportion of the sources of value come from demonstrably defensible sources, we can move away from the notion that ascribing value must inevitably be a job exclusively for stakeholders. Stakeholder input will certainly be useful in the interpretation of certain findings, particularly because some people are in an excellent position to inform the evaluation team about the issues at hand. But now that it is clear that the sources of value are not just personal, we no longer need to view their application as an act of personal judgment.

To be sure, the evaluation team is made up of people, and people are fallible. Again, we must remember three things. First, we are not usually (if ever) looking for 100% certainty in our conclusions; rather, we are seeking just enough to meet the requirements for certainty in the relevant decision-making context. Second, the more the evaluation team keeps this part of the process carefully documented and justified, the less likely its members are to slip carelessly into sloppy evaluation. Third, all evaluations, especially high-stakes ones, should be meta-evaluated (i.e., the evaluations themselves should be evaluated), preferably by independent evaluation experts. The more transparent the application of values, the easier it will be for a meta-evaluator to see any areas of slippage such as false assumptions. That is our goal as we move into the next few chapters on evaluation-specific methodology.

NOTES

1. Note that the Methodology checkpoint has not been included in Exhibit 6.1. People often argue that methodological choices are very "value laden" and that values or preferences used to choose these should be made explicit. This is very good advice. However, this discussion really falls within the realm of *intradisciplinary* evaluation (Scriven, 1991) in that it pertains to what makes the evaluation design and methods good or bad rather than what makes the evaluand good or bad. This chapter focuses on the latter, whereas the former is addressed throughout the book as we evaluate the merits of different evaluation design issues and strategies and how to match them to the information needs of stakeholders.

2. The more extreme forms of subjectivism, which hold that there is no such thing as *any* kind of objective fact (whether descriptive or evaluative), are being excluded here. Although that is arguably included by purists in the thinking of this doctrine, the more frequent—and important—point of contention in evaluation is about the subjectivity of the values part.

3. Some clients specifically ask that evaluation teams *not* draw explicitly evaluative conclusions because the clients prefer to make up their own minds about the findings and (if necessary) to argue for those interpretations within their organizations. That may be a reasonable request in some cases, but it is often true that a particular stakeholder has insufficient time or expertise to wade through all of the relevant findings and often has a vested interest in a particular interpretation. Therefore, the evaluation team needs to carefully consider the likelihood of evaluation misuse when making a decision not to provide explicitly evaluative conclusions.

ADDITIONAL READINGS

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

- Bias
- Illicit values
- Neutral
- Objective
- Social science approach
- Subjectivity
- Value-free doctrine
- Value judgment
- Valuephobia

Eagly, A., Makhijani, M., & Klonsky, B. (1992). Gender and the evaluation of leaders: A meta-analysis. *Psychological Bulletin, 111*, 3–22.

Guba, E. G. (Ed.). (1990). *The paradigm dialog.* Newbury Park, CA: Sage.

Guba, E. G., & Lincoln, Y. S. (2001). *Guidelines and checklist for constructivist (a.k.a. fourth generation) evaluation.* Available online: http://evaluation.wmich.edu/checklists/

House, E. R., & Howe, K. R. (1999). *Values in evaluation and social research.* Thousand Oaks, CA: Sage.

Scriven, M. (1975). *Evaluation bias and its control.* Occasional paper series #4, The Evaluation Center, Western Michigan University. Available online: http://www.wmich.edu/evalctr/pubs/ops/ops04.html

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

Stufflebeam, D. L. (2001). *Evaluation Values and Criteria Checklist.* Available online: http://evaluation.wmich.edu/checklists/

EXERCISES

1. Suppose that a client does not like the findings of your evaluation and says, "Well, that's just *your* opinion about the program. Evaluations are always just so subjective." How would you respond? (A suggested answer to this question is provided in the "Answers to Selected Exercises" section.)

2. Suppose that you have been asked to evaluate a counseling program that is designed to help people overcome phobias. To really understand the participants' experiences, you make extensive use of in-depth interviewing techniques as part of your work. Shortly after you publish the study, an experienced researcher who specializes in (mostly quantitative) applied psychology criticizes it due to its weak measurement, which she claims is "too subjective." The journal editor has given you 150 words of space in which to respond to this accusation. What would you write? (Be diplomatic.)

3. Imagine that you are an internal evaluation specialist who has been asked to devise a performance appraisal system. In your initial meetings with the union and various staff members, there are strong objections to the idea. "After all," says one person, "whose values will you use to decide what counts as good performance?" How would you respond (on half a page or less)?