

# Phase 1

## **IDENTIFY STAKEHOLDERS, EVALUATORS, AND EVALUATION QUESTIONS**

“Chance favors prepared minds.”

—*Louis Pasteur*

### **Chapter Objectives**

- Remind the evaluation team of the many types of stakeholders whose views might need to be considered during the assessment
- Review important considerations that must be addressed when assembling an evaluation team
- Offer evaluators general methods that can be used to generate effective program evaluation questions

The recommendations that derive from an HR program evaluation will be driven by the nature of the questions that are posed and the stakeholders involved. For example, in evaluating a succession planning process, the chief executive officer may have expressed concerns about the organization’s current mix of leaders and how the upcoming retirements of key players could limit achievement of strategic business objectives and depress shareholder value. He wants immediate answers as to how the organization’s succession program can be positioned to populate the talent pool for all fifteen critical leadership positions within nine months of implementation. The outcome of this evaluation will likely focus on what enhancements need to be made and resources need to be allocated to the current succession program

to achieve this high exposure and aggressive goal. On the other hand, a directive from your manager to evaluate all of your training programs and make recommendations for cutbacks will result in a different series of questions related to the overall worth of each program and their differential impact on the organization's bottom line. The outcome of this type of evaluation would focus on which of the training programs should be dropped based on its relative contribution.

In many cases, stakeholder groups, evaluators, and evaluation questions will be evident from the content area of the program and the events that led to the program evaluation. Any number of events could have precipitated the start of an HR program evaluation. These events can vary from a periodically scheduled review of a program that seems to be running smoothly (a review of the compensation system every five years) to the need for a program to be certified (safety in a nuclear power plant) to a major revamping of a program caused by a catastrophic (a death in the workplace) or a high-visibility (publicized race-based incident in one restaurant for a chain) event.

Regardless of the reason for the evaluation, three key types of identification activities must occur, before even planning to carry out the evaluation. Specifically, the first phase of our HR program evaluation focuses on the identification of stakeholders, the evaluation team or evaluator, and the questions that need to be answered. This chapter examines factors that can help HR professionals systematically determine each of the three.

## **Identify Stakeholder Groups**

A stakeholder in an HR program evaluation is any person potentially affected by the HR program being assessed or the subsequent decisions (for example, to leave a benefits program unchanged) that result from the assessment. Stakeholders can provide important perspectives about the various aspects of an HR program. With in-depth knowledge about some parts of the program or the total program, stakeholders are well-suited to help evaluators focus on the most critical issues to be addressed in the assessment. In addition, they may be able

to identify some of the criteria for judging how efficiently and effectively the HR program is operating.

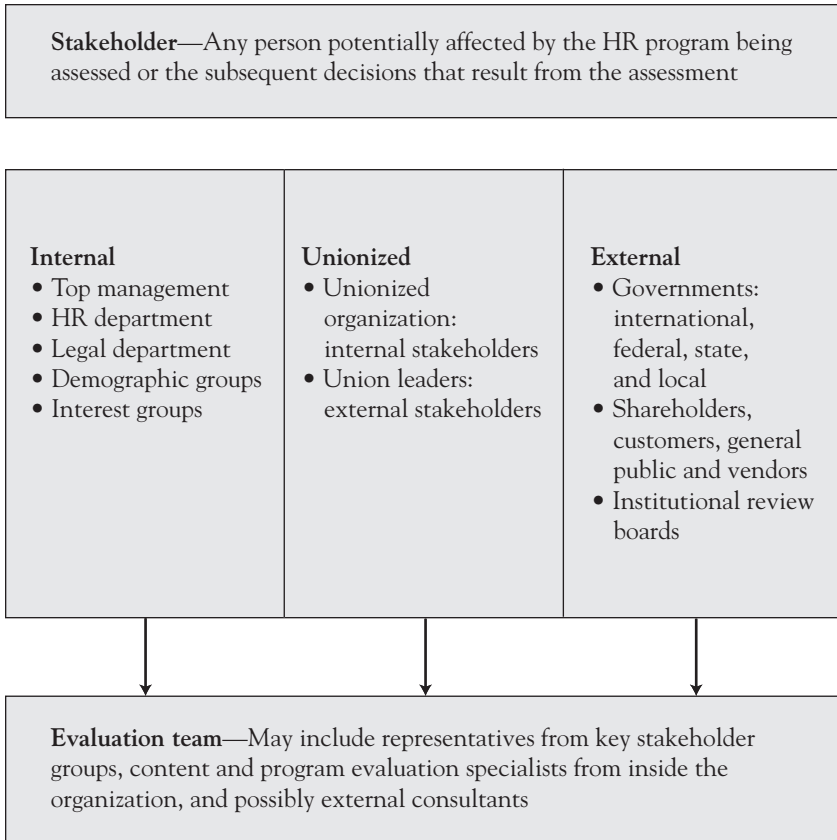
The evaluation team should involve as many types of stakeholders as possible in initial discussions. Expanded involvement enhances buy-in to the evaluation processes, takes into account the many different perspectives among stakeholder groups, and lessens potential resistance to change. It is also sometimes wise to include a representative group of stakeholders as an advisory panel for the evaluation, or even as part of the evaluation team of content and program evaluation specialists. Rossi and Freeman (1993) maintained that evaluators must understand both their relationships with the stakeholders and the relationships among stakeholders in order to conduct their evaluation effectively.

A convenient way to identify stakeholders for a specific evaluation is to think of three general categories of such individuals: internal, unionized, and external stakeholders. Figure 1.1 shows some of the more common types of groups that we have encountered in each category. Those groups will be discussed in the remainder of this section.

### **Decide on Internal Stakeholders First**

Because HR programs generally affect everyone in the organization, nearly all the organizational members will have vested interests in any changes that result from a program evaluation. Some internal stakeholders might resist such change. For example, an evaluation examining possible workplace redesign could result in an organizational unit being moved to another area of the headquarters and that unit's middle managers having smaller offices, less privacy, and less opportunity to interact with upper management. To minimize the apprehension resulting from possible change, selection of the advisory stakeholders (spokespersons for each stakeholder group) should be made apparent to all stakeholders and progress as quickly as possible. People often deal better with change when they are involved in the process, and the selection of recognized leaders from the key

**Figure 1.1**  
**Common Stakeholders in HR Program Evaluations**



stakeholder groups offers an opportunity for those groups to have input into the evaluation.

Membership in one internal stakeholder group typically overlaps with membership in other internal stakeholder groups. The net result of the overlapping membership with regard to organizational change is that members may lose in one way and gain in another way. Using the redesign example again, members who drive to work now have the post-redesign benefit of having parking spaces next to the entrance to their office rather than much farther way. But the net re-

sult of the change varied according to other types of group membership. At one extreme, lower-level employees' office spaces remained much like before, and those employees got better parking. At the other extreme, up-and-comer middle managers who took mass transit lost informal face time with upper management and will have to walk farther to get to their office from the mass transit stop.

The prior paragraphs hinted at the composition of some internal stakeholder groups. Let's look at five types of internal stakeholder groups that will be important in all or almost all HR program evaluations in larger organizations. These internal stakeholder groups are (1) top management, (2) staff in the HR department and (3) the legal department or outside counsel, (4) different demographic groups, and (5) special interest groups. Figure 1.1 schematically shows these and other types of stakeholder groups. Smaller organizations may not include all of these stakeholder groups, such as legal or union representatives.

***Recognize the Value of Top Management Stakeholders.*** Organizational leaders constitute one of the most valuable and readably recognizable stakeholder groups for an HR program evaluation. In some critical program evaluations, top management stakeholders (the vice-president level) could be on the evaluation team. Their inclusion would depend on factors such as the visibility of the HR program being evaluated, cost and scope of the program evaluation, and the impact of the program on other units in the organization.

While inclusion of a top management stakeholder on the evaluation team has advantages, it can also bring concerns such as perceptions that this stakeholder will unduly influence the evaluation's findings and recommendations. The inclusion of a top management stakeholder on an advisory panel is a strategy that allows for upper-level input. At the same time, it could lead to the negative consequence of perceptions that the evaluation team is merely going to recommend changes that top management wanted, regardless of what an unbiased evaluation would have shown.

Top management stakeholders may provide the actual evaluation questions that they want answered or indicate the general issues

that are to be addressed. The interactions with top management may include both individual and group meetings. The individual meetings allow each upper-level management member to provide his or her unique insights. In addition, one or more group meetings between these managers and the evaluation team leader might be needed to iron out differences that the various stakeholders have about issues such as milestones, budgets, and researchable questions.

***Gain Expertise from the Program Staff in the HR Department.***

Another easily recognizable group of stakeholders is the HR department, with the staff who deal with the strengths and weaknesses of their programs each day. The HR department staff will be interested because they will be directly affected by any changes that result from the program evaluation. Thus, conflicts of interest can occur for program staff. Moreover, serious negative findings from the evaluation could result in adverse consequences—including replacing an HR program, undesired personnel transfers, and poor performance evaluations—for members who are in charge of or work in the HR program that receives low marks at the end of the evaluation. Although rare and extreme, a program evaluation might result in termination of HR staff if incompetence or malfeasance is found. Conversely, HR staff may wholeheartedly embrace the evaluation as a way to upgrade an out-of-date or ineffective HR program that they have heretofore been unable to change, despite their desire to improve it.

Because at least one member of this stakeholder group will typically be on the evaluation team as either a member or as a liaison, the HR department will probably receive periodic feedback about what is being done and found during the evaluation. Information exchange between the evaluation team and the HR department stakeholders is essential during all stages of the program evaluation because implementation of any changes to an HR program will almost surely require participation by members of the HR department. Besides, if HR program staff learn about the findings and recommendations at the last minute, momentum for implementing changes may be slowed.

***Involve the Legal Department to Address Prohibited Practices Proactively.*** This stakeholder group will have insights into potential contractual and legal limitations that the evaluation team and the whole organization will face during and after the program evaluation. More often than not, the legal department (which could be outside counsel for smaller organizations) will want to be aware of the evaluation's scope, subsequent findings, and the potential impact of recommended changes on employees and the organization as a whole. This last statement is particularly true for HR areas (for example, selection and compensation programs) prone to employee challenges and litigation. Therefore, one of the organization's attorneys should be assigned to the HR program evaluation team as either a full member or liaison. Smaller organizations without internal legal representation may want to engage an outside employment attorney for this purpose.

The legal department should have detailed knowledge regarding the types of changes that might be limited by labor agreements (for example, alteration of the conditions of work) or contracts with vendors (for example, that prevent the customization of HR-related software). Additionally, the attorney should be able to advise top management and the evaluation team on applicable international, national, state, and local regulations that govern an organization's HR practices. More will be said about labor, vendor, and regulatory concerns in later sections on unions and external stakeholders.

***Ask Whether Program Participants from Some Demographic Groups Will Be Affected More Than Others.*** While most program evaluators would consider it a given to include evaluation inputs from participants who belong to various demographic groups, this might not be done in some (for example, paternalistic) organizations. For example, Posavac and Carey (2003) noted that participant input may not occur in situations in which reliance is placed on program staff's opinions to determine what is needed because there is a belief that the participants do not know fully what they need.

The organization's membership is made up of stakeholders from various demographic groups that might perceive either the need to keep the program as it is or to change it. A prime example of this point

Individuals selected to represent demographic groups (for example, as a member of the evaluation team or in individual or focus-group interviews) have much pressure on them. Many of these stakeholders may feel uncomfortable speaking for all organization members who share their demographic characteristic(s). Moreover, others may feel that they are treated as tokens. Therefore, the procedures for selecting demographic group representatives as well as the evaluative role of the representatives should be transparent and objective.

relates to equal employment opportunity. Some stakeholder groups may feel that any gains made by another demographic group will negatively affect the group to which they belong.

Retirees are a quasi-internal demographic group. While they are not at work in the organization each day, they may be entitled to participate in some of the organization's HR benefit programs. These programs might include subsidized health insurance, retirement income plans, and special company-specific benefits (such as discounted fares for former airline employees and purchase discounts for former store employees). As has been seen in recent high-profile curtailments of retirees' benefits, legal, public-relations, and ethical concerns can be voiced when organizations change longstanding HR programs that affect retirees' quality of life.

Therefore, the evaluation team should think broadly when identifying demographic subgroups that could be affected by findings and recommendations from its HR program evaluation.

***Ask Whether Members of Some Interest Groups Need to Have Their Perspectives Highlighted.*** Similarly, transparency and objectivity should be present in the identification and roles of repre-

sentatives for interest groups (for example, parents with day care needs, employees who use mass transit, or people who participate in the company's stock plan). These stakeholders' interests may be tied directly to an HR program being evaluated.

While the stakeholders in some interest groups might willingly identify themselves, others may not. Individuals with drug problems would be an example of those who may fit into the latter category. There is little probability that drug abusers will identify themselves for interviews during a program evaluation to determine whether or not a confidential drug treatment benefit needs to be added to the employee assistance program. As a result, the organization may have little internal information upon which to base the potential need for or benefits derived from changes to some programs. In such cases, where the data central to the evaluation's objectives may be questionable, it is important to identify methodological limitations early in the evaluation process and then determine whether or not an evaluation with sufficient rigor and probability of providing meaningful findings can be performed.

### **Consider the Perspectives of Unions and Their Members**

Unions and their members fall into a gap between internal and external stakeholder groups. Rank-and-file union members are internal stakeholders, but upper-level leaders of the union(s) probably are not. The situation is complicated further because stakeholders from an affected union may receive support from internal stakeholders in other unions that will not be affected by the program evaluation or its findings. An example of such an action is illustrated when pilots, flight attendants, or mechanics unions support each other in negotiations on wage and benefits reductions.

The often adversarial relationship of an organization's management and union officials and their members can cause a program evaluation to be longer and more sensitive than it might be in an organization without a union. Obtaining unionized members' perceptions may

require pre-approval of surveys and attendance by a union representative when the members are interviewed. For example, Berk and Rossi (1990) suggested that an evaluation of an educational program would be seriously impeded if the teachers' organization recommended that its members not cooperate with the evaluator. Also, recommendations at the end of the evaluation might seek to change conditions of work (for example, safety or compensation) governed by a labor agreement. In such cases, implementation could be delayed or prevented, especially when bargaining units have not been kept informed throughout the HR program evaluation.

### **Don't Forget That There Are External Stakeholder Groups**

Although the external groups discussed in this section are not involved in the day-to-day HR operations of an organization or the program evaluation, they can exert significant influence on internal HR decisions. Therefore, the evaluation team should cast its net widely to identify external stakeholders who might be affected by the way that an HR program is administered or the decisions that result from an evaluation of such a program. The list of potential external stakeholders includes governmental agencies, stockholders, customers, the general public, vendors, and institutional review boards.

***Assess Governments' Positions.*** An evaluation team would rarely seek direct input from governmental agencies, but the Joint Committee on Standards for Educational Evaluation (1994) correctly noted that laws and regulations are one basis for interpreting program evaluation findings. Consequently, an assessment of compliance to applicable rules and regulations needs to be integrated into HR program evaluations. Among other things, governments often influence HR programs by establishing minimum HR program standards that organizations must meet. These minimum standards could be established through laws, executive orders, court cases, and other materials. Depending on how dispersed their operations are,

organizations may need to comply with legal standards enforced by international, national, state, and local governmental agencies.

- *International regulations.* With more and more organizations having offices in multiple countries, organizations must be cognizant of international HR regulations. For example, the European Privacy Directive could greatly impact program evaluations for multinational companies because this directive provides common rules for collecting, maintaining, and transmitting personal data about employees, even when the data are transferred to non-European Union countries. The U.S. Department of Commerce's Safe Harbor Provisions aid in the uninterrupted flow of such data for U.S. companies that agree to abide by certain constraints regarding the collection and dissemination of data.
- *National regulations.* Organizations that operate only within the United States still must comply with HR-related regulations from myriad federal agencies. Let's examine three of the many types of compliance concerns that an organization might face during an HR program evaluation. A federal agency like the Equal Employment Opportunity Commission may take legal action to correct a workplace situation that is thought to violate civil rights standards. Another federal agency monitors workplace health and safety, with serious violations potentially resulting in the immediate closure of a work site. Other agencies such as the Bureau of Labor Statistics might require the periodic submission of data on the organization's labor force as part of its efforts to generate national labor statistics.
- *State and local regulations.* State and local governments sometimes issue minimum standards that are more restrictive than those issued by international and federal governments. For example, some U.S. municipalities have established minimum wage laws that provide for a higher minimum wage than

that required by federal regulations. Such wage considerations would be of concern during the evaluation of a compensation system.

***Consider the Needs and Views of Stockholders, Customers, the General Public, and Vendors.*** Wood (2005) provided a succinct but thorough list of other types of organizational stakeholders. Among the others that she noted are financial analysts and markets, financial institutions, customers, local communities, environmental and consumer protection groups, and suppliers who provide the input materials for the organization. We have addressed these stakeholders as four groups: stockholders, customers, the general public, and vendors.

Stockholders' expectations of dividends and capital growth indirectly limit the scope of some HR programs in private-sector organizations. In particular, changes in market share, downturns in stock price, and other economic influences may result in expectations that the organization must tighten its belt. That belt tightening might include a rollback of company benefits or the elimination of jobs. Many times, the belt tightening is done without the benefit of a formal program evaluation; rather, it is done when upper management *thinks* that the most money can be saved with the least negative impact.

If the organization receives bad publicity because of events such as actual or perceived racial discrimination against employees or customers or violations of environmental regulations, demographic groups of customers could decide to refrain from buying the organization's goods or services. Similarly, the general public is an external shareholder in a more general sense. For example, various publications provide lists of the best employers. Being named a best organization can make it easier for an organization to recruit and retain top talent. Many of the characteristics used to identify top organizations pertain to HR practices. Program evaluation can make the HR programs more attractive by making them work more efficiently and effectively.

Vendors are important external stakeholders because organizations may have contracts to purchase a vendor's HR-related goods and services across multiple years. Also, the vendors may have contracts

that prevent the modification and duplication of HR-related software for such things as an HR information system or an automated 360-degree feedback program. In those cases, HR program evaluations may be constrained with regard to recommendations for immediate improvement and therefore might need to design both shorter- and longer-term strategies for improving a program.

***Engage Institutional Review Boards When They Are Present in the Organization.*** In some organizations, internal review boards blur the lines between HR research and practice. While these boards have typically monitored an organization's research using human participants, some governmental agencies have questioned whether all HR-related surveys need board approval. This could be a growing trend for HR programs in academic and healthcare institutions that are required to maintain institutional review boards to protect the rights of human participants in research projects.

### **Identify the Evaluation Team**

When feasible, we believe that HR program evaluations should be performed by teams rather than by a single evaluator. A single evaluator may have limited the choice of evaluation methods to those with which the individual feels most comfortable, rather than to the most appropriate methods for the given situation and program being evaluated. Another problem encountered when using a single evaluator could involve the interpretation of findings. Often, more than one explanation can be given for a finding, but the ability to see patterns and alternative interpretations is probably more limited when a single person performs the whole program evaluation. Because of these and other problems that result when a single evaluator performs an assessment, the remainder of this book assumes that a team will be used for all HR program evaluations. Of course, the size of the organization or program being evaluated may dictate that a single evaluator conduct the evaluation. Successful evaluations can certainly be conducted with a single evaluator. We are simply recommending

that, when it is possible, a team-based evaluation brings with it a number of advantages, particularly in large-scale, high-profile evaluations. For example, the evaluation of a compensation system within a large organization that is experiencing gender bias claims would minimally require that the evaluator possess content expertise in compensation, performance assessment, and statistics; have the time and resources to conduct the necessary policy reviews, data collection, and analyses; and have well-developed oral and written communication skills to present the findings, conclusions, and recommendations to key groups and individuals, such as top management in the organization and attorneys inside and outside of the organizations. The required skill sets and diversity of tasks involved with this sort of evaluation are more simply accomplished by a team. We recognize, however, that not every organization will have the luxury of using an evaluation team.

When determining the evaluation team, consideration needs to be given to the team's size, composition, leadership, and charter.

### **Ask, "How Big Should the Team Be?"**

The answer to this question is a very ambiguous "as big as it needs to be." The authors have worked on teams that had as few as one primary member (one of the authors) who examined a very limited HR issue to teams that have had as many as ten members. A team that is too large may have trouble coordinating its activities, even with something as straightforward as identifying times when everyone can meet. Also, if the evaluation team is using consensus as its means for decision making, large teams could use more time than smaller teams. Conversely, a team that is too small may have trouble completing the evaluation in a reasonable period of time. Having a very small evaluation team decreases the likelihood that some tasks can be done simultaneously. For example, the person who is conducting the interviews may be the only evaluation team member who has experience designing surveys. In such a case, tasks must be done serially, rather than simultaneously.

Some of the factors that are particularly important for determining the size of the evaluation team are the size of the organization and program, how long it has been since the program was last evaluated, and how soon the evaluation has to be completed. The size of the team may increase and decrease at different times during the program evaluation if temporary members are used to perform certain activities. While it is important to obtain inputs from a wide variety of stakeholders, those inputs can be obtained during interviews and other data collection efforts. More specifically, the formal evaluation team does not need to include members of all stakeholder groups as part of the day-to-day working group.

### **Ask, "Who Should Be on the Team?"**

Decisions have to be made about where the organization will obtain its evaluation team members. At minimum, most teams will probably include evaluators with skills in the program content area, general program evaluation procedures, and statistical analysis. The need for additional specific skills such as interviewing and survey design will become apparent as the evaluation team starts identifying the data collection and analysis methods that it intends to use. Love (1991) offered six tips for building effective internal evaluation teams. Those tips were to:

- Keep the team small,
- Emphasize the need for high performance,
- Reward team members and the leader,
- Focus on people instead of methodology,
- Inventory team members' skills, and
- Use project management tools to benchmark success.

The need to balance team size considerations against the needs for diverse skill sets suggests that the net should be cast widely to obtain the best combination of evaluators for the team. Fink and Kosecoff

(1978) suggested that, for small evaluations (and therefore, probably small evaluation teams), evaluators should possess a wide variety of skills. For larger evaluations (where more evaluators will probably be involved), they recommended that the team include members with specialized skills in statistics and report writing. The team should also represent all key stakeholder groups. Some of the representation may be obtained by including someone from one stakeholder group on the evaluation team. Other stakeholder groups may be represented indirectly by periodically seeking input from and providing feedback to an advisory panel of representatives from the other stakeholder groups.

According to Berk and Rossi (1990), another important characteristic of evaluation team members is that they should not be people “who prefer to avoid controversy, or who have difficulty facing criticism” (p. 14). They also noted that criticism is often political and that methodological flaws in the evaluation are one way in which the attacks are made. We believe that the best way to ameliorate these often real concerns is (a) through the selection of a team leader—and to a lesser extent, team members—who is well-respected and perceived to be objective throughout the organization (for example, government blue-ribbon panels to study politically sensitive problems) and/or (b) with the addition of external consultants who have expertise in program evaluation in the content area in order to minimize methodological weaknesses. Regardless, the evaluation will go more smoothly if all of the evaluation team members have good interpersonal skills that allow them to be collaborative as well as firm if/when the need arises.

***Select Team Members from Three Categories of Evaluators.*** In general, members of the evaluation team fit into three categories: HR staff, other organizational members (that is, those not from the HR department), and external consultants. Each type of evaluator brings potential strengths and weaknesses to the evaluation process. As might be expected, one category of evaluators might bring strength to the evaluation for one issue, but use of the same type of evaluator for a different evaluation activity could result in a potentially flawed finding.

Members of the HR department constitute one category of potential evaluators. Since they work every day with the various HR programs and HR program administrators, they may have program-related insights that are not available elsewhere. At the same time, their closeness to the program may prevent them from recognizing opportunities for improvement.

Non-HR organization members are another category of potential evaluators. The evaluation team leader might choose other internal staff for a variety of reasons:

- The views of one or more of the five types of previously identified internal stakeholders should be represented.
- A unit (for example, manufacturing) is particularly affected by the quality of the program (for example, safety). Therefore, the evaluation requires a team member with specific content knowledge.
- An individual possesses a desired skill set (for example, an accountant's skill with budgets or a quality control manager's experience with sampling).

External consultants are a third category of potential evaluators. Organizations can contract with external consultants to (a) perform the entire evaluation by themselves, (b) work as part of the evaluation team during the entire process, or (c) complete specified evaluation tasks. If the organization sees implementation of recommendations as a continuation of the program evaluation, an organization may choose to exclude the external evaluators from participating in Phase 6 of our approach. While such a step might result in less biased input from the external evaluators, it could also deprive the organization from using consultants who know much about the organization and the HR program.

***Consider the Relative Strengths and Weaknesses of the Three Categories of Evaluators.*** Worthen and his colleagues (1997, p. 44) noted that “No professional association or government agency has yet

assumed or accepted broad responsibility for licensing or certifying the competence of evaluation practitioners. . . . In the absence of certification and licensure, uninformed amateurs and unprincipled hucksters can do much mischief and greatly tarnish the image of program evaluation in the process.” While this conclusion should serve as a caution, most organizations should find a wealth of qualified professionals, possibly internally or externally, who can contribute to their HR program evaluations.

Table 1.1 shows eight issues that are important to an evaluation and our general assessment of the relative strengths and weaknesses that each category of evaluator might bring to a program evaluation. Our generalized conclusions in Table 1.1 are designed as guidelines for an organization embarking on an HR program evaluation. The situations considered in our generalized conclusions may, however, differ from the actual situation in an organization because of the particular characteristics of the organization’s staff or the consultants. For example, if the HR staff includes a recognized expert for the type of program being evaluated, some of our general statements may prove incorrect. Likewise, our conclusions about external consultants may be incorrect for such issues as frankness if the external consultants sold the HR program being evaluated to the organization and continue to be paid for maintaining the program.

One thing is readily noticeable from the shading in each column of Table 1.1. Each category of evaluators brings certain strengths and weaknesses to the program evaluation.

To aid readers, we have shaded the relative strengths and weaknesses so that the darker shading indicates there is often a major concern, lighter shading denotes a caution, and no shading suggests that the category of evaluator is usually well-suited to handle the issue. For example, relative to the other categories of evaluators, HR department personnel have the advantage in knowing the HR program and the organization, but the evaluation team might want to exercise caution before using HR staff

in evaluation tasks for which personal frankness about aspects of the HR program is a primary concern.

The team leader should consider our generalizations for the eight issues, along with potential team members' actual strengths and weaknesses, when choosing team members and the tasks that each member will perform. Such planning can go a long way toward ensuring the validity of evaluation findings by maximizing team members' strengths and minimizing their weaknesses. Introspection of the team's strengths and weaknesses during this initial phase can avoid problems that it will be too late to address in the final phases of the HR program evaluation.

### **Ask, "Who Should Lead the Evaluation Team?"**

The team leader is often an internal person or an external consultant who has the best combination of position power and technical skills, with position power typically carrying more weight—rightly or wrongly—than technical skills. The leader must be someone who has immediate access to top management and other key stakeholders. Having the ability to get onto people's calendars with little prior notice is particularly important when (a) providing periodic updates about the progress of the evaluation, (b) obtaining more time or resources if something unplanned occurs, or (c) negotiating entry into organizational units where managers may not want to provide their subordinates for program evaluation tasks such as surveys or interviews.

The team leader will be particularly important for the planning that occurs during the second phase of the program evaluation and for keeping the team within its timeline and budget. As with other members of the team, the leader may be devoted full- or part-time to the program evaluation.

### **Ask, "Should the Evaluation Team Write a Charter?"**

It is a good idea for the evaluation team to create a team charter that specifies the evaluation team's mission, membership, and general evaluation plan, including milestones and resource needs. In essence, it is

**Table 1.1 Assessing Various Types of Program Evaluators Against Numerous Criteria**

Criteria Against Which Evaluators Are Judged	Types of Program Evaluators		
	HR Staff	Other Internal Staff	External Consultants
Knowledge of the organization and the organization's HR program	Have in-depth knowledge of both the organization and the HR program being evaluated	Have in-depth knowledge of the organization but may have limited knowledge of HR program specifics	Will lack this initially if it is the first time that they have worked with the organization
Knowledge of industry best practices for specific HR program and program evaluation	May be aware of best practices; but awareness depends on networking, keeping up knowledge of the literature, and other such factors	Are unlikely to know about best practices for an area outside their own areas of expertise	Are probably exposed to the various practices of multiple clients and general practices for program being evaluated
Credibility with decision makers/stakeholders	May have their knowledge, skills, and abilities underestimated by their organizational peers	May partially compensate for their lack of HR expertise by being (a) a high-level stakeholder or (b) a representative of a stakeholder group	Have high credibility at time of entry because the organization formally picked them—supposedly for their expertise with the HR program
Potential constraints to a frank evaluation	May not be totally objective because the program belongs to the HR department, and a negative assessment may reflect badly on the department (or even get them fired)	Have less-direct reasons for biasing the evaluation findings than do the other two types of evaluators	May not be totally objective because they (a) are also focused on building client relationships or (b) helped design the HR program being evaluated

Time and dollar cost of the evaluation	May work on pieces of the evaluation as part of their normal HR program duties, and program familiarity may result in more rapid completion of those pieces	Take time away from their normal jobs to do the evaluation, and doing work outside one's area of expertise probably results in poorer return relative to cost	May be (a) more because consulting rates are high compared to internal salaries or (b) less because expertise decreases the time needed for an evaluation
Design of a practical implementation plan for program changes	Can identify practical concerns better than others, but may attempt to limit the changes because of (a) internal HR turf concerns and (b) a desire to keep expectations low to reach objectives	Probably have limited HR-specific knowledge and thereby have limited ability to design the plan to fit with other HR programs	May wish to add features that result in (a) "Cadillac" version of program or (b) steps that do not fit the organization's culture
Program evaluation and HR-program knowledge left in the organization after the evaluation	Keep some evaluation expertise in-house because a member of the HR staff will have been a member of or a liaison to the evaluation team	Will lose their HR knowledge as a result of disuse, and transfer of new evaluation skills will probably be limited because transfer is to new content domain	Take their knowledge with them when they exit the organization unless additional consulting occurs
Fresh perspective	Could be overly conscious of former constraints that no longer exist	Might consider internal constraints but could stimulate out-of-the box thinking	Are not bound by perceived organizational constraints

Adapted from J.C. Scott, J.E. Edwards, & N.S. Raju's April 2002 workshop, *Program evaluation for human resources: The art and science of measuring success*, and later published and copyrighted (in 2003 by Sage Publications) in D.S. Rose & E.J. Davidson's chapter, Framework for human resources program evaluation, in J.E. Edwards, J.C. Scott, & N.S. Raju's book, *The human resources program-evaluation handbook*.

a contract between the team and top management about how the HR program evaluation will be conducted. Love (1991) provided extensive lists of contract responsibilities and terms of reference that evaluation teams could use. These lists are useful as reminders of potential issues that might need to be included in either a legally binding contract with external consultants or a psychologically binding contract between an internal evaluation team and management.

Steinhaus and Witt (2003) noted that a charter signed by top management lends tremendous influence to the team's efforts in securing the necessary organizational resources. While a key purpose of the charter is to ensure buy-in from top management, a secondary goal of the charter development is to force team members to think through key questions and issues in advance of taking action.

### **Identify Evaluation Questions**

How the evaluation questions are posed has implications for the kinds of data to be collected, the source of the data, data analyses, and the conclusions that can be drawn. Therefore, the evaluation team must arrive at evaluation questions that are doable within the constraints that the team will face. Still, the constraints should not typically be the primary concern at this early stage in the HR program evaluation process. If a sufficient business case can be made for a more in-depth evaluation, top management may choose to devote more resources to the effort and lengthen the timeline.

Berk and Rossi (1990) noted that program evaluation questions can be addressed at varying levels. "When great precision is needed and ample resources are available, the most powerful evaluation procedures can be employed. When the occasion demands approximate answers or when resources are in short supply, 'rough and ready' (and, usually, speedier) procedures can be used. Correspondingly, the answers supplied vary in quality: The findings of some evaluations are more credible than others, but all genuine evaluations produce findings that are better than speculation" (p. 34).

The team needs to be aware of the three general types of evaluation questions as they begin developing the questions. The ability to distinguish among the types is a key to generating and refining questions to meet the information needs of key stakeholders. Also, the team must identify the measures of success—that is, criteria. We discuss desirable characteristics of criteria in the last part of this section, but it is impossible for us to identify in this book the criteria for each specific type of HR program. Our *Human Resources Program-Evaluation Handbook* (Edwards, Scott, & Raju, 2003) contains actual criteria that subject-matter experts identified for evaluating over twenty types of HR programs, including personnel selection, appraisal, and training.

### **Determine the Types of Evaluation Questions That Match the Evaluation Objectives**

Different types of questions lead to different types of answers. And most large-scale HR program evaluations are likely to include more than one of the three types of questions: descriptive, normative, and impact.

*Use Descriptive Questions If Simple Information Is Needed.* This type of question results in information about the HR program's specific conditions, processes, or contexts. An example of a descriptive question is: "What procedures are used to report the different categories of work accidents?"

Descriptive questions are important in providing key stakeholders with basic knowledge or a context for interpreting more complex information about another part of the HR program. At first blush, it would appear that organizations have already answered such questions with documents such as previously prepared reports and policy or procedural manuals. However, the information in the reports and manuals may be out-of-date or otherwise inaccurate. In evaluating a diversity program, descriptive questions might include the following:

- What topics are covered in the diversity awareness training program, and how is the information conveyed to training participants?
- What percent of the employees have participated in the training program, and does it vary according to organizational level?
- How does the organization track discrimination charges, audits, and lawsuits?
- Has the company collected fairness perception data (administered as diversity climate survey)?
- Last year, how many discrimination complaints were resolved at each stage in the organization's four-stage problem-resolution system?

***Use Normative Questions to Compare the HR Program Against Standards.*** This type of question results in data that identify how the organization's evaluated HR program compares to internal or external standards. An example of a normative question that uses an internal standard is: "How have the organization's injury rates changed during the last five years?" An example of a normative question that uses an external standard is: "How do this organization's injury rates compare to industry-wide averages?"

Continuing with the diversity program evaluation, normative questions may include the following:

- How have the diversity climate ratings changed over the past three years?
- How do the diversity climate ratings compare to those in best-practice companies in your industry?
- How has spending on diversity vendors changed over the past two years?
- For each of the last three years, how many discrimination complaints were resolved at each stage in the organization's four-stage problem-resolution system?

- How have your customer demographics changed since implementation of your diversity training program? How does this change compare to your competitors?

Normative questions that use internal standards let key stakeholders determine whether conditions in the organization's HR program are improving, staying the same, or deteriorating. A major constraint in asking such questions is that the organization may not have (a) the same kind of data from prior years or (b) prescribed standards or procedures. Data from surveys often have the former type of problem. If the items have not been asked before in one of the organization's surveys or have been modified to correct previously detected deficiencies, the organization is left with only baseline data for a future trend analysis. For the latter type of problem, the HR program may never have laid out prescribed objectives or standards (for example, amount of time allowed for issuing a travel reimbursement) such as those commonly found in a program's mission statement. When either type of problem exists, the evaluation team will be unable to answer a normative question that looks at internal standards.

Normative questions that use external standards let key stakeholders determine where their organization's HR program fits relative to programs in other organizations in the same industry or across all organizations. For some HR issues, this type of information is available from the government (for example, Bureau of Labor Statistics), industry clearinghouses/consortia (for example, national associations), or other sources (for example, newspapers to see the prevailing rate of compensation for locally recruited workers). HR issues with readily available external standards include salary, benefits, injury rates, and workforce availability in a given geographic area. Normative data are not readily available for a wide range of other HR issues (for example, per capita expenditures on training). However, certain associations, such as the American Management Association, the Conference Board, and consulting firms, conduct HR benchmarking surveys to provide a context for determining whether aspects of a given type of HR program are above average, average, or below average relative to those for other organizations that the association monitors.

Caution is needed when making comparisons to external standards. Some vendors of surveys, tests, and other HR-related products provide “norms” as part of their product. For example, survey vendors frequently offer clients their national or international norms as part of the survey package. Internal company norms that are developed as part of the survey data collection may be more useful when identifying strengths and weaknesses as well as prioritizing actions; however, the appeal of external norms cannot be denied.

One major problem with external norms is that the organizations included in the normative database are not necessarily representative of an identifiable population, much less representative of the specific workforce against which it is being compared. In such instances, the only population to which an organization conducting the program evaluation can compare itself is the population of other organizations that happened to use the diagnostic instrument. As a result, conclusions about being above or below average on the diagnostic instrument could be an artifact caused by differences in the size, profitability, geographic location, industry, and other characteristics of the organizations in the “norms” and have little to do with the characteristics of an HR program as such.

Likewise, concerns may also arise as the evaluation team attempts to determine how applicable the external norms are when the team compares findings for its HR program against those present in best-practices organizations. Underlying this point is an acknowledgement that what one expert judges as a best practice may not be similarly judged by another expert. For instance, some of the HR procedures by formerly esteemed dot.com organizations in existence around the turn of the century illustrate this point. The dot.coms were recruiting disproportionate numbers of the highly valued talent from more staid organizations and the pool of new college graduates. It would be easy to conclude that the dot.coms had superior recruiting programs, when one of the real reasons for their success in recruiting was probably the high levels of compensation that seemed all-but-certain from the stock options and a variety of unusual perks. A few years later, many of these “best-practices” organizations were no longer in exist-

tence, and their state-of-the-art recruiting practices were seen in a less-positive light. In these cases, some of their initial successes may have been less about the recruiting practices and more about the technology buzz that lured this top talent.

***Use Impact Questions When Cause-and-Effect Can Be Reasonably Assumed.*** This type of question focuses on data that reveal whether the observed conditions are attributable to the HR program. An example of an impact question is: “What effects have the organization’s new safety program had on injury rates?” In the diversity program evaluation context, impact questions might appear as follows:

- What effect has the diversity awareness training program had on employees’ perceptions of fairness?
- What impact has the diversity vendor program had on changing customer demographics?
- How has the leadership development program affected the diversity of applicants for executive positions?

Various degrees of uncertainty will enter into the answers to almost all impact questions used in HR program evaluations. The uncertainty is caused by the myriad factors that could be influencing results. Foremost among the factors is the fact that clear cause-and-effect conclusions require laboratory-type methods (for example, random assignment to intervention and control groups) that are seldom feasible in applied settings. For example, consider the following impact question: “What effects has the organization’s new safety program had on injury rates?” A decrease in injury rates could be the result of (a) an improved safety program, (b) a newer, more restrictive definition of what constituted a reportable accident, (c) new pressure from supervisors to not report some types of accidents, (d) a less harsh winter that led to fewer falls due to icy conditions, or (e) many other factors.

When answering an impact question, an evaluation team can try to eliminate alternative possibilities for program-related changes by looking at the central issues from different perspectives. Using different methods and sources to obtain fact-based data allows evaluators to determine with greater certainty (than would a single method or source) whether the program or other factors caused the observed impact on the outcome measure. This technique—sometimes referred to as triangulation—can be illustrated for diversity program evaluation. We could use surveys and individual and group interviews (different methods) to ask a number of the same questions about serving customers to both employees and customers (different sources) throughout the organization's sales regions. Consistent findings across different methods and sources would carry more credence than would findings that surfaced from only one method or source.

### **Develop and Refine Evaluation Questions**

While top management may dictate the program and possibly the general issues to be evaluated, the team is often responsible for translating the general issues into researchable questions that address stakeholders' information needs. A useful two-step process for generating researchable evaluation questions is to brainstorm potential questions and then to refine them.

**Brainstorm.** Brainstorming is an excellent way to identify a large number of initial questions for each issue. In their book on focus groups, Stewart and Shamdasani (1990) suggested that brainstorming appears to be most useful for generating ideas when there is no single best solution and that creativity in identifying ideas is facilitated by the airing of the different perspectives of the group members. Group members are encouraged to build on the ideas of others as well as generate new ideas. The questions generated in brainstorming sessions can be developed using the framework of the program's goals, processes, and desired outcomes. Including questions suggested by key stakeholders at this time avoids problems of

getting near the end of the evaluation and learning that some stakeholders' concerns were not surfaced, much less addressed.

Brainstorming in the development of researchable questions can be accomplished in many ways. We have found the following process to be useful across a wide array of organizational types. The first step is to assemble a group of individuals who represent the key stakeholders' perspectives and are knowledgeable of the HR program under review. Once assembled, the facilitator—usually the team leader or a team member with strong facilitation skills—asks the group to generate an exhaustive list of questions that address the HR program's goals, objectives, roll-out procedures, communication strategies, implementation processes, and anticipated outcomes. The facilitator's goal here will be to maximize the likelihood that all stakeholders' perspectives are heard and that the group agrees that an exhaustive set of questions has been generated.

***Refine Questions After Initial Issues Have Been Identified.***

Now that the team has identified questions that might be used to address the key stakeholders' initially identified issues, it is time to refine the list. A team could begin the refinement process by clustering all of the brainstormed questions into groups of two or more questions that address aspects of the same larger issue. This process might include listing questions on small separate sheets of paper and moving them around on a table until every question is assigned to an issue.

Grouping the questions provides several benefits. The team can eliminate redundant questions and identify evaluation gaps. Also looking ahead to the end of evaluation, it is easier to explain two or three findings within each of one to four major issues than it is to talk about eight to twelve seemingly independent findings. It is not enough to brainstorm and refine the questions; instead, grouping the questions starts the team on a path to developing their vision about how they will present the results, conclusions, and recommendations at the end of the evaluation. Envisioning the end state at each step in the six-phase HR program evaluation will maximize

the odds of producing accurate and meaningful findings, conclusions, and recommendations.

Finally, each question should be categorized as descriptive, normative, or impact. This step forces the team to consider again whether

The importance of starting with the right question is summed up by an unreferenced quote that is often attributed to the well-known statistician John Tukey: “An approximate answer to the right question is worth a good deal more than an exact answer to an approximate question.”

it has gaps in the types of information that *should* versus *will* be gathered and analyzed.

### **Attend to Desirable Characteristics When Selecting Criterion Measures**

A program evaluation will only be as good as the measures that are used to answer the researchable evaluation questions and judge program effectiveness. The ultimate goal for the evaluation team is to

deliver the most useful and accurate information to key stakeholders in the most cost-effective and realistic manner possible. Deficient criteria could hamper the evaluation team’s efforts to obtain stakeholders’ buy-in that the proposed methods will provide valid findings, conclusions, and recommendations when the evaluation is complete.

Although identification of specific criteria for every type of HR program is beyond the scope of this book, many researchers have provided lists of desirable characteristics for criteria as well as recommendations for the specific criteria. We have already referred to the work of the Saratoga Institute and researchers in the field of HR metrics (Fitz-enz, 2002; Sullivan, 2002). These metrics can provide a running start for establishing the criteria against which many HR programs can be evaluated. Obviously they should be checked against stakeholder goals and expectations and adapted as needed for the particulars of the program. Chances are, no criterion will embody all the desired characteristics to a high degree because of the complexities found in

human behavior, differences across organizations, legal restrictions, and other factors important to a successful HR program. Therefore, a key to good HR program evaluation is choosing multiple criteria and methods that lessen the shortcomings of any single criterion or method.

The Joint Committee (1994) identified, defined, and gave examples for thirty desirable standards or characteristics for program evaluation criteria. That particularly thorough list of standards is suggestive of important characteristics that should be present in the criteria or measures that will be used in all types of HR program evaluations. Our list of desirable characteristics for criteria overlaps significantly with the list that Steinhaus and Witt (2003) put forth. Like their list, our list of ten desirable characteristics can be grouped into three categories. Our categories are *measurement quality*, *relevance*, and *practicality*. The three categories, the ten desirable characteristics, and a description of each characteristic are presented in Table 1.2. To illustrate these characteristics further, the following paragraphs also describe conditions that occur when the criteria are deficient of the desired characteristics.

***Build Measurement Quality into the Criteria.*** Basic to any successful evaluation is the need to have faith in the quality of the criterion data used to arrive at the findings, conclusions, and recommendations. While individuals in professional fields such as psychometrics and survey development spend years learning techniques to maximize measurement quality, even less experienced members of an HR program evaluation team can help their more experienced colleagues optimize data quality. They would accomplish this by placing due emphasis on using criteria that are reliable, valid, measurable, observable, and unbiased.

Emphasizing the development and use of high-quality criteria could, however, be difficult if no one on the evaluation team has training or experience in the professional field of measurement. Pedhazur and Schmelkin's (1991, p. 3) comment about students is equally applicable to the same people who are later part of an evaluation team:

**Table 1.2 Desirable Characteristics for Criteria**

<b>Desirable Characteristics</b>	<b>Descriptions and Actions to Increase the Desirable Characteristics</b>
<b>Measurement Quality</b>	
Reliable	<ul style="list-style-type: none"> <li>• Measures are stable across individuals, time, and circumstances</li> <li>• The focus is on accuracy or precision of measurement</li> </ul>
Valid	<ul style="list-style-type: none"> <li>• Measure must address key evaluation objectives</li> <li>• Lack of validity implies that an area of interest is not being effectively isolated and measured</li> <li>• A valid question (a) focuses on a single issue, (b) is unambiguous, (c) is directly linked to key criteria, and (d) requires little interpretation by respondents</li> </ul>
Measurable	<ul style="list-style-type: none"> <li>• Measurements can be either quantitative or qualitative</li> <li>• “Triangulating” in on an answer by using multiple methods is one way to lessen the uncertainty caused by the measurement limitations for any one method</li> </ul>
Based on Observable Events	<ul style="list-style-type: none"> <li>• Others viewing the same data, events, and so forth should discern the same qualities/levels without relying on assumptions, judgment, or intuition</li> <li>• Standardizing procedures minimizes the influences of unfounded assumptions, judgments, or intuition</li> </ul>
Unbiased	<ul style="list-style-type: none"> <li>• Evaluation questions are reviewed for potential bias</li> <li>• Measures should not be biased toward certain groups</li> </ul>
<b>Relevance</b>	
Meaningful to Stakeholders	<ul style="list-style-type: none"> <li>• The criterion data are credible for key stakeholders</li> <li>• Criterion variables are selected with stakeholder concerns in mind</li> </ul>
Focused on Value Added	<ul style="list-style-type: none"> <li>• Criteria should emphasize how the program contributes to effectiveness or efficiency of the organization</li> <li>• Return on investment estimates are desirable, but questionable estimates can harm the program and evaluation</li> </ul>
Actionable	<ul style="list-style-type: none"> <li>• Criteria should be aimed at facilitating program improvement and structured to reveal improvement opportunities</li> <li>• Data collection should emphasize obtaining information that can result in program changes that are feasible and can be implemented</li> </ul>
<b>Practicality</b>	
Practical and Cost-Effective	<ul style="list-style-type: none"> <li>• Targeted results of program evaluation need to be realistic and not overzealous in their goals</li> <li>• Criteria should be accepted by key stakeholders</li> <li>• Costs need to be balanced with expected benefits</li> </ul>
Organizational Politics	<ul style="list-style-type: none"> <li>• Hidden agenda can derail even the best evaluation</li> <li>• Anticipating all stakeholders’ points of view can help a team mitigate the effect of organizational politics</li> </ul>

“Many students get the impression that no special competencies are necessary for the development and use of measures. . . . Unfortunately, many readers and researchers fail to realize that no matter how profound the theoretical formulations, how sophisticated the design, and how elegant the analytic techniques, they cannot compensate for poor measures.”

*Assess How Reliable the Criteria Are.* The measures being used to evaluate the criteria must be reliable in order for the team and stakeholders to have confidence in the accuracy of the HR program evaluation results. A reliable measure is one that will obtain consistent results, regardless of who is administering the measure and when it is being administered. Reliability can be built into criterion measures by standardizing instructions, structuring questions, pilot-testing the measure to ensure clarity, training evaluators to gather the data in the same way, periodically testing and calibrating evaluators to make sure they judge the same information in the same way, and collecting data from the population or a representative sample. Phase 3 includes a section that reviews common steps that evaluators can take to enhance the reliability of information during data collection.

There are at least four ways to quantify reliability. One, measurements might be taken with the same instrument at two or more times to see whether there is consistency in the attribute being measured (technically referred to as test-retest reliability). Two, measurements of an area of interest such as job satisfaction may be obtained using more than one instrument (all designed to measure the same variable) to see whether individuals who score relatively high or low on one instrument also score at the same relative position on the other instrument (alternate/parallel forms reliability). Three, we could have two or more raters evaluate the same individuals, processes, or other organizational issues of concern, and then we could determine the degree to which the different raters agreed on the assessments (inter-rater reliability). Four, if multiple raters or instruments are not available, reliability can be assessed through a measure of internal consistency for a single set of homogeneous items on an instrument such as a survey or test.

*Ensure That the Criteria Are Valid.* Validity refers to the degree to which a score on an assessment reflects what the assessment is designed to measure. In order to ensure that an HR program evaluation produces relevant and meaningful recommendations, it is essential that the criterion measures used to generate findings, conclusions, and recommendations are directly linked to the key evaluation objectives. A criterion measure is valid if it accurately assesses the evaluation design questions. A lack of validity implies that an area of interest is not being isolated and accurately measured.

Careful planning and analysis are positive steps that the evaluation team can take toward ensuring that the criterion measures are adequately addressing the key design questions. Again, pilot-testing is important for ensuring that the criterion measures are providing information that truly answers the questions about the HR program being evaluated. In addition, each of the other desirable characteristics of criterion measures impact validity to varying degrees. For example, a measure can be reliable but have no validity (for example, height can be reliably measured, but is not a valid predictor of employee turnover). However, an unreliable measure cannot be a valid predictor.

*Make Sure That the Criteria Are Measurable.* Both quantitative and qualitative criteria need to be measurable. The absence of measurable criteria opens an evaluation team's findings to criticism. It is very difficult for an evaluation team to defend its conclusions when the findings were based on general impressions drawn from unstructured interviews and other unsystematic methods. It would be much easier to defend the same conclusions if they were based on measurable characteristics obtained with multiple methods that systematized the data gathering to provide quantifiable information.

Quantitative measures (for example, performance ratings or years with the organization) provide countable information describing an HR program. They yield relatively straightforward measures that are generally understood when presented to stakeholders. In an evaluation of a safety program, a quantitative measure might be the num-

bers of accidents occurring during the six months immediately before and immediately after a safety awareness program was provided.

Qualitative information (such as attitudes and opinions) can be measured with data collection techniques such as interviews, surveys, and naturalistic observations. Quantification of qualitative information helps the evaluation team understand an HR program and the context in which it operates. Such measurement often involves coding open-ended or narrative answers or observations. For example, team members might code narrative answers from structured interviews by assigning them to different categories based on the issues raised in the answers, or the team members might rate the answers based on how positive or negative the answers were in tone. The evaluation team can then summarize the various aspects of the qualitative data by calculating the frequency with which each category of narrative answer or observation occurred.

*Base the Criteria on Observable Events.* Different evaluation team members are likely to arrive at different conclusions about the impact of a program if they are relying on assumptions and intuition, rather than on observations or hard data. To arrive at defensible conclusions, criterion data should be systematically observed and collected. In the process, the evaluation team should allow long enough for the full range of behaviors to be observed. Too short an observation period (for example, only the month before the performance appraisal ratings are submitted) may not allow sufficient opportunity to gather information on a process that uses a longer cycle time (for example, a performance appraisal year). Another potential observational problem is the Hawthorne effect—when “group members’ knowledge that they are being observed changes the quality and frequency of their behavior” (Zaccaro & Marks, 1996, p. 158). An HR concern like discrimination in promotion may challenge an evaluation team with regard to finding a way to “observe” the behaviors of concern, but accurate conclusions about potential discrimination can only be drawn through a systematic evaluation of the program’s various components and an analysis of its potential differential impact on protected groups.

Continuing with an example of possible discrimination, some stakeholders may assume that an organization's performance appraisal program is discriminatory and should be changed if protected group employees perceive the system is unfair. Before drawing such a conclusion about whether discrimination is or is not occurring, it would be necessary to define specifically what discrimination means in this context (that is, how it might manifest itself or be observed). For this type of situation, some evaluation team members might interview protected group members to determine their specific concerns and the observed and perceived discriminatory behaviors that they and others have experienced, while other team members could review hard data (differences in performance ratings, compensation, promotion rates, and tenure) to see whether the perceptions are supported by other data. If differences are found, the evaluation team would also determine whether they are isolated in a particular department or exist organization-wide. These are only a few of the steps that would need to be taken to conduct a thorough evaluation for this HR issue, but they highlight the primary point—criteria need to be operationally defined and conclusions must be based on observable measures.

*Make Sure the Criteria Are Unbiased.* From a measurement perspective, bias is a consistent type of error that occurs when a data collection, analysis, or interpretation process results in a score that is *consistently* too high or too low. While bias is often discussed in its narrower context of discrimination, the current context uses the broader meaning of bias that includes all types of consistent errors in measurement. A bathroom scale that is out of adjustment is a simple example of an instrument resulting in a biased measure—that is, one that is consistently too high or low. In an HR program evaluation, findings can similarly be biased for/against a total population or some group within the population. Depending on the severity of the bias, a biased finding can alter the conclusions and recommendations drawn from the data. It is therefore important that evaluators who are designing criterion measures be aware of the different types of bias that can impact criterion measures and take the necessary steps to

eliminate them during the design phase. The presence of bias should also be specifically assessed during the pilot-test of instruments and procedures.

One type of bias pertains to the use of culturally specific language. At its extreme, this issue would be of concern when a multinational organization wants to solicit information from its members who speak a language other than the one in which the original questions were written and answered. This issue may also be salient when different groups in the same country are asked to respond to written or verbal interview or survey questions. Some questions may be interpreted differently by members of the different groups. To minimize the chance that bias will enter the evaluation, a subgroup—representative of the population—should review the measures for this type of bias, either as part of the pilot-test or as a separate task.

Another type of bias that can impact data accuracy is the influence of evaluators' or stakeholders' assumptions about the criteria under study. This influence can lead to criterion measures being designed to consciously or unconsciously confirm preconceived notions. The end result is a measure that is less informative or accurate than it could be.

***Establish the Relevance of the Criteria.*** Suggesting that criteria need to be relevant might seem like stating the obvious. Much harder is operationally defining the relevance with characteristics that are equally transparent to stakeholders. Steinhaus and Witt (2003) listed and discussed three such characteristics: criteria that are meaningful to the stakeholders, focus on value added to the organization, and result in actionable recommendations.

*Consider the Meaningfulness of the Criteria to Stakeholders.* Early in the program evaluation, the evaluation team must gain key stakeholder buy-in on what is meaningful to them and will be used as the criteria for judging the effectiveness of the program. To gain that buy-in, the team must find criteria that are transparent and phrased in the stakeholders' language. Stakeholders' buy-in from the beginning of

the evaluation will lessen the likelihood that they will later raise questions about whether the “correct” criteria were studied. The buy-in is particularly important when the results do not come out as the stakeholders would have liked.

As we mentioned earlier in our section on identification of stakeholders, different stakeholder groups may hold very different perspectives from those of others about the evaluation, its purposes, and what the outcomes should say. Since the criteria used to conduct the evaluation will influence the eventual evaluation outcomes, arriving at meaningful criteria for all stakeholders can be a challenge.

Evaluations of compensation systems offer an example of how the meaningfulness of the criteria might vary by stakeholder type. For a city government examining whether it provides equal pay for equal work, some stakeholder groups might argue for higher or lower weight to be given to criteria that support how much they are paid. For example, white-collar stakeholders might push for more weight—and therefore more compensation—to be given for post-secondary education and the competing salaries paid to white-collar employees in local private-sector organizations. Other types of workers such as garbage collectors might advocate for more weight to be given to criteria that acknowledge the need to work in very cold and hot environments and perform physically strenuous work. Failure to address all stakeholders’ concerns with criteria that all of them agree are meaningful can doom the credibility of the program evaluation findings, conclusions, and recommendations.

*Focus on Value Added.* Value can be measured in many ways, but monetary value is typically the preferred metric for top management. While it is desirable to put a dollar value (for example, return on investment) on every HR function, that is sometimes not possible. When an HR department or program evaluation team tries to put a questionable dollar value on a function, the questionable estimate can sensitize those receiving the information to other assumptions that were made. And this additional sensitivity can re-

sult in criticisms of other findings and conclusions. A whole field of study has emerged in utility analysis, which explores ways to quantify the effect of different HR interventions and actions.

A good way to couch findings is to phrase the results in terms of the goals and outcomes stated for the HR program. For example, an on-site day care center's performance might be judged by looking at the turnover and absences of parents who used the center versus the turnover and absences of both parents who do not use the organization's center and non-parents. While it would be possible to place a dollar value on the avoided costs such as those for recruiting and training a new employee to replace departing employees, it may be sufficient to merely report other meaningful metrics that still show value to the organization without requiring the involvement of accountants and others to arrive at the dollar-valued metrics. Making assumptions about the dollar-based replacement costs for departing employees can result in the evaluation team opening its evaluation up to criticisms that could have been avoided by sticking with a different value metric.

*Determine Whether the Criteria Will Result in Actionable Recommendations.* A key to determining whether many HR program evaluations are deemed to be successful is to look at whether the evaluation resulted in the organization taking actions to make the program more efficient and effective. Information for information's sake is often not useful, and in some cases may be harmful. Among other things, the harm may be a denial that anything can be done to improve the program because the evaluators could not or did not identify alternative actions to improve or replace the HR program. Also, data collected without concern for how the data can be used in a subsequent action could set up those supplying information with unrealistic expectations that changes will be made based on findings.

To illustrate, if most organization members indicate on a survey that they are dissatisfied with their pay, the respondents might expect increases in pay. If respondents do not see changes, an additional

adverse effect could be decreased response rate on subsequent surveys. In past applied organizational research, focus group participants told one of the authors that they do not send back surveys any more because no one ever does anything with the information when a problem is identified. Our advice: Don't gather the data if you don't (a) *really* want to know the answer to what you asked and (b) hope to take some action based on the data.

***Focus on the Practicality of the Criteria.*** Teams will face many hurdles in the course of an evaluation, and they must arrive at practical means for dealing with such obstacles. Common hurdles addressed by our third category of desirable characteristics for criteria are the needs to select cost-effective criteria and to minimize the effects of organizational politics.

*Emphasize Practical and Cost-Effective Criteria.* With organizations emphasizing efficiency and cost-consciousness in all of their functions, it stands to reason that the criteria used in an evaluation must be practical and worthwhile relative to the resulting expense required to obtain the information. Attempts to gather perfect data can result in excessive time or monetary costs. For example, moving from an 80 percent response rate to a 100 percent response rate for an organizational survey may require more resources (for example, evaluator time as well as respondents' goodwill and time) than the additional data will be worth. Answers very different from those already obtained would have only minor effects on the *overall* organization findings. The remaining 20 percent could, however, be important if the non-respondents fell disproportionately into the same *subgroup* and findings for that *subgroup* were to be examined and reported separately.

Phillips (1997, p. 3) similarly noted that, at a more global level, the question, "Is it worth it?" should be asked before undertaking an evaluation. He cited an example in which a training course for ten supervisors cost \$5,000 and an evaluation of the course could easily

cost \$50,000. He concluded that the evaluation might not be of merit for a training program that cost \$5,000, but a \$50,000 training evaluation might be a worthwhile investment for a training program that cost \$750,000.

*Beware of Organizational Politics.* To some extent, the information for this issue has been alluded to when we discussed obtaining key stakeholder buy-in; but because of its importance to the success of an HR program evaluation, a few more words about this issue are warranted in this section on desirable characteristics for criteria. The effects of organizational politics on the choice of criteria and the ways that those criteria are defined conceptually and operationally can greatly impact the outcome of an HR program evaluation.

Royse, Thyer, Padgett, and Logan (2001, p. 326) stated, “agency directors, managers, employees may have strong motivations to present their programs or organizations more favorably than might otherwise occur in an unbiased program evaluation. The actions of these individuals can seriously affect treatment fidelity and make a complete mess of evaluation efforts.” In contrast to this desire to make a program appear favorable in an evaluation, many of us in the HR program evaluation field have also encountered situations in which key stakeholders have wanted evaluation results to show that a program should be dropped or some aspect of the program should be greatly modified.

## Conclusions

While we have dealt with the identification of stakeholders, evaluators, and evaluation questions as three discrete tasks in order to limit confusion, the tasks are interrelated. Moreover, the steps required to accomplish one of the identification tasks can be performed at the same time that steps for another of the identification tasks are being performed. The next chapter on planning the evaluation will address some of the specifics that will start the team on its journey toward a successful HR program evaluation.

## **Suggested Resources**

- Berk, R.A., & Rossi, P.H. (1990). *Thinking about program evaluation*. Thousand Oaks, CA: Sage.
- Edwards, J.E., Scott, J.C., & Raju, N. S. (Eds.). (2003). *The human resources program-evaluation handbook*. Thousand Oaks, CA: Sage.
- Joint Committee on Standards for Educational Evaluation. (1994). *The program evaluation standards: How to assess evaluations of educational programs* (2nd ed.). Thousand Oaks, CA: Sage.
- Love, A.J. (1991). *Internal evaluation: Building organizations from within*. Thousand Oaks, CA: Sage.
- Steinhaus, S.D., & Witt, L.A. (2003). Criteria for human resources program evaluation. In J.E. Edwards, J.C. Scott, N.S. Raju (Eds.), *The human resources program-evaluation handbook* (pp. 49–68). Thousand Oaks, CA: Sage.