EIGHT

THE MERIT DETERMINATION STEP

Evaluation is the systematic determination of the quality, value, or importance of something (Scriven, 1991). That "something" can refer to an entire evaluand (e.g., a program or product), or it can refer to aspects (i.e., dimensions or criteria) or pieces (i.e., components) of an evaluand. The previous chapter outlined several strategies for determining the *importance* of evaluand components or dimensions. This chapter explores how to determine the *quality or value* of performance on these components or dimensions (i.e., the merit determination step).

> **Merit determination** is the process of setting "standards" (definitions of what performance should constitute "satisfactory," "good," etc.) and applying those standards to descriptive data to draw explicitly evaluative conclusions about performance on a particular dimension or component.

The merit determination step is where we apply the contents of the Values checkpoint to the descriptive data we gather to draw evaluative conclusions under the Sub-evaluations checkpoints of the Key Evaluation Checklist (KEC) (Exhibit 8.1). Note the explicitly evaluative questions under each checkpoint.

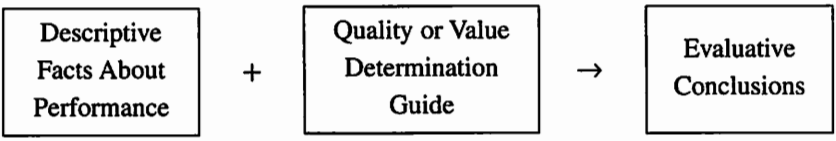
The "big picture" question of how we should determine the quality, value, or importance of an evaluand overall is addressed later in Chapter 9, where we talk about synthesizing all of our findings to draw an overall evaluative conclusion.

Exhibit 8.1 The KEC Checkpoints Where the Merit Determination Step Appears

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?

DETERMINING MERIT: WHAT AND WHY

Merit determination involves two steps: (a) defining what constitutes poor, adequate, good, very good, and excellent performances on a particular dimension (or for a particular component) and (b) using that definition to convert empirical evidence of evaluand performance (descriptive facts) into evaluative conclusions (i.e., saying something explicit about quality or value). Here we are applying a basic evaluation formula:



Merit Determination Using a Single Quantitative Measure

In the special case where performance is being measured on a single quantitative dimension, the quality or value determination guide would simply

be a set of cutoff scores (e.g., for a test, > 90% = A/excellent, 80%–89% = B/good, 70%–79% = C/adequate). In some cases, it might be just one cutoff, that is, the line between satisfactory and unsatisfactory performance.

The difficult issue when converting scores to grades is determining where the cutoff score should be placed.

Pop Quiz: Using Fixed Cutoffs to Ensure Consistency

In the United States, school and university exam and course grades are frequently determined using the following cutoff scores: > 90% = A, 80%–89% = B, 70%–79% = C, 60%–69% = D, < 60% = F. Many argue that mandating such cutoffs is one way in which to ensure objectivity and consistency of grading across courses. Is this true? If so, why? If not, why not?

In New Zealand, school and university exam and course grades are frequently determined using the following cutoff scores: > 80% = A, 65%–79% = B, 50%–64% = C, 35%–49% = D, < 35% = F. Does this mean that it is easier to get A's in New Zealand universities than in U.S. universities? If so, why? If not, why not? (Puzzled? Make contact with someone who has been to a university in New Zealand, or in a country that uses different cutoffs from those in your home country, and ask him or her about it.)

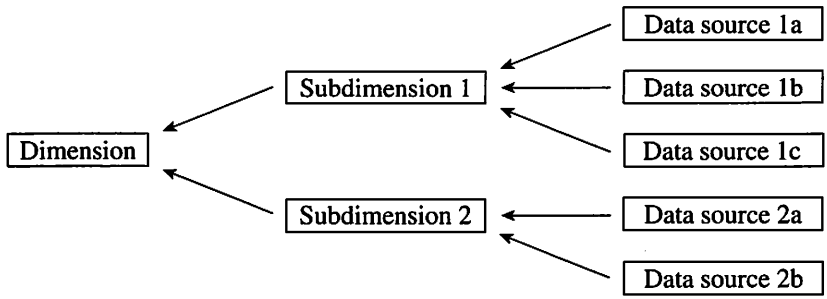
Merit Determination With Qualitative or Multiple Measures

As mentioned previously, the use of a single measure to assess performance on a particular dimension is not generally good practice. This means that for most evaluations, the evaluation team will be faced with a much more complex set of data (often a mix of qualitative and quantitative data) that must be converted to evaluative conclusions.

Many of us are trained in either the social sciences or the hard sciences, so breaking things down into their component parts comes fairly naturally. Once we have done that, we can go out and gather the data while applying our knowledge of research methodology. The tricky part comes when all of those data come in and we are left with a mass of information that needs to be

packed back together to find answers to the question of how well the evaluand did on a particular dimension or component (Exhibit 8.2).

Exhibit 8.2 Synthesizing Multiple Data Sources to Draw Evaluative Conclusions About Performance on Dimensions and Subdimensions



A more concrete example of what the problem looks like is provided in Table 8.1. Here we have multiple sources of data pertaining to the performance of a hypothetical graduate program on just one dimension: job placement. To assess the program's performance on job placement, four **subdimensions** have been defined: (a) speed and ease of placement, (b) level and quality of jobs obtained, (c) prestige and desirability of organizations where graduates find work, and (d) match of positions with graduates' interests and aspirations. For each of the subdimensions, multiple measures and indicators have been collected.

In the next section, we look at how to make sense of a mix of data such as this. But first, let's tackle a couple of key points that are important to bear in mind as we do this.

Merit Determination and the Futility of Seeking Precision

For each of the subdimensions pertaining to the quality of job placement in our hypothetical graduate program, there are several sources of data (Table 8.1). Some of these point in the same general direction, whereas others seem to partially contradict each other. Either way, we are in need of some strategy for figuring out whether the information we have obtained about, say, speed and ease of placement (one of the subdimensions) should be considered "excellent," "very good," "good," "barely acceptable," "poor," or "completely unacceptable."

Table 8.1 Example Data on Subdimensions Related to Quality of Job Placement

Speed or ease of placement	Three quarters (75%) of graduates who sought work found employment within 3 months of graduation (mean = 6 weeks). Nearly one third (30%) had job offers by the time they graduated. Only 15% were still unemployed, underemployed, or in jobs unrelated to their degrees 12 months after graduation. Most graduates (85%) complained that finding work was considerably more difficult than they had expected. The average graduate sent out 22 applications, was invited in for three or four interviews, and was offered one or two jobs.
Level or quality of jobs obtained	Most jobs obtained were entry level, with 10% of the class making it into "senior associate"-level positions (all of these people had 5 or more years of previous experience). The starting salary mean was $38,000 (range = $29,000–$84,000). Feedback from graduates indicated that jobs were generally moderately challenging relative to their skills.
Prestige or desirability of organizations where graduates find work	Among graduates in the for-profit sector, 12% found work in Fortune 500 companies. A small percentage (2%) found work in companies rated in the top 20 "best places to work" lists. Expert assessments showed that the reputations of most graduates' new organizations were moderate to weak, although one or two graduates found work in very highly regarded institutions and organizations.
Match of positions with graduates' interests and aspirations	Analysis of pregraduation areas of specialization against job and/or company type showed that 65% of graduates found work in desired areas of specialization, 20% were in appropriate industries but not appropriate job areas, and 15% were in jobs that were only marginally related to their areas of specialization. Most graduates expressed at least some disappointment regarding the degree of match between their jobs and their career aspirations and interests.

Of course, a particular evaluation might not require such fine gradations. This is just an example using a number of quality categories (six) that tends to be workable in most cases. In the author's experience, there is not much to be gained from trying for greater precision (e.g., 10 categories of merit) unless the evaluand really lends itself to that (this may be possible in some product evaluation tasks). Attempts to strive for high levels of precision in evaluation usually result in a lot of time-consuming debate about what should go where.

When it comes to the evaluation of programs, policies, and/or other things that involve people (this would include most evaluands), evaluation is sometimes a fairly blunt instrument. We can usually attain a reasonable broad-brush level of accuracy that is good enough to meet the informational needs of the client and other right-to-know audiences. But attempts to achieve a level of precision that outstrips the tools and knowledge at our disposal can do little but undermine credibility and increase accusations of arbitrariness.

Remember that just because we can measure something to four decimal places does not mean that we can rate its quality or value to the same level of precision. Performance on all dimensions or criteria of merit should be assessed using a mix of data, usually both qualitative and quantitative, plus a mix of information drawn from the needs assessment and other relevant considerations (e.g., professional standards). Rarely is this cut-and-dried. But neither is it a hopelessly impossible task so long as we can stay comfortable with a certain amount of fuzziness around the edges and we do not oversell the precision of our work.

The practical point to remember here is that it is usually impossible to obtain high levels of precision on the merit determination step. However, you should keep in mind the following:

- Providing a well-supported broad-brush answer to an important question is generally far more valuable to clients (and to other audiences) than is telling them that the answer is impossibly mired in subjectivity and so they will have to work it out for themselves.

- It is perfectly appropriate to give an answer that still has a certain amount of fuzziness or uncertainty associated with it. (As mentioned earlier in the discussion of causation, one does not always need a very high degree of certainty about the answer. In any case, it may be extremely helpful indeed to get even part of the way there, e.g., by dealing with the parts that are reasonably straightforward, narrowing the options, and clarifying the trade-offs surrounding any dangling issues.)

USING RUBRICS FOR DETERMINING "ABSOLUTE" MERIT

So, how are we going to convert a mix of quantitative and qualitative data into some rating of the quality or value of that attribute or level of performance? One tool that can be incredibly useful (and a good conversation starter with the evaluation team and stakeholders) is a rubric.

> A **rubric** is a tool that provides an evaluative description of what performance or quality "looks like" at each of two or more defined levels.
>
> A **grading rubric** is a rubric that is used to determine *absolute* quality or value, whereas a **ranking rubric** is used for questions of *relative* quality or value.
>
> (See Chapter 2 for a review of absolute versus relative quality or value.)

A generic example of a grading rubric that can provide a good starting point for the merit determination step is shown in Table 8.2. The rubric shown in the table is merely a starting point for rubric development. It does, of course, take a significant amount of additional work to define terms such as *exemplary performance* and *serious weakness.* This usually requires a combination of

Table 8.2 Generic Rubric for Converting Descriptive Data Into "Absolute" (rather than "relative") Determinations of Merit

Rating	*Explanation*
Excellent	Clear example of exemplary performance or best practice in this domain; no weaknesses
Very good	Very good or excellent performance on virtually all aspects; strong overall but not exemplary; no weaknesses of any real consequence
Good	Reasonably good performance overall; might have a few slight weaknesses but nothing serious
Barely adequate	Fair performance; some serious (but nonfatal) weaknesses on a few aspects
Poor	Clear evidence of unsatisfactory functioning; serious weaknesses across the board or on crucial aspects

background research and extensive discussions with experts and/or key stakeholders. (Two simple examples of this are provided later in the chapter.)

Developing Rubrics in a Participatory Evaluation

This process of defining "how good is good" can be an incredibly valuable exercise for helping all sorts of organizations to think through what they mean by *quality* or *value.* In a participatory evaluation, this part of the process forms an important part of the groundwork for the evaluation and doubles as an intervention that helps people to focus on what is really important about the work they do.

Whether the evaluation is being conducted in participatory mode or not, it is very important to talk to consumers at this point when developing a merit determination rubric. After all, the program, policy, or product is presumably designed to create value for them. This can help organizational staff to identify incorrect assumptions they might have been making about needs and other issues.

Sample Grading Rubric 1

To give an example of a more fully fleshed-out merit determination rubric, Table 8.3 shows what an early draft might look like for one of the subcriteria identified for a hypothetical master's program in evaluation. It is important to note that the example given in the table is merely a sample rubric that has not been subjected to discussion with key stakeholders or job placement experts. In nearly all cases, rubrics such as this need considerable refinement based not only on, for example, student or graduate expectations but also on expert (e.g., recruiter, job placement specialist, employer) input regarding the job market and what expectations would be reasonable for graduates with this particular mix of qualifications and experience.

Recall the data collected for our hypothetical master's program:

- Three quarters (75%) of graduates who sought work found employment within 3 months of graduation (mean = 6 weeks).
- Nearly one third (30%) had job offers by the time they graduated.
- Only 15% were still unemployed, underemployed, or in jobs unrelated to their degrees 12 months after graduation.
- Most graduates (85%) complained that finding work was considerably more difficult than they had expected.
- The average graduate sent out 22 applications, was invited in for three or four interviews, and was offered one or two jobs.

Table 8.3 Rubric for Determining the Merit of a Master's Program in Evaluation on the Subcriterion "Speed or Ease of Job Placement"

Rating	*Description*
Excellent	All students had evaluation-relevant job offers on graduation or soon after (within 2 months excluding those who were not actively seeking such employment), and several students had more than one strong job offer. Several high-profile organizations recruited on campus or sought recommendations through program faculty to identify the best recruits.
Very good	The vast majority of students (> 80%) had evaluation-relevant job offers on graduation or soon after (within 2 months excluding those who were not actively seeking such employment), and several students had more than one strong job offer. A small number of high-profile organizations recruited on campus or sought recommendations through program faculty to identify the best recruits. Most students had to be quite proactive about networking and applying for jobs.
Good	Most students (> 70%) had job offers on graduation or soon after (within 2 months excluding those who were not actively seeking such employment), although some of these were not directly related to evaluation, and some students had more than one reasonable job offer. Most students had to drive their own job-seeking agendas quite hard, although some assistance was provided. Those students without job offers tended to be those who were more passive about job seeking.
Barely adequate	Most students (> 70%) had at least one job offer within 3 or 4 months of graduation, although many of these were not related to evaluation. Job-seeking efforts had to be very intensive to obtain decent job offers. Many graduates reported that employers were not at all familiar with their university or the program.
Poor	With only a few exceptions (< 30%), most graduates of the program took up to 6 months to obtain placements (or promotions in their current jobs) that were only slightly better than what they had left to enroll in the program (or that were only slightly better than what bachelor's-level graduates were getting). Most positions were not related to evaluation but rather were related to the cognate areas.
Completely unacceptable	Graduates of the program found it difficult even to obtain positions equivalent to the ones they had left to enroll in the master's program.

Based on this information, the program in question seems to fit most closely with a rating of "good" on the rubric in Table 8.3. In a real evaluation, you might need to do some further digging to make sure that the rating is justified. For example, there might be a high proportion of graduates who have decided to go on to doctoral programs, in which case they should not be counted as having been unable to find full-time jobs.

Sample Grading Rubric 2

Here is another example of a grading rubric that draws on much more qualitative (i.e., nonnumerical) information to draw conclusions about merit. In this case, the rubric is used for performance appraisal; however, the logic is the same as for program evaluation. The rubric was designed for evaluating the performance of clerical staff in a small accounting office on their management of monthly accounts (one of several duties). It was developed in discussion with the two business owners/partners, who defined the expectations. The rubric starts with a description of the scope of duties, lists the main performance indicators, and then defines each level of performance—in this case, from "unacceptable" to "excellent" (Table 8.4).

Table 8.4 Performance Appraisal Rubric for a Specific Set of Tasks in a Small Accounting Firm

Monthly Support Packages (clerical)
Scope: • Preparing monthly financial reports • Responsible for data entry • Liaising with clients • Critically analyzing results • Systematically reviewing income tax liabilities
Performance indicators: • Timeliness and efficiency • Accuracy • Clarity of communication, tact, and diplomacy • Use of Inland Revenue Department (IRD) or Internal Revenue Service (IRS) compliance knowledge
Instructions: Choose the description that best fits how well the objective has been met, and check the appropriate box.

Rating	*Description*
Totally unacceptable performance (score = 1)	Any one or more of the following: (a) inadequate checking and/or following up of queries or missing information, leading to serious inaccuracies in data entry and/or monthly reports; (b) failed to report one or more major problems or issues to partners; (c) inadequate documentation, making auditing extremely difficult or impossible; (d) frequently rude or abrupt with clients; (e) failed to inform clients of important obligations on one or more occasions
Mediocre (substandard) performance (score = 2)	Any one or more of the following: (a) inadequate use of communication skills, checking or following up of queries, or missing information, leading to minor inaccuracies in data entry and/or monthly reports; (b) failed to report one or more minor problems or issues to partners; (c) failed to inform clients of minor obligations on one or more occasions, causing inconvenience; (d) barely adequate documentation and/or auditing trails, making quality checking possible but somewhat difficult; (e) inadequate prioritizing of time, leading to one or more jobs being completed outside budgeted time frames (except when delay was out of the accounting firm's control)
Good performance (expected level) (score = 3)	Efficient checking of data entry, allowing preparation of accurate monthly reports supported by clear work papers and audit trails; clients always informed of their obligations and requirements; partners kept informed of any problems or issues as they came to light; time prioritized so that all jobs were completed within budgeted time frames unless delays were out of the organization's control; queries and missing information always documented and followed up quickly and efficiently to ensure that jobs were not held up; all clients handled professionally and courteously with excellent communication skills displayed; thorough documentation, allowing for rapid evaluation of clients' overall financial positions and internal record keeping and systems

(Continued)

Table 8.4 (Continued)

Rating	*Description*
Performance exceeded expectations (score = 4)	All of the above in addition to the following: excellent use of communication skills and time management, ensuring that clients had an excellent understanding of their financial situations and that statements were 100% accurate and consistently completed well within budgeted time frames; meticulously organized work papers and audit trails, allowing any staff member to quickly ascertain the current state of any work in progress and to check the accuracy of work completed; constantly worked to streamline procedures for both clients and the accounting firm
All-around excellent performance (score = 5)	All of the above in addition to the following: superb professional service to clients, enhancing the reputation of the accounting firm and resulting in positive feedback and/or new clients through word-of-mouth advertising; innovative approach to managing monthly support packages, resulting in a smooth-running and error-free system that allowed jobs to be completed significantly more efficiently than time frames budgeted for (levels to be agreed on between partners and employees)

USING RUBRICS FOR DETERMINING "RELATIVE" MERIT

In some cases, the evaluation team will need to determine the relative merit (rather than the absolute merit) of performance on a particular dimension. Relative merit evaluations (i.e., ranking) tell us little or nothing about how good the performance was in any absolute sense. They simply tell us how the person or program did relative to peers or competitors, respectively.

"Grading on the Curve"

Perhaps the simplest example of this is the practice called "grading on the curve." Although the term *grading* is used (and letter grades may even be given), the instructor is actually ranking rather than grading evaluees. Table 8.5 shows a hypothetical rubric that might be used to generate grades for student performance in a large class.

The main problem with grading on the curve is that the letter grades imply that there is some sense of absolute merit (e.g., A = excellent, B = good, C = satisfactory). But the reality is that this system forces the instructor to fail

Table 8.5 Hypothetical Rubric for "Grading on the Curve" (actually ranking)

Score Falls in:	*Grade Assigned*
Top 10%	A
Next 25%	B
Next 50%	C
Next 15%	D
Bottom 5%	F

a certain proportion of the class, whether those students are performing at an unsatisfactory level or not. In addition, it forces the instructor to give A's to 10% of the students, regardless of whether their performance was truly excellent. In general, if ranking is being used, the terminology used to label the categories should make it clear that this is ranking (e.g., "top 10%" instead of "A").

Standardized Tests

Most standardized tests, such as the Scholastic Aptitude Test (SAT), the Graduate Record Examination (GRE), and the Graduate Management Admission Test (GMAT), also determine relative merit rather than absolute merit, expressing scores in percentile terms that indicate the test taker's percentile rank (i.e., what proportion of all test takers scored lower). One recently added exception is the analytical writing section of the GRE, which provides a numerical rating that corresponds to a description of absolute merit. Tests of intelligence quotient (IQ) are another example of tests that determine where someone falls relative to the population. Unlike the aforementioned standardized tests, IQ score ranges are assigned explicitly evaluative labels such as "gifted" (see Table 8.6 for the conversion rubric).

Relative Merit and Experimental and Quasi-Experimental Designs

Determination of the relative merit of outcomes is particularly important for experimental and quasi-experimental evaluation designs, that is, designs that incorporate the use of a control or comparison group. For example, student achievement scores for a particular school are often interpreted relative to state

Table 8.6 Rubric for Interpreting IQ Scores

Evaluative Intelligence Rating	*IQ Score*	z *Score*[a]	*Percentage Below*
Exceptionally gifted	160	+4	> 99%
Highly gifted	145	+3	99%
Very superior/gifted	130	+2	98%
High average	115	+1	84%
Average	100	0	50%
Low average	85	−1	16%
Borderline	70	−2	2%
Mild mental retardation	55	−3	1%
Moderate mental retardation	40	−4	< 1%

NOTE: a. A z score indicates how many standard deviations a score is above or below the mean.

SOURCES: www.psychologicaltesting.com/iqtest.htm and http://iq-test.learninginfo.org/iq04.htm

averages or by comparison with schools from areas with a similar demographic and socioeconomic makeup.

The usual approach of researchers using experimental and quasi-experimental designs is to assume that a statistically significant difference in the right direction is evidence of merit, whereas failing to attain statistical significance implies a nonmeritorious outcome. In evaluation, there is a need to look further than statistical significance—to practical significance.

A **statistically significant** result tells us only that any observed difference (or statistical relationship) is unlikely to be due to chance (e.g., a fluke sample yielding unusual data).

A **practically significant** result is one that translates to real impact on people's lives (e.g., the difference has a noticeable and nontrivial effect on functioning or performance).

When determining the merit of a particular outcome, it is important to take into consideration both its practical significance and its statistical significance (or the qualitative equivalent).

Using Comparisons to Determine Relative Merit

To determine the relative merit of a process, an outcome, or a cost criterion, it is important to identify useful comparisons. For example, the evaluation team might "benchmark" process, outcome, and cost criteria against what has been achieved elsewhere (e.g., by other evaluands of a similar scope).

Benchmarking is a systematic study of one or more other organizations' systems, processes, and outcomes to identify ideas for improving organizational effectiveness. It has been used in manufacturing for years and is now widely used throughout business and industry.

Benchmarking most commonly refers to a process of gathering comparison data about what organizations in similar or related industries are achieving. This approach to benchmarking focuses primarily on collecting quantitative data about process efficiency, outputs, outcomes, and costs.

Sometimes organizations undertake their own benchmarking studies. In such cases, two or more organizations (often doing business in different sectors) each agree to allow teams from the other organization(s) to come in and study their practices, compare results, and discuss how improvements were made. These studies are typically heavier on qualitative data gathering (e.g., observation of processes, interviews with key stakeholders), although they still look at specific quantitative data.

The following example, taken from an evaluation of a large-scale organizational change effort, illustrates the use of a simple rubric to determine the relative merit of evaluand components. In this case, the components were clusters of initiatives that formed part of the change effort. The rubric is designed to assess the relative cost-effectiveness of each cluster of change interventions (Table 8.7).

After drawing up the simple rubric, the next task was to rate each component (i.e., cluster of organizational change interventions) on the given scale. As an example, one of the components was a set of interventions intended to create a more strategic and constructive work environment. To this end, the organization had implemented a culture survey and a climate survey, that is, quantitative instruments sourced from separate providers that were to deliver periodic "snapshots" of the organizational culture and climate that managers would then reflect on before making changes in their business units.

Table 8.7 Rubric for Determining the Relative Merit of Organizational Change Interventions

Relative Merit	*Description*
Superior practice	Clearly the most cost-effective of the available alternatives
Above average	Considerably more cost-effective than most alternatives
Average	Approximately as cost-effective as most of the alternatives
Below average	Considerably less cost-effective than most alternatives
Inferior practice	Clearly the least cost-effective of the available alternatives

Table 8.8 outlines the main costs and benefits of the two surveys and provides a list of alternative options given the resources the organization had at its disposal. By applying this information to the rubric in Table 8.7, this component of the organizational change effort was rated "below average."

MAKING PRACTICAL USE OF MERIT DETERMINATION RUBRICS

There you have it—the basics of rubric methodology for synthesizing qualitative and/or quantitative data to draw conclusions about the merit of an evaluand on a particular dimension, subdimension, or component. Naturally, a lot of the finer points, such as how you might go about talking to various informants (e.g., evaluation team members, experts, other stakeholders) about the definitions of levels within the rubrics and resolving any differences of opinion, had to be skipped over here. The main thing to remember is to keep an open mind, draw in the views of all those who can help you to make good sense of the data, and do not let any one group drive the agenda to the exclusion of other points of view.

Table 8.8 Costs, Benefits, and Alternatives for Two Related Components of an Organizational Change Intervention

Component	*Costs and Benefits*	*Alternatives*
Improvement of work environment: 1. Culture survey 2. Climate survey	• Large financial outlay exists for each survey (several hundred thousand dollars each). • Culture and climate surveys are scheduled to alternate each year for several years. • Organization receives information with a lot of overlap. • Some managers believe that the information is interesting; only a few managers believe that it is useful and report having used it to make some changes. • Employees report seeing no substantial change arising from the surveys, which suffer from a poor response rate and widespread cynicism.	• Just one of the surveys would give sufficient information about the organization's climate or culture (there is a large overlap between organizational culture and organizational climate when both are assessed using quantitative methods). • A focused questionnaire would address only the issues that the organization faces (and those aspects of the climate that affect such issues) rather than provide a full cultural profile. This could be developed with managerial input and reported back in a session that generated ideas for improving problem areas. • One of the above could be combined with manager and employee focus groups and interviews. These would provide two different perspectives that could be integrated to provide a much more complete picture of the organization's underlying culture than two surveys ever could.

By now, you may find yourself in one of two camps. Either you are saying, "Well, that's not exactly rocket science," or you are holding your head in your hands and saying, "You go to all that trouble for just one teeny tiny subdimension? How on earth am I going to get an evaluation done in real time?" The methods described in this chapter are not intended to plunge you into paralysis by analysis. Sure, there will be times on very extensive evaluation projects where this level of analysis (right down to rubrics on each subdimension) is appropriate. But there will be other times when you simply cannot go into that level of detail.

The intent here was simply to present a tool that can be used when appropriate to find answers to evaluation questions or that can be used to help clearly explain the rationale behind your conclusions to a client or critic. The most important point is not to either (a) resort to "smoke and mirrors" in your interpretation of evaluative data (i.e., present determinations of merit as just your own judgments or impressions) or (b) throw your hands in the air and proclaim that it cannot be done. We use rubrics again in the next chapter, not only to determine merit for another example but also to blend multiple sources of data more systematically so as to draw evaluative conclusions about an entire evaluand or its components (i.e., the synthesis step).

ADDITIONAL READINGS

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

- Benchmarks, benchmarking
- Merit
- Rubric
- Worth

Camp, R. C. (1995). *Business process benchmarking: Finding and implementing best practices.* Milwaukee, WI: ASQC Quality Press.

EXERCISES

1. Recall the earlier example of a training program that you have been asked to evaluate. The program was designed to help young unemployed people to seek and obtain work effectively. Although the "proof of the pudding"

is whether the participants actually find jobs, one other important outcome (a little further upstream) is the extent to which they apply the skills they have learned when they hunt for jobs. One of these skills is effectively tailoring résumés for different job applications. Draw up a rubric to show how you would translate into a rating (e.g., "very good") evidence of how well the program recipients as a group did on tailoring their résumés to the specific jobs for which they applied (on half a page or less).

2. Choose the most important dimension of merit for your evaluand. Draw up a rubric that shows how you will interpret data from at least three different sources (including at least one qualitative source and one quantitative source) to determine the merit of your evaluand's performance on that dimension.