SEVEN

DETERMINING IMPORTANCE

A frequent argument from those who *oppose* the notion that part of the evaluation team's job is to be explicit about quality, value, or importance is that no valid methodologies exist for doing so (Lawler, Seashore, & Mirvis, 1983). It is true that the average research methods text leaves the reader pretty well in the dark on this topic. But it is equally true that there has been significant headway made on the evaluation-specific methodologies available for the tasks of importance weighting, merit determination, and synthesis.[1] That is where we are headed in this chapter as well as the next few chapters.

Importance determination is defined here as the process of assigning labels to dimensions or components to indicate their importance.

When referring to importance determination, the term **importance weighting** is sometimes used. Conceptually, this is reasonably accurate. However, it does tend to make people think immediately of numerical weighting systems, which comprise only a small slice of the possibilities here. Whether one uses numbers, words, or symbols to signify importance matters little until we get to the synthesis step.

Importance determination is most relevant to the Sub-evaluation checkpoints and to the Overall Significance checkpoint of the Key Evaluation Checklist (KEC) (Exhibit 7.1). Under the Sub-evaluation checkpoints, the evaluation team needs to determine the relative importance of the various aspects of the evaluand investigated in addition to determining the merit of performance on each of those aspects (this is covered later, in Chapter 8). Under the Overall Significance

Exhibit 7.1 The KEC Checkpoints Where Importance Determination Is Used

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, 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?

checkpoint, all of these strengths and weaknesses are combined together based on their relative importance to draw overall conclusions. Methods for combining these are covered in Chapter 9.

DETERMINING IMPORTANCE: WHAT AND WHY

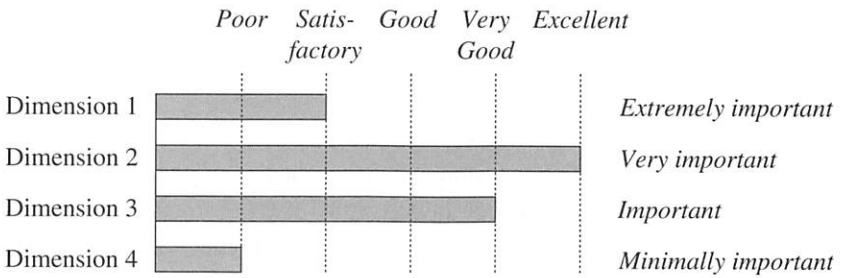
As we look down any list of evaluative criteria, it is intuitively obvious that not all of the criteria are equally important. The same is true when looking at

the performance of an evaluand across various components. Knowing which criteria and/or components are more important is essential for being able to (a) prioritize improvements, (b) identify whether identified strengths or weaknesses are serious or minor, and/or (c) work out whether an evaluand with mixed results is doing fairly well, quite poorly, or somewhere in between. In this section, we examine the distinction between **dimensional evaluation** and **component evaluation** as well as how it affects the importance determination task.

Determining the Importance of Dimensions or Criteria of Merit

Information about the importance of criteria can be used when profiling the performance of an evaluand on several different dimensions or criteria, as shown in Exhibit 7.2. Weak performance on a minor criterion (e.g., Dimension 4) may be no big deal, but weak performance on something really important (e.g., Dimension 1) would be very bad news indeed. Without this information about importance, one might think that Dimension 4 represented the most pressing area for improvement or the evaluand's most serious weakness, when in reality, Dimension 1 should probably be the primary cause for concern.

Exhibit 7.2 Hypothetical Dimensional Profile With Dimension Importance Indicated



Determining the Importance of Evaluand Components

The same logic applies to component evaluation, where the evaluand is first broken down into components (or pieces), which are considered separately before looking at the overall picture. Both component evaluation and

dimensional evaluation are analytical approaches and are distinguished from **holistic evaluation** (which involves considering the evaluand as a whole rather than breaking it down for analysis).

Quick Explanation: Dimensional, Component, and Holistic Evaluation

Dimensional evaluation: A form of **analytical evaluation** in which the quality or value of the evaluand is determined by looking at its performance on multiple dimensions of merit (also called criteria of merit) that pertain to the evaluand as a whole rather than separately to its parts

Component evaluation: A form of analytical evaluation in which the quality or value of the evaluand is determined by evaluating each of the evaluand's components (or parts) separately and then (usually) synthesizing these findings to draw conclusions about the evaluand as a whole. (Each component is usually evaluated on several dimensions of merit that pertain specifically to that component rather than to the evaluand as a whole.)

Holistic evaluation: An approach to evaluation that is either not analytical or not explicitly so and where the quality or value of the evaluand is determined at the whole evaluand level, without explicit analytical consideration of separate evaluand components or dimensions of merit

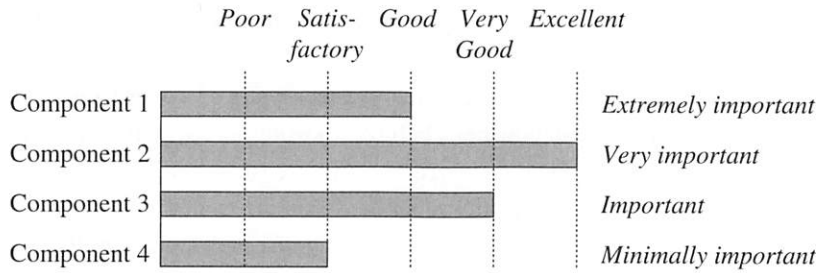
SOURCE: These terms were coined by Scriven (1991).

Component evaluation is common in the evaluation of policies, programs, or interventions that have several quite distinct parts. For example, suppose that a government policy is introduced with the aim of reducing juvenile delinquency. To achieve this goal, the government might implement several **policy instruments** (components or interventions) such as after-school programs for high-risk youth, more frequent police patrols in areas where juvenile delinquency is rife, counseling and guidance for first-time offenders, and tougher sentences for juvenile recidivists. When evaluating a multifaceted policy or program such as this, it makes sense to make the task more manageable by first breaking the evaluand out into components and considering each one

separately before looking at the interactive effects and overall merit of the entire set of policy instruments.

Information about importance may be used when profiling the performance of the evaluand on each of its components (as shown in Exhibit 7.3) and/or when synthesizing the performances on multiple components to draw an overall conclusion about evaluand effectiveness (i.e., figuring out what all of the strengths and weaknesses add up to). Again, the information about importance allows us to (a) identify the components in most urgent need of improvement (if the evaluation is formative) and (b) have some basis for determining the overall merit of a package of interventions, some of which are working better than others (for drawing overall conclusions in a formative or summative evaluation).

Exhibit 7.3 Hypothetical Component Profile With Component Importance Indicated



Determining When to Use Dimensional Versus Component Versus Holistic Evaluation

Component evaluation is more appropriate when evaluating policies, programs, or interventions that have several quite distinct parts that are experienced separately by consumers. Some typical examples might include a large-scale international development program consisting of projects implemented in different locations, a government policy that includes multiple policy instruments, and an organizational transformation effort that includes several distinct interventions.

Dimensional evaluation should usually be used when evaluating entities whose quality or value is experienced by consumers on multiple dimensions that

pertain to the evaluand as a whole. This approach is typical for single-component programs or interventions and those whose components are experienced by recipients as a package rather than separately. Products also are almost always evaluated dimensionally rather than by their components. For example, cars are usually evaluated with respect to several overall dimensions experienced by the driver (e.g., safety, handling, reliability, fuel economy) rather than by looking separately at the quality of their various components (e.g., engine, braking system, suspension). Of course, evaluative information about product components may be useful, but this information by itself is usually inadequate as a good product evaluation due to the excessive emphasis on technical specifications with little real link to users' needs (see also Scriven's [1991] entry on "technicism").

Holistic evaluation is unusual in the evaluation of programs, policies, and other large complex evaluands. One exception is seen in **connoisseurial evaluation** (also called expertise-oriented evaluation), where an expert provides an overall assessment of the evaluand without explicitly breaking it out analytically. Although we call this form of evaluation *holistic,* it is likely that something implicitly analytical is going on in the mind of the person judging the quality or value of the evaluand. In other words, the individual may be consciously or subconsciously considering the evaluand's merit on several dimensions (or by components) before drawing an overall conclusion. However, in holistic evaluation, merit is being determined in a way that goes beyond subconscious analytical evaluation. A simple example is judging the overall quality of a sample of writing. Although the person making the judgment may consider various aspects of the writing quality (e.g., interesting content, a clear thesis statement, logic arguments, correct grammar and spelling), the overall judgment is inherently holistic and does not consist of merely summing the merit on several dimensions to come to an overall conclusion. In many cases (e.g., grading essays), holistic evaluation actually yields more reliable and valid evaluations than does an analytical approach.

Holistic evaluation is more common in personnel, product, and service evaluation. It is most appropriate when the quality or value of the evaluand is experienced as an entire package, where it is either not possible or not economical to identify dimensions or components that will give a complete picture, and where reasonable (or better) reliability and accuracy can best be obtained by using judgments. Examples include the evaluation of student essays, customer service, classroom teaching, leadership potential, athletic performance, and cosmetics.

DETERMINING IMPORTANCE: SIX STRATEGIES

There are basically six strategies available for determining the importance of evaluative criteria or components:

1. Having stakeholders or consumers "vote" on importance
2. Drawing on the knowledge of selected stakeholders
3. Using evidence from the literature
4. Using specialist judgment
5. Using evidence from the needs and values assessments
6. Using program theory and evidence of causal linkages

Each of these strategies has advantages and disadvantages that make it a better choice in certain situations than in others. In this section, we look at how each one works and when to use it.

Strategy 1: Having Stakeholders or Consumers "Vote" on Importance

Many evaluators tackle the importance issue by gathering **stakeholder** input and using a kind of voting system to decide what is important. This may be a single-step, "cast your vote" approach (e.g., asking people in a survey or an interview to rate the importance of certain evaluand aspects or components), or it may be a consensus-seeking approach (e.g., an initial vote followed by a facilitated discussion among key stakeholders until agreement is reached).

Who gets to vote? Inclusion of a full range of stakeholder opinions is typical in evaluations with a strong democratic and inclusive focus and is also more likely in evaluations employing data collection methods that allow broad coverage (e.g., surveys). A small-group participatory evaluation that is more limited in scope might involve only a selected subgroup of staff within a particular unit or program. In product and service evaluation, it is more common for input to be gathered exclusively from consumers, using more of a "market research" approach. For example, one might ask people with disabilities what service attributes they consider most important in the delivery of community-based health care.

The "stakeholder vote" approach is probably the most commonly used one in both participatory and **nonparticipatory evaluations** that tackle the importance weighting issue. However, it is by no means the only option, whether the evaluation is being conducted in participatory mode or not. Before deciding whether the stakeholder vote approach is appropriate, it is important to consider the following *assumptions* that one makes when opting for it:

- Each person asked is sufficiently well informed and interested in the issues to make an assessment of importance (perhaps after some discussion if a deliberative method is used).
- The most important aspects of the evaluand are whatever aspects the participating stakeholders *believe* are important. There is no more valid way in which to determine importance in this case.
- Of those individuals whose input is sought, no particular stakeholder's or stakeholder group's assessment of importance is more credible or well informed than that of another.

It is important for the evaluation team to consider the circumstances under which these assumptions would and would not be valid. In some evaluations, certain stakeholders might not be sufficiently well informed or deeply interested to work out what *should* be considered important (e.g., due to lack of technical or content knowledge and/or time to devote to deliberation). For example, not all stakeholders would have an opinion as to the relative importance of criteria in the evaluation of a technology-based knowledge management system; instead, some (or most) might provide some initial input about the information they need from such a system and then leave the details to those with some technical expertise.

In other situations, stakeholder beliefs about what is important might be at odds with the facts. Suppose that you asked students who have just entered a doctoral program in evaluation to rate the relative importance of the knowledge and skills they will learn in the program. Although students with significant work experience in evaluation might be able to give well-based opinions on the matter, many will be not at all attuned to the kinds of knowledge and skills they will need to practice effectively in the profession. In cases like this, what certain stakeholders believe to be true may quite simply be misguided or based on insufficient knowledge.

Similarly, there are occasions where a certain group of stakeholders' views should be given greater consideration than a "one person, one vote" strategy would allow. For example, the viewpoints of elders in an indigenous community

might provide a deeper understanding of the local issues and priorities than would the viewpoints of younger people in the same community or of a government worker designing programs for that community. In situations where certain stakeholders are better informed than others, it may be advisable to weight input differentially or to filter out input that is less reliable. But more often than not, it is best to consider some alternative methods of weighting performance such as the remaining five strategies detailed in this section.

Strategy 2: Drawing on the Knowledge of Selected Stakeholders

An alternative to the stakeholder vote method of importance determination is the strategy of using selected stakeholder input to *guide* the assignment of importance weightings by the evaluation team. In the stakeholder vote method, we assumed that "importance" was roughly equivalent to "whatever most people think is important." But here we are seeking to go beyond that. We are not simply collecting opinions and reporting the "average" opinion; instead, we are selectively collecting input from various well-informed sources and combining that information in an attempt to determine what *really is* important, taking all of the relevant perspectives and considerations into account.

The methodology for doing this can range from fairly simple to quite complex, depending on the need for precision in the particular decision-making context. At the simplest level, one might identify the one or two best-informed stakeholders and conduct a brief interview that probes their relevant knowledge. For example, when determining the relative importance of the main skill sets needed by doctoral students, the evaluation team might identify certain stakeholders who are in a position to be particularly well informed (e.g., employers of graduates; university, community college, and polytechnic professors or lecturers), gather their input, and then take it into consideration when importance weights (either numerical or nonnumerical) are assigned.

When using this method for determining importance, it is necessary to consider, from an evaluation perspective, what it is that makes a particular aspect of an evaluand "important." Figuring out what importance should mean in a particular evaluation is not a trivial exercise. However, one relatively simple and usually valid option is to conceptualize importance in terms of potential impact. For example, the evaluator might ask stakeholders the following two questions: "How beneficial would it be overall if the evaluand did very well on this dimension or component?" and "How detrimental would it be overall if the evaluand did very poorly on this dimension or component?" The idea here is

that the most important dimensions or components are the ones that can make or break the evaluand, whereas the less important ones are just pluses and minuses, that is, nice if you have them but no big deal if you do not.

When determining the importance of a particular criterion, an additional consideration is identifying whether or not there is any level of performance that would be unacceptably low, regardless of how well the evaluand did on other criteria. The minimum acceptable level of performance on a particular criterion is called the **bar** (Scriven, 1991). For example, most programs, products, policies, and job vacancies have a bar on cost. Even if quality is extremely high on all other dimensions, there are limits to what can be spent (in terms of time, money, and other resources).

It is possible to set up a matrix to help with importance determination and identification of bars (or minima). A sample that could be adapted for use in various evaluations is shown in Table 7.1.

Table 7.1 A Simple Matrix for Determining the Relative Importance of Components or Criteria Based on Stakeholder Knowledge

		How detrimental would it be overall if the evaluand did *very poorly* on this dimension or component?		
		Not Noticeably Detrimental	*Noticeably Detrimental*	*Unacceptably Detrimental*
How beneficial would it be overall if the evaluand did *very well* on this dimension or component?	*Somewhat Beneficial*	Somewhat important	Important	Important (and set a bar)
	Very Beneficial	Important	Very important	Very important (and set a bar)
	Extremely Beneficial	Very important	Extremely important	Extremely important (and set a bar)

NOTE: A *bar* is a defined minimum level of criterion performance below which the evaluand is considered completely unacceptable, regardless of performance on other criteria.

When deciding whether to use stakeholder knowledge *alone* to guide importance determination, it is important to consider the assumptions that one must make (or the conditions that must be met) to use this methodology:

- The particular stakeholders who provide input regarding the probable impact of various evaluand attributes or outcomes must be sufficiently well informed to provide valuable relevant information.
- The combination of stakeholder input gathered will, as a package, provide sufficient certainty about importance for the given decision-making context; that is, no other information will be required to supplement stakeholder input.

Whenever one or both of these conditions are shaky, the evaluation team would be well advised to consider alternative or additional options for determining importance.

Strategy 3: Using Evidence From the Literature

In some situations, the evaluand may be too complex and/or stakeholder knowledge may be insufficient to allow the determination of importance with the degree of precision needed in the particular decision-making context using either of the previous two methodologies. In such cases, it may be advisable to either replace or supplement stakeholder input regarding importance with evidence from the empirical literature. Useful sources of evidence include the following:

- Meta-analyses or literature reviews addressing the effectiveness of this type of evaluand, success factors, and/or common weaknesses
- Evaluations of similar evaluands, especially in similar contexts
- Research documenting the key drivers (or strongest predictors) of success or failure with this type of evaluand

It is easy to get sidetracked in the literature when using this method, so it is important to keep a focus on evidence pertaining to the importance of certain evaluand characteristics or components as opposed to the myriad details about evaluand functioning that may have been researched. For many evaluands, we can use the potential impact rule of thumb that we employed for gathering and interpreting stakeholder knowledge (Strategy 2). In some cases, this might need to have other considerations incorporated.

However importance is defined for the particular evaluand, it will be helpful to draw up a matrix or rubric to help guide importance determination and to document how this was done for potential readers of the evaluation report. A sample matrix is shown in Table 7.2.

Table 7.2 A Simple Matrix for Determining the Relative Importance of Criteria Based on Evidence From the Literature

		What evidence exists that it would be detrimental overall if the evaluand did *very poorly* on this component or criterion?		
		Little or No Evidence of a Potential Detrimental Impact	*Clear Evidence That Some Detrimental Impact Would Be Possible*	*Evidence That Unacceptably Detrimental Impact Would Be Possible*
What evidence exists that it would be beneficial overall if the evaluand did *very well* on this component or criterion?	*Clear Evidence That Some Beneficial Impact Would Be Possible*	Somewhat important	Important	Important (and set a bar)
	Clear Evidence of Substantial Beneficial Impact	Important	Very important	Very important (and set a bar)
	Consistently a Major Determinant of Evaluand Quality or Value	Very important	Extremely important	Extremely important (and set a bar)

NOTE: A *bar* is a defined minimum level of criterion or component performance below which the evaluand is considered completely unacceptable, regardless of performance on other dimensions or components.

The use of evidence from the empirical literature alone for determining the relative importance of evaluative criteria or components depends on the following assumptions or conditions:

- The volume and quality of the available research is sufficient to allow inferences about importance to be drawn.
- The context in which the other research was conducted is sufficiently similar to that of the evaluand that the findings can reasonably be assumed to apply in this case.

Because of the questionable comparability of context in nearly all cases, it is a good idea to supplement the evidence from the literature with other evidence such as stakeholder input, specialist judgment (which is covered next), evidence from the needs assessment, and any other evidence that can be applied economically. The particular mix that works best in a given evaluation will depend on the level of certainty required and the costs of obtaining the various kinds of evidence about importance.

Strategy 4. Using Specialist Judgment

Suppose that you are evaluating something relatively complex on a fairly tight timeline, do not have sufficient expertise among the stakeholders to provide solid enough evidence of importance, and are not able to locate enough in the way of really relevant literature within the time you have available. What options are available for a "fairly quick and fairly clean" assessment of the relative importance of various criteria?

One extremely useful strategy in a situation like this is to identify one or two well-known specialists who have spent many years evaluating or studying numerous examples of this type of evaluand in contexts that are at least partially similar to yours. These kinds of seasoned evaluation (and/or research) practitioners have seen numerous examples of successes and failures and have become attuned to the things that can make or break an evaluand such as yours.

As with all undertakings in evaluation, it is always risky to base any part of the evaluation on input from just one source or even on input from two very similar sources. For this reason, you should deliberately choose two or more specialists who have quite different theoretical perspectives or who have worked in somewhat different contexts. Their assessments of importance should, wherever possible, be supplemented with other evidence gathered by the evaluation team, preferably evidence that speaks to the applicability of the specialists' judgments to this particular evaluand and setting and to these particular recipients or consumers.

Strategy 5: Using Evidence From the Needs and Values Assessments

The first four importance determination methods just discussed draw on a combination of stakeholder and expert judgment and the existing empirical research. There are pros and cons to each of these. Stakeholders are more in touch with the context at hand (or at least a specific aspect of it), although they might sometimes be a little too deep into the forest to see the trees. However, they usually lack both specific content expertise and experience with similar evaluands in different contexts. The literature and content experts, on the other hand, can usually contribute more in-depth knowledge of content (i.e., subject matter expertise) as well as evidence from multiple contexts, but they have less familiarity with the current context.

Perhaps the most relevant and powerful method for importance determination is to use evidence directly from the needs assessment (and the assessment of other relevant values). The use of this method may vary somewhat depending on whether one is determining the importance of criteria (or dimensions) of merit or the importance of evaluand components.

Determining the Importance of Criteria

What kind of evidence would we draw from a needs and values assessment, and how could we make sure that the assessment was designed to capture the information needed to establish the importance of criteria? Let's take an example of a master's program in evaluation. Based on the methods described earlier, we can list a number of sources of evidence for identifying the relevant performance needs (Table 7.3).

The needs assessment outlined in Table 7.3 is designed to identify key knowledge, skills, and other capabilities that constitute an important set of outcome criteria for the evaluation. But how do we know which of the criteria identified are the most important? At a conceptual level, we can say that the most important outcomes will be those that are particularly pivotal for distinguishing top-notch graduates from those who do poorly when they move into evaluation careers. In other words, the most important skills, knowledge, and abilities are the ones that make a huge difference in how effective an evaluator someone is. Less important outcome criteria are those that are a plus if you have them but not a serious problem if you do not.

How can we obtain information about importance from a needs assessment? One strategy might be to go back over the information collected and identify the

Table 7.3 Strategies for Identifying Different Kinds of Performance Needs for Students in (or graduates of) a Master's Program in Evaluation

	Conscious Needs	*Unconscious Needs*
Met needs	Ask evaluators (especially high performers) what skills, knowledge, and experience they gained in graduate school have been most useful for helping them to succeed.	Ask employers and clients about the skills, knowledge, and other characteristics of the best evaluators with whom they have ever worked.
Unmet needs	Ask evaluators what skills, knowledge, and experience they really needed when they first started working after graduation but had not learned in graduate school. Ask evaluators about instances when they have seen other relatively new evaluators do poorly. What skills or knowledge were the new evaluators missing that were most problematic?	Ask employers and clients for examples of people they have hired who turned out to be incapable of doing the jobs they were hired to do. What was missing from these people's repertoires? Ask employers and clients what knowledge, skills, and abilities are hardest to find when they are looking for good evaluators.
General sources (all four types)	Identify the top evaluation contracting organizations and ask what knowledge, skills, and other capabilities have made them successful and what they seek out when hiring. Talk to top evaluation theorists, identify the elements of a high-quality evaluation, and map out the knowledge and skills required to complete each one. Do some "job shadowing"; that is, observe evaluators with different skill levels in action doing their jobs. Note areas of excellence and of problematic performance. Look at examples of evaluation reports produced by master's-trained evaluators. In what area(s) were they lacking? Apply one or more meta-evaluation checklists to identify weaknesses.	

characteristics that are most frequently mentioned when respondents talk about the best and worst evaluators with whom they have ever worked. Although this is a relatively straightforward way in which to start, when using this strategy, it is important to bear in mind its key underlying assumption, that is, that *more frequently mentioned = more important.* This is not necessarily true given that the frequently mentioned dimensions might simply be the ones on which there is greatest variation, and this is quite different from importance.

A second (and better) option is to look for the characteristics of poor-performing evaluators that cause the most serious problems and, conversely, the characteristics of top-notch evaluators that have dramatic impacts on success. This information might be collected in interviews or focus groups by asking people to identify critical incidents with very serious or highly beneficial consequences (e.g., when an entire evaluation was derailed, when an evaluation created breakthrough valuable knowledge). Critical incidents are mined for information using probing questions to find out what skill, knowledge, and ability deficits or advantages appeared to be the causes of those incidents.

Determining the Importance of Components

Now suppose that we are separately evaluating several different components (i.e., distinct parts) that make up a single evaluand. What makes one component more important than another, and what relevant evidence could we gather in a needs and values assessment?

One central consideration should certainly be the severity of the needs addressed by a particular component. Interventions or services that address serious and life-threatening needs are more important than those that alleviate inconvenience. In many cases, one should also take into account whether the component in question is the only viable means of meeting those needs. The logic here is that a service or an intervention that provides the only available relief from certain problems is more important than one that is merely an option within a range of viable alternatives. Of course, the hypothetical availability of alternatives is only one part of that equation. If consumers are unlikely to seek out those alternatives, this lessens their viability and increases the importance of the evaluand component in question.

This was precisely the logic behind the importance determination task in an evaluation of a multicomponent school-based intervention (Mersman, 1999), a project for which this methodology was originally developed by the evaluation contractor and the author. This U.S. inner-city school-based

program was designed to address health, mental health, and social problems of students and parents. Seven different services (program components) were offered to students: nutrition education, education about reproductive anatomy and safer sex, mental health counseling, transportation to and from school, legal services, case management of pregnant and parenting teens, and direct health services delivered by a nurse.

The client in this case (the school principal) needed to know how well each of the seven components was meeting students' needs. Thus, this was a component evaluation with no need for an overall synthesis, that is, no need to combine the performances of all the components to draw an overall conclusion about the value of the program as a whole. However, it was important for the client to know which services were most important to allow effective prioritization of improvements and/or to know which services to retain in the event of budgetary cuts.

We began by clearly defining the determinants of component importance as follows:

- *Severity of dysfunction addressed (primary consideration):* the extent to which the component targets a serious need, that is, a source of potentially severe dysfunction and/or highly beneficial effective functioning
- *Scarcity of alternatives (secondary consideration):* the extent to which there are no alternative options for addressing the needs in question
- *Intent to seek out alternatives (secondary consideration):* the extent to which potential recipients would actually bother to seek out alternative options if the evaluand component in question did not exist

Next, we set up rubrics to clearly state how we would classify each of the components with respect to these considerations. Table 7.4 shows the rubric for rating each service (program component) on the severity of need that it addressed.

The second determinant of importance was the scarcity of alternative ways in which to meet the presenting need. In the evaluation of this multicomponent school-based intervention, this was determined by asking students whether they thought that they could receive the same services elsewhere for free (all services were free). The rubric for scarcity ratings is shown in Table 7.5.

Finally, each service was given a rating based on the intent or energy of the students to seek out any available alternatives if the services in question were not available. The more likely students were to bother seeking out an

available alternative if it existed, the more important the service was inferred to be. Like the scarcity of alternatives, this too was a secondary consideration in the determination of component importance. The rubric for determining intent ratings is shown in Table 7.6.

Table 7.4 Description of the Ratings for Severity of Dysfunction

Level of Severity	*Description of Dysfunction*
Low (■)	Not integral for survival Minor inconvenience if service not received Can function effectively without receiving service
Moderate (Δ)	Substantial inconvenience from not receiving service Implications of not receiving service probably do not include death
Severe (●)	Addresses serious dysfunction in physical, mental, or emotional health In some cases, implications of not receiving service can include death

SOURCE: Modified version reprinted with permission from Mersman (1999).

Table 7.5 Description of Scarcity Ratings

Level of Scarcity	*Percentage Reporting That They Could Not Receive the Services for Free Elsewhere*
Low (– –)	0% to 30% (clear majority believe that services for free exist elsewhere)
Slight (–)	31% to 50% (most people believe that services for free exist elsewhere)
Moderate (+)	51% to 65% (most people believe that services for free dos not exist elsewhere)
High (++)	66% to 100% (clear majority believe that services for free do not exist elsewhere)

SOURCE: Modified version reprinted with permission from Mersman (1999).

Table 7.6 Description of Ratings for Intent to Use Alternatives

Level of Intent	*Percentage Reporting That They Would Bother to Get Services Elsewhere*
Low (– –)	0% to 25% (majority would not bother to get services elsewhere)
Slight (–)	26% to 50% (some would bother to get services elsewhere, but most would not)
Moderate (+)	51% to 75% (majority would bother to get services elsewhere)
Strong (++)	76% to 100% (vast majority would bother to get services elsewhere)

Note that the symbols used to depict low (– –) through high (++) intent to seek out alternatives (Table 7.6) mirror those used to rate the perceived scarcity of alternatives (Table 7.5), with both being secondary considerations in the determination of importance. These symbols differ from the ones used to depict the severity of dysfunction addressed (●, Δ, and ■ in Table 7.4). The two different kinds of symbols come into play when the three considerations are combined to determine the importance of each program component. The severity rating provides the initial anchor (●, Δ, or ■). Then the ratings for (a) availability of alternatives and (b) the likelihood of their being used (– –, –, +, and ++) are used to "adjust" the importance rating up or down slightly to yield an overall importance rating on the scale shown in Table 7.7.

To illustrate how this methodology was used to determine the importance of each component of the school-based intervention, a full working illustration is shown in Table 7.8.

The three criteria for determining importance of the components of the school-based intervention—severity of dysfunction addressed, scarcity of alternatives, and student intent or energy to seek out those alternatives—were designed to fit the particular evaluation in question. For other evaluations, there might be a need to identify a different or modified set of determinants. The intent here was to provide an illustrative example to show the thinking behind the use of needs assessment data to determine importance.

Table 7.7 Description of Component Importance Ratings

Severity + Scarcity + Intent	*Level of Importance*
●$^{+}$	Critical
●	Extremely high
●$^{-}$	High
Δ^{+}	Moderately high
Δ	Moderate
Δ^{-}	Moderately low
■$^{+}$	Low
■	Very low
■$^{-}$	Trivial

Strategy 6: Using Program Theory and Evidence of Causal Linkages

The use of information from the needs assessment to determine importance works fairly well in dimensional evaluation when the dimensions (or criteria) relate very directly to identified needs. For example, certain skills, knowledge, and abilities of evaluation master's program graduates can be linked very directly with good or poor performance as an evaluator. The same is true in component evaluation where each evaluand component clearly addresses a need whose severity can be determined quite clearly (as was the case with the components of the school-based intervention mentioned earlier).

There are some other cases, however, where the criteria or components in evaluations are linked to needs through a more complex logic chain. For example, although the importance of certain job-related skills (e.g., time management, technical know-how, communication skills) can be determined by direct reference to how they affect performance, the importance of "soft" skills and attributes (e.g., inspirational leadership, self-esteem, stress management) is much more difficult to determine. This is because these criteria are not valuable in their own right (i.e., they have no intrinsic value) but are valuable to the extent that they lead to something else that is verifiably valuable (i.e., they have "instrumental"[2] value).

Table 7.8 Determination of Importance of Various Student Services in the School Health Program

Program Component	*Types of Dysfunction Addressed*	*Severity Rating*[a]	*Scarcity of Services*	*Scarcity Rating*[b]	*Intent/Energy to Obtain Available Alternatives*	*Intent/ Energy Rating*[b]	*Overall Importance*
Health education in nutrition	• Malnutrition • Weight problems • Illness	→ ●	• Slight—most people (58%) believe that they could receive services elsewhere	→ –	• Moderate—majority (59%) would bother to seek out available alternatives	→ +	→ ● Extremely high
Case management of pregnant or parenting teens[c]	• Poor health for self or baby • Lack of parenting skills • Inability to plan academically and vocationally	→ Δ/●	• Slight—most people (50%) believe that services exist elsewhere	→ –	• Strong—mass majority (100%) would bother to seek out available alternatives	→ ++	→ $●^{-}$ High
Transportation[c]	• Lack of access to child care, case management, and existing services • Economic hardship • Threatened safety if forced to walk	→ ■/Δ	• High—clear majority (67%) report that they could not receive services elsewhere	→ ++	• Moderate—majority (67%) would bother to seek out available alternatives	→ +	→ $●^{-}$ High
Mental health	• Suicide • Depression • Inability to focus on schoolwork • Absence from school • Lack of referrals (reduced access to other services)	→ ●	• Slight—most people (60%) report that they could receive services elsewhere	→ –	• Slight—most (60%) would not bother to seek out available alternatives	→ –	→ $Δ^{+}$ Moderately high

(Continued)

Table 7.8 (Continued)

Program Component	*Types of Dysfunction Addressed*	*Severity Rating*[a]	*Scarcity of Services*	*Scarcity Rating*[b]	*Intent/Energy to Obtain Available Alternatives*	*Intent/ Energy Rating*[b]	*Overall Importance*
Health education in reproductive anatomy and safer sex	• Spread of sexually transmitted diseases and AIDS • Pregnancy • Absence from school	→ Δ/●	• Slight—most people (60%) report that they could receive services elsewhere	→ –	• Slight—most people (58%) would not bother to seek out available alternatives	→ –	→ Δ Moderate
Legal services	• Economic hardship • Legal problems • Inattention to schoolwork	→ ■	• High—vast majority (70%) report that they could not receive services elsewhere	→ ++	• Slight—most people (57%) would not bother to seek out available alternatives	→ –	→ ■+ Low
Health clinic	• Lack/Delay of emergency care for minor illness • Inability to access care • Lack of referral system to other services	→ Δ	• Slight—most people (54%) believe that services exist elsewhere	→ –	• Slight—most people (58%) would not bother to seek out available alternatives	→ –	→ ■+ Low

NOTES:

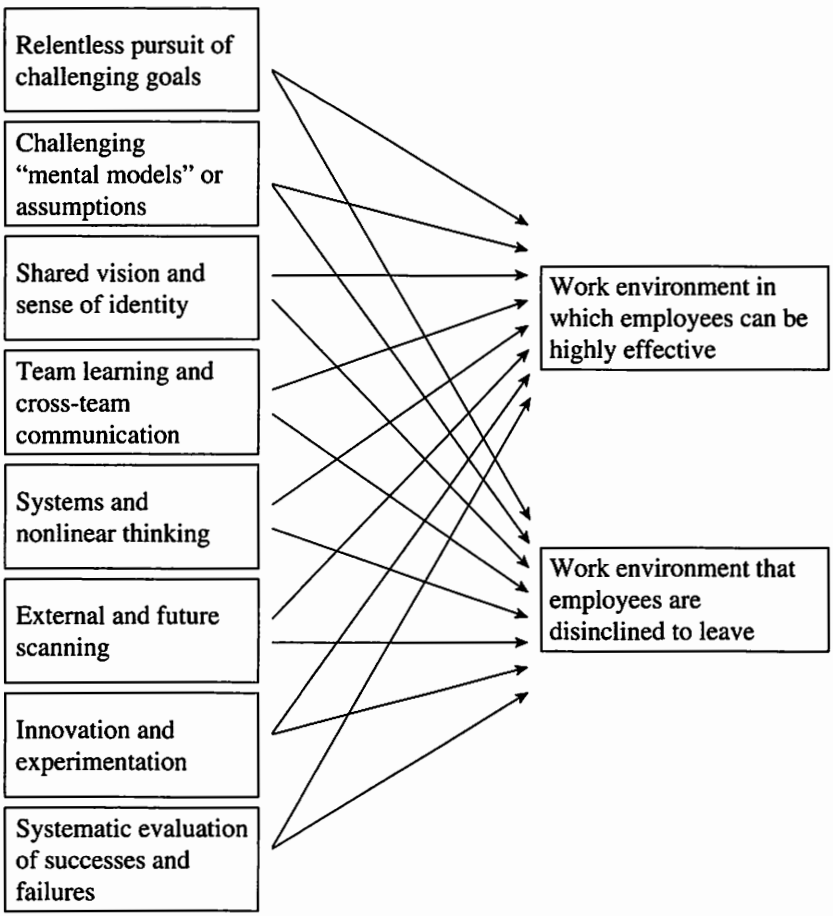
a. ● = severe, Δ = moderate, ■ = low severity.

b. ++ = strong, + = moderate, – = slight, – – = low.

c. $N < 10$ (warrants caution in interpretation).

Another way in which to conceptualize these criteria with instrumental value is as "upstream variables" in a program logic model. To illustrate, Exhibit 7.4 shows an example of a logic model used in an evaluation of the learning capacity of a small biotechnology start-up company on the U.S. East Coast (Davidson, 2001).

Exhibit 7.4 Logic Model Linking Aspects of Organizational Learning Culture to Performance Needs at the Individual Level of Analysis



Eight dimensions of organizational culture had been identified from the literature as being the elements that distinguish "learning-enabled" organizations from "learning-impaired" organizations. The underlying logic was that if

these elements of the organizational culture were strong, this would create an environment where employees could be highly effective in their jobs while at the same time making them disinclined to leave the organization. The latter variables can be considered the *performance needs* of the organization at the individual level of analysis[3] (for a review of performance needs, refer back to Chapter 3).

The importance of these two criteria was determined using Strategy 2 (drawing on the knowledge of selected stakeholders). Based on an in-depth interview with the company's owner-manager, it was determined that it was somewhat more important for this organization to create an environment where employees could be highly effective (in this case, creative), given the innovation-intensive nature of the biotechnology industry, than it was to create an environment that would ensure employee retention (Table 7.9).

Table 7.9 Determining the Importance of Criteria Related to Performance Needs

Criterion	*Potential Impacts of Excellent or Poor Performance on This Criterion*	*Importance*
Individual performance	Quality/Productivity of the creative process (which drives product quality and, therefore, is central to organizational survival) is heavily dependent on each individual's ability to add maximum value in his or her position.	Extremely high
Employee retention (especially of top performers)	Continuity is important for the development of a particular product and is desirable across multiple projects due to cumulative learning effects. Employees have high levels of very specific expertise, are hard to replace, and carry significant organizational knowledge.	High

For this evaluation of organizational learning capacity, the goal was to produce a profile of organizational "learning-enabledness" on the eight dimensions of organizational culture. To help the organization prioritize any efforts to improve the organization's learning culture, it was also necessary to provide some indication of the relative importance of each dimension.

Although determining the importance of the downstream criteria was relatively simple, the challenge in this evaluation was determining the importance of the dimensions of organizational learning culture, that is, the upstream variables. The organizational stakeholders had little or no knowledge about organizational learning, which ruled out Strategy 1 (stakeholder vote) and Strategy 2 (using stakeholder input). In addition, because this was a relatively new area of study, the lack of relevant empirical evidence ruled out both Strategy 3 (using evidence from the literature) and Strategy 4 (using specialist judgment). In any case, it seemed likely that the importance of the eight dimensions of organizational learning culture would vary from organization to organization. Finally, because the organizational learning culture dimensions were not performance needs, Strategy 5 (using evidence from the needs assessment) was also inappropriate.

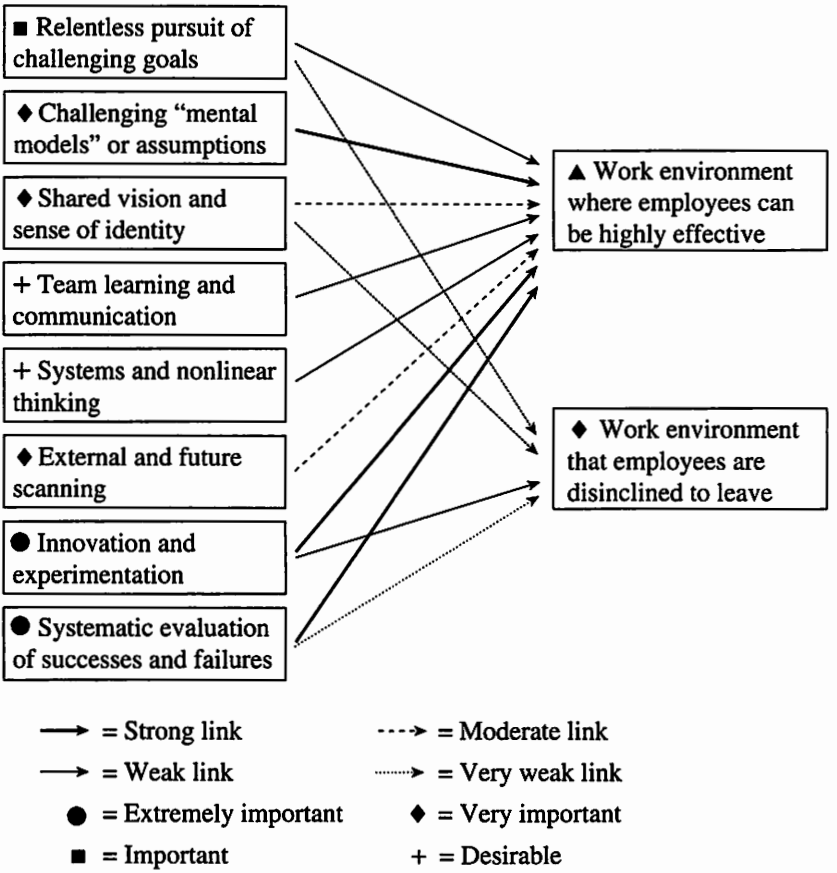
The challenge posed by this evaluation sparked the development of a new methodology for determining the importance of upstream variables. The underlying logic was as follows: The more important upstream variables are those that are most strongly causally linked to the most important downstream variables. Accordingly, the first task was to draw on the tools outlined in Chapter 5 to assess the strengths of the causal links between each of the organizational learning culture dimensions and the two performance needs-related criteria.

For this particular evaluation (which was conducted on a shoestring), all that was required was a broad-brush estimate of dimension importance rather than a high degree of accuracy. The causal analysis drew on several pieces of data to estimate the strengths of the links:

- Visual analysis of two-dimensional scatterplots
- Correlation coefficients (excluding outliers where appropriate)
- Employee responses to open-ended questions about the most important determinants of their ability to perform effectively and of their intent to stay with or leave the organization
- Owner/Manager accounts of the work environment variables that appear to affect employee performance and turnover

Based on this mix of information, the strengths of the links were added into the logic model, and the information was used to determine the importance of the organizational learning culture dimensions (Exhibit 7.5). Extremely important organizational learning culture dimensions were those that were strongly linked to the most important need and had at least a weak

Exhibit 7.5 Determination of the Importance of Upstream (learning culture) Dimensions Using a Logic Model and Evidence of Causal Links

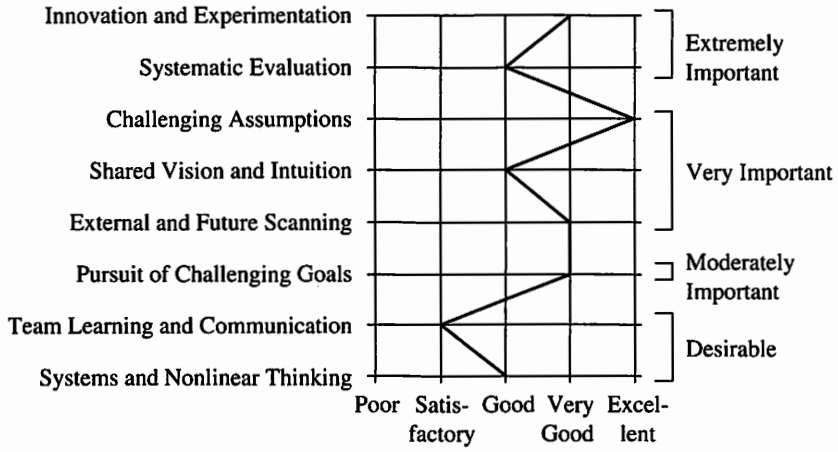


link to the other need. Very important dimensions had a strong or moderate link to the most important need. Important dimensions had weak or very weak links to both needs, whereas the least important (desirable) dimensions had weak or very weak links to both needs.

Of course, it is always possible to argue for slightly different cutoffs when classifying the dimensions into the four importance categories. Some debate about this is healthy to ensure that the classification system is justifiable and to help stakeholders understand how it works. However, it is important to bear

in mind the purpose of the exercise, that is, to provide a broad-brush assessment of dimension importance that will allow the client to identify priorities and make effective decisions more easily. This is illustrated in Exhibit 7.6, which shows the learning culture profile presented to the client in this case. Although many evaluations will require more accuracy than was used in this example, it is important not to obsess about achieving a far higher level of precision than what is really required.

Exhibit 7.6 Learning Culture Profile Arranged According to Importance Weightings



Strengths and Weaknesses of the Six Strategies

The six importance determination strategies outlined in this chapter vary considerably in their complexity and in the kinds of situations to which they are most applicable. Each has its own set of advantages and challenges, as outlined in Table 7.10. Often the best option is to employ the principles of critical multiplism (Shadish, 1994; Shadish, Cook, & Campbell, 2002) and to choose two or three complementary strategies with different weaknesses.

As noted in this chapter, one of the main uses of importance determination is to allow more illustrative profiling of findings for a client. However, there is another application: synthesizing mixed findings on several dimensions or components to draw an overall conclusion about evaluand quality or

Table 7.10 Advantages and Challenges of the Six Importance Determination Strategies

Strategy	*Advantages*	*Challenges*
1. Having stakeholders or consumers "vote" on importance	Is inclusive and democratic; maximizes buy-in to the evaluation process; requires relatively little expertise in importance determination on the part of the evaluation team	Assumes that all voters are equally well informed; assumes that *popularity* = *importance* (which weighs minority opinions more lightly); may be expensive if many opinions are sought; opens the "who chooses the voters?" can of worms
2. Drawing on the knowledge of selected stakeholders	Targets those stakeholders in the best position to know about importance; combines stakeholder expertise with evaluator expertise; relatively cost-effective compared with Strategy 1; gets buy-in from the top	Requires more skill on the part of the evaluation team; needs stakeholders to have sufficient defensible knowledge of importance; requires careful justification of the choice of stakeholder informants
3. Using evidence from the literature	Avoids reinventing the wheel; provides justification that is independent of those with a vested interest in the evaluand; is a good method to complement stakeholder input or other methods	Requires sufficient literature that addresses this issue (often not available for very innovative evaluands); can be time-consuming if literature is widely dispersed; can be seen as overly academic and undervaluing of local knowledge

Strategy	*Advantages*	*Challenges*
4. Using specialist judgment	Is considerably quicker than a literature search; does not rely on either stakeholder expertise or evaluation team expertise; can help with credibility (if the specialist has "brand recognition"); is a good method to complement stakeholder input or other methods	May yield information that represents just one line of thought in the body of knowledge on this topic; can be seen as undervaluing of local knowledge
5. Using evidence from the needs and values assessments	Provides independently verifiable evidence of importance that is directly related to this evaluand in this context	Requires expertise in needs and values assessments; works only for those criteria for which there is direct evidence of importance
6. Using program theory and evidence of causal linkages	Provides independently verifiable evidence of importance that is directly related to this evaluand in this context; can also determine importance of upstream outcomes where there is no direct evidence of "intrinsic" importance	Requires substantial expertise on the part of the evaluation team (or hiring an expert to help); requires a defensible program theory; can be time-consuming and expensive; may be difficult to explain to stakeholders and/or to use in participatory mode

value. We discuss this challenging task later in Chapter 11. But first, we need to tackle another explicitly evaluative task: merit determination.

NOTES

1. By far the greatest contributions in this area to date, especially with respect to conceptualizing what it is we do when we infer merit or worth, have been from Scriven (in particular, see Scriven, 1991). He certainly deserves credit for providing much of the conceptual "grist for the mill" that the author has used to develop these methodologies and put them into a form that we can all run with—and, hopefully, that he and others will help the author improve on.

2. Many thanks go to colleague Christopher Nelson for suggesting this terminology.

3. It is true that the organization has other performance needs at the group/team, business unit, and organizational levels of analysis. However, these are immaterial for the purposes of demonstrating this importance determination methodology.

ADDITIONAL READINGS

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

- Analytic/Analytical
- Component evaluation
- Dimensional evaluation
- Global
- Going native
- Incestuous relations
- Interactive evaluation
- Perspectival evaluation
- Stakeholder
- Technicism
- Therapeutic role
- Utilization

Davidson, E. J. (2002). Organizational evaluation: Issues and methods. In R. L. Lowman (Ed.), *Handbook of organizational consulting psychology* (pp. 344–369). San Francisco: Jossey-Bass.

Davidson, E. J. (2003). Linking organizational learning to the bottom line: Methodological issues, challenges, and suggestions. *Psychologist-Manager Journal, 6*(1), 54–67.

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.

EXERCISES

1. In your own words, briefly define and explain the differences among (a) holistic evaluation, (b) component evaluation, and (c) dimensional evaluation. For each one, (i) give a real-world example of an evaluation in your profession for which you would choose one over the others and (ii) indicate why—for example, not just why (a) but also why not (b) or (c). The examples should not be taken from the texts, from one of your assignments, or from your project. (A suggested answer to this question is provided in the "Answers to Selected Exercises" section.)

2. What are the two main applications of importance determination as outlined in this chapter? Are there any other possible applications?

3. List and explain clearly in nontechnical terms the six different strategies for determining the importance of criteria of merit and/or evaluand components. For each one, describe a hypothetical evaluation in your field where that particular strategy would probably be the best option. Justify your choice in each case.

4. List the main criteria (under the headings of Process, Outcomes, and Cost) that you will be using to determine the quality or value of your evaluand. Outline at least two or three strategies that you would use to determine their relative importance and indicate how you would go about using them (e.g., If you use stakeholder input, exactly whose input would you seek and why?). Explain your choices, particularly why you did not use the other three or four importance determination options.