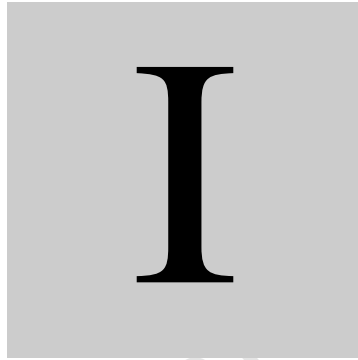


PART



**THE IBSTPI®
EVALUATOR
COMPETENCIES:
DEVELOPMENT,
INTERPRETATION,
AND APPLICATION**

CHAPTER

1

THE PRACTICE OF EVALUATION IN ORGANIZATIONS

This chapter will enable you to accomplish the following:

- Identify several major evaluation approaches used within organizational settings
- Describe some similarities among the approaches
- Describe some differences among the approaches

Various professionals and managers within organizations are increasingly called on to undertake evaluations of organizational programs and processes. This emphasis comes from a variety of sources. In some cases, it arises because of pressure from executives. For example, a study by the Conference Board (Gates, 2005) found that executives plan to make increasing use of human capital metrics. The explanation for this interest by executives involves the role that such measures and evaluation play in helping the organization achieve its strategic goals. In other cases, line managers, including those in human resources (HR), human resource development (HRD), and human performance technology (HPT), recognize the importance of measuring the effects

4 Evaluator Competencies

and impacts of their interventions. Such information can be used to determine what changes, if any, are needed in the interventions to meet their objectives. Thus, knowledge and skills in evaluation appear to be of increasing importance within organizations.

Given the importance of evaluations done within organizational contexts, this book provides some background on such work. More important, this book presents and discusses competencies needed by those undertaking evaluations within organizational settings, whether these are for-profit, nonprofit, governmental, or military organizations. Such evaluators are, however, focused on examining internal organizational programs, processes, and products. Thus, the competencies described in this book are not necessarily relevant to evaluators who are examining local, state, or national programs. So as an example, the national-level Department of Education may provide funding for programs aimed at reducing adult illiteracy, and there may be an evaluation being undertaken of that program. The competencies and standards that the evaluator should use for such an evaluation are typically described by some national evaluation association, such as the American Evaluation Association, the Australasian Evaluation Society, the Canadian Evaluation Society, and so forth. At the same time, however, that same Department of Education may provide internal staff training on conflict management and may ask a staff person or an external consultant to evaluate it. The competencies described in this book are relevant and applicable to such an evaluator. Furthermore, it should be recognized that although the example describes a governmental organization, such internally focused evaluations take place in a variety of organizations that may be small or large and that may provide products or services locally, nationally, or globally.

The purpose of this chapter is to provide a brief overview of the theory and research that exists concerning evaluations within organizations. It should be recognized that evaluation work focused within organizational settings has tended to follow a different evolution and history than that characterized by program evaluation. Russ-Eft and Preskill (2001) provided a comprehensive examination of the development of the field of evaluation and of the specialized area focused on evaluations within organizational settings. This chapter does not replicate the description of that history, but instead provides a brief overview focused on evaluations within organizations. Furthermore, it presents this overview following the recent suggestion by Wang and Spitzer (2005) that the field can be characterized by three types of approaches: (a) “practice-oriented

atheoretical,” (b) “process-driven operational,” and (c) “research-oriented, practice-based comprehensive” (p. 6).

PRACTICE-ORIENTED ATHEORETICAL APPROACHES

Much of the work on evaluation within organizational settings began with Donald Kirkpatrick, who focused on whether programs are achieving their objectives. He created a taxonomy that he called a “four-step approach” to evaluation and has more recently called “levels of evaluation” or the “four-level model.” His work was initially published in a series of articles in 1959 and 1960, and it was updated in 1994. Some other practice-oriented approaches, many of them variants of the Kirkpatrick approach, will also be mentioned.

Kirkpatrick’s Four-Level Evaluation Model

Kirkpatrick (1959a, b; 1960a, b) described the outcomes of training as focused on *reactions*, *learning*, *behavior*, and *results*, and he proposed that evaluations of such training should measure each of these outcomes. A level 1 evaluation gathers reactions to the training, a level 2 evaluation determines trainees’ learning, a level 3 evaluation measures the behavior of trainees (typically on the job), and a level 4 evaluation examines the business results from training.

Kirkpatrick’s taxonomy has had extensive application in training evaluations, and this primarily stems from the fact that it is easy to understand. In 1993, Kraiger, Ford, and Salas asserted that “Kirkpatrick’s recommendations continue to represent the state-of-the art training evaluation” (p. 311); however, they did suggest some improvements in the taxonomy. Later, Hilbert, Preskill, and Russ-Eft (1997) reviewed 57 journal articles in the training, performance, and psychology literatures that discussed or mentioned training evaluation models. Of those, 44 (or 77 percent) included Kirkpatrick’s model (either alone or in comparison with another model). A mere 13 articles discussed a model other than Kirkpatrick’s.

Although Kirkpatrick’s approach has been discussed in the literature, organizations have not tended to implement all four levels. Training interventions, in particular, are typically evaluated at the reaction and learning levels, with only some attention paid to the behavioral outcomes. For example, Taylor, Russ-Eft, and Chan (2005) undertook a meta-analysis of studies, both published and unpublished, that evaluated the effectiveness of behavior-modeling training. Of these

6 Evaluator Competencies

studies, 52 measured attitudes, 14 measured declarative knowledge, 32 measured procedural knowledge, 66 measured job behavior, and none measured business results. The reliance on reaction and learning measures may be due to the perceived difficulty and cost in measuring performance or behavior and organizational benefits. Some would consider such a result to be ironic because they would view improvements to the business as representing the business case for a learning or training investment.

Kirkpatrick's approach to evaluation has served as the basis for the development of other similar models for use in corporate, military, or small organizational settings. The remainder of this section describes some of these models, approaches, and taxonomies.

Navy Civilian Personnel Command Model

Similar to the Kirkpatrick approach, the Navy Civilian Personnel Command model examines knowledge and competencies gained during training (Erickson, 1990), but the measurement of those knowledge and competencies involves intensive testing. In this particular evaluation, trainees in staffing and placement experience intensive interviews three to six months after training. The interviews present real-life situations, and trainees must explain how to handle those situations to a subject-matter expert. Such an approach more closely approximates the transfer situation than does a paper-and-pencil test.

Training Effectiveness Evaluation System

The Training Effectiveness Evaluation System (Swanson & Sleezer, 1987) recommends measuring participants' and supervisors' satisfaction; trainees' knowledge and skills; and organizational, process, job, and financial performance. Four separate tools, then, are suggested: two satisfaction measures (participant and supervisor), a trainee learning measure, and a financial performance measure. Scores from each of the tools are then assessed before and after an intervention.

Hamblin's Five-Level Model

Hamblin (1974) proposed a five-level model similar to Kirkpatrick's, suggesting that evaluations measure reactions, learning, job behavior, and organizational impact (noneconomic outcomes of training). In addition, however, he indicated that there should be a level 5; this level should measure "ultimate value variables" or "human good" (economic outcomes).

Kaufman, Keller, Watkins Five-Level Model

Kaufman and Keller (1994) and Kaufman, Keller, and Watkins (1995) proposed a model that expanded on the ideas of both Kirkpatrick and Hamblin. They suggested the levels of Enabling and Reaction, Acquisition, Application, and Organizational Outputs—again, conceptually similar to Kirkpatrick. In addition, they recognized the societal impact and included Societal Outcomes as a fifth level. By adding it, they take into account the societal impact of training or of any HRD intervention. Such a model recognizes that organizations and the programs and processes within those organizations can affect clients and the larger society. According to Kaufman, Keller, and Watkins (1995), such Societal Outcomes or “megalevel” (p. 375) provides evidence of the ways in which the organization benefits the society.

Swanson and Holton's Results Model

Swanson and Holton (1999, p. 8) claimed that “assessment and evaluation are different. Assessment of results is a core organizational process. Evaluation is optional.” They then proceeded to describe approaches to measuring performance results in terms of systems outcomes and financial outcomes, learning results in terms of knowledge and expertise outcomes, and reaction results in terms of participant and stakeholder outcomes. They also detailed specific measurement approaches such as the critical outcome technique, auditing program practices and effectiveness, certification of core expertise, and assessing performance drivers.

PROCESS-DRIVEN OPERATIONAL APPROACHES

The Wang and Spitzer (2005) conceptualization of process-driven operational approaches tends to limit examination to issues related to return on investment (ROI). By taking a broader view, however, we can identify various other approaches that focus on processes, either within programs or within the evaluation.

Brinkerhoff's Stage Model

Brinkerhoff (1988, 1989) suggested a cyclical approach in which every phase of a program can be evaluated. His six-stage model begins with goal setting or needs analysis. In this stage, the evaluation involves identifying the training needs before designing a program. In the next stage, the evaluation examines the program's design. The third stage evaluates

8 Evaluator Competencies

the training program's operation or implementation, and the later stages may be considered similar to the outcomes in Kirkpatrick's approach. This stage model provides a type of formative evaluation in that the results can be used to aid in decision making and improvement through the design and implementation of the program.

Input, Process, Output Model

Bushnell (1990) described the Input, Process, Output Model as IBM's corporate education strategy for the year 2000 and, similar to Brinkerhoff's stage model, views the evaluation as a cyclical process. This model also appears related to the context, input, process, and product (or CIPP) evaluation suggested by Stufflebeam (1983, 2000). The Input, Process, Output model begins by examining the input factors that may affect a program's effectiveness, such as trainee qualifications, program design, instructor quality and qualifications, materials quality, facilities, and equipment. The process factors are then examined, and these include such variables as the planning, developing, and delivery of the training. The results can then be evaluated and divided or organized into the outputs and the outcomes. Outputs are considered the short-term results and include trainee reactions, knowledge and skill gains, and job performance improvement; outcomes, or long-term results, include what might be considered business results, such as profits, customer satisfaction, and productivity.

Stages of Transfer Model

Foxon's (1994) Stages of Transfer Model focuses on the transfer resulting from a training intervention, and it views transfer as a process rather than an outcome. The stages of transfer move from conscious intention to unconscious maintenance, with each stage affected by supporting or inhibiting factors.

Stage 1: Intention to transfer begins with the decision and the motivation to apply newly acquired knowledge and skills. The training environment, work environment, and organizational environment support or inhibit this intention.

Stage 2: Initiation occurs with the first attempt to apply new knowledge and skills at the job. The organizational climate, trainee characteristics, training design, and training delivery can support or inhibit this initiation.

Stage 3: Partial transfer takes place when some of the knowledge and skills are learned or are applied inconsistently. Skill mastery, the opportunity and motivation to utilize the learning, and the confidence to apply skills and knowledge support or inhibit this partial transfer.

Stage 4: Conscious maintenance occurs with the conscious application of what was learned in training, and this is supported or inhibited by the trainee's motivation and skills.

Stage 5: Unconscious maintenance occurs when the new knowledge and skill are integrated into the work routine.

Return on Investment

Literature describing the evaluation of business results, financial results, and ROI has become increasingly popular. Phillips and Phillips (2005, 2006) have written extensively about ROI and consider this to equate to Kirkpatrick's level 5 evaluation. At this level, "impact measures are converted to monetary values and compared with the fully loaded program costs" (Phillips & Phillips, 2006, p. 3). Evaluation at this level is typically represented by the benefit-cost ratio (BCR) or by ROI. The following provides the basic formula:

$$\text{BCR} = \text{Program Benefits} / \text{Program Costs}$$

$$\text{ROI} = [\text{Program Benefits} / \text{Program Costs}] \times 100$$

Despite the interest in ROI and cost-benefit kinds of outcomes, well-designed and well-documented ROI evaluation efforts are hard to find, particularly in relation to programs in which outcomes are focused on attitudes rather than knowledge or skill. It can even be difficult to measure the ROI of leadership development programs, which is where many executives are concerned about evaluation. Bartel (1997), however, identified some key attributes for determining ROI, specifically of training interventions. One key recommendation is that evaluators interested in financial benefits or ROI use the net present value or the internal rate-of-return method when determining the ROI of training. Both of these methods take into account the time value of money.

RESEARCH-ORIENTED, PRACTICE-BASED COMPREHENSIVE APPROACHES

Wang and Spitzer (2005) suggest that some of the more recent approaches tend to have a practice base but are also research and theory driven. A more appropriate characterization, however, might be that they are focused on systemic issues within the organizational setting. In this section, we review some of those approaches.

Systemic Model of Factors Predicting Employee Training Outcomes

Richey (1992) described a Systemic Model of Factors Predicting Employee Training Outcomes. The model focuses on factors affecting training outcomes, particularly the trainee characteristics and perceptions of the organization. Thus, it posits that the trainee attitudes are affected by such background characteristics as age, education, previous training, ability to learn, and motivation. Furthermore, these attitudes are also affected by the working conditions and management approach. Although instructional design and delivery may affect training outcomes, it is these trainee attitudes that have a direct effect on knowledge and behavior resulting from training.

Learning Outcomes Model

Kraiger, Ford, and Salas (1993) suggested that training evaluations should focus on the learning outcomes. Furthermore, they argued that training evaluation lacked theoretically based models. They used cognitive, social, and instructional psychology and human factors to determine the relevant outcomes. These were identified as cognitive, skill-based, and affective learning outcomes, and potential measures were suggested for each.

Training Efficiency and Effectiveness Model

Lincoln and Dunet (1995) recommended that evaluation should take place throughout the training process, with results continuously informing next steps. Their model consists of evaluation stages as analysis, development, delivery, and results. The approach suggests that the evaluator identify all stakeholders in the program and the evaluation. These views and information needs of these stakeholders need to be considered in the design, development, and implementation of the evaluation.

Brinkerhoff's Impact Map

Brinkerhoff and Gill (1994) suggested that too often the wrong people get sent to the training or that the right people attend but there are factors preventing their use of the training, such as poor program design, inadequate instructors, the lack of supervisory or peer support, a fear of failure, or a system that punishes the new behaviors. In such cases, the intervention can have little or no impact. As a result, Brinkerhoff and Gill introduced the notion of an *impact map*, which can be viewed as similar to the idea of evaluability assessment introduced by Joseph Wholey (1975, 1976, 1979). In this case, the evaluator creates a “map” showing the entire process from an input phase to the desired outcomes. This map can then help to identify both the process and the factors that can affect the outcomes.

Brinkerhoff's Success Case Evaluation Method

More recently, Brinkerhoff (2003, 2006) introduced what he calls “the success case method.” This method enables the evaluator to examine the ways in which the training is or is not aligned with the business strategy. It begins by recognizing that “programs are almost never completely successful such that 100% of the participants use learning on the job in a way that drives business results. Similarly, almost no program is ever a 100% failure such that no trainee ever uses anything for any worthwhile outcome” (Brinkerhoff, 2005, p. 92). Rather than gather superficial information regarding all trainees, the method involves examining the successful and the unsuccessful cases. This information can then be used to document the individual and business effects and can identify the factors that support or hinder those effects. It also provides impact data demonstrating the measurable impact of the intervention on the organization.

Holton's HRD Evaluation Research and Measurement Model

Holton's model (1996) argued that the Kirkpatrick model is really a taxonomy. As a more comprehensive model, he proposed that the three outcomes of training (learning, individual performance, and organizational results) are influenced by primary and secondary factors. More recently, Holton (2005) has elaborated on the original model. Although the outcomes remain as learning, individual performance, and organizational performance, the various factors influencing these involve ability,

12 Evaluator Competencies

environment, motivation, and secondary influences on each of these outcomes.

Preskill and Russ-Eft Systems Model

Preskill and Russ-Eft (2003, 2005) and Russ-Eft and Preskill (2001, 2005) recommended that evaluations within organizations use a systems model. Such a model recognizes not only the various factors affecting the individuals and the program or process but also the factors that influence the evaluation itself. Thus, the knowledge and competence of the evaluator and the procedures used in the evaluation can affect the outcomes and the findings as much as the design of the program, the motivation of the participants, and the support from peers and supervisors. In addition, the context affecting the organization, such as the competitive environment and the views of customers, may also affect the evaluation and its outcomes. Finally, such a systems model recognizes that these factors both influence the evaluation and the outcomes and are influenced by the evaluation and its outcomes.

CONCLUSIONS

A variety of approaches to evaluation exist for use by evaluators working within many different organizational settings, including small, medium, and large companies; local, state, and national government agencies; and local, regional, national, and international nongovernmental organizations. As more and more of these organizations undertake evaluation efforts, the various approaches to evaluation continue to evolve. Furthermore, this evolution now suggests that organizations and the programs within them are complex systems that are difficult both to examine and to evaluate. Thus, there is a need to determine the competencies required by those who undertake evaluations within such contexts.

QUESTIONS FOR CONSIDERATION

What are two different evaluation approaches that you might use, and on what basis would you make a choice?

Choose two of the evaluation approaches that you could use in your organization and identify the benefits and drawbacks of each approach.

Describe how you would combine two or more of these approaches for use in an organization.