EVALUATION THAT GOES THE DISTANCE

BY PAUL R. BERNTHAL

Four-level training evaluations often stop short of reaching meaningful, long-term results. Here's a road map for adding on measures that can go the distance.

In 1959, Donald Kirkpatrick published a paper that classified training outcomes into four levels: reaction, learning, behavior, and results. Kirkpatrick's classic model has weathered well. But it has also limited our thinking regarding evaluation and possibly hindered our ability to conduct meaningful evaluations.

Too often, trainers jump feet first into using the model without taking the time to assess their needs and resources, or to determine how they'll apply the results. When they regard the four-level approach as a universal framework for all evaluations, they tend not to examine whether the approach itself is shaping their questions and their results.
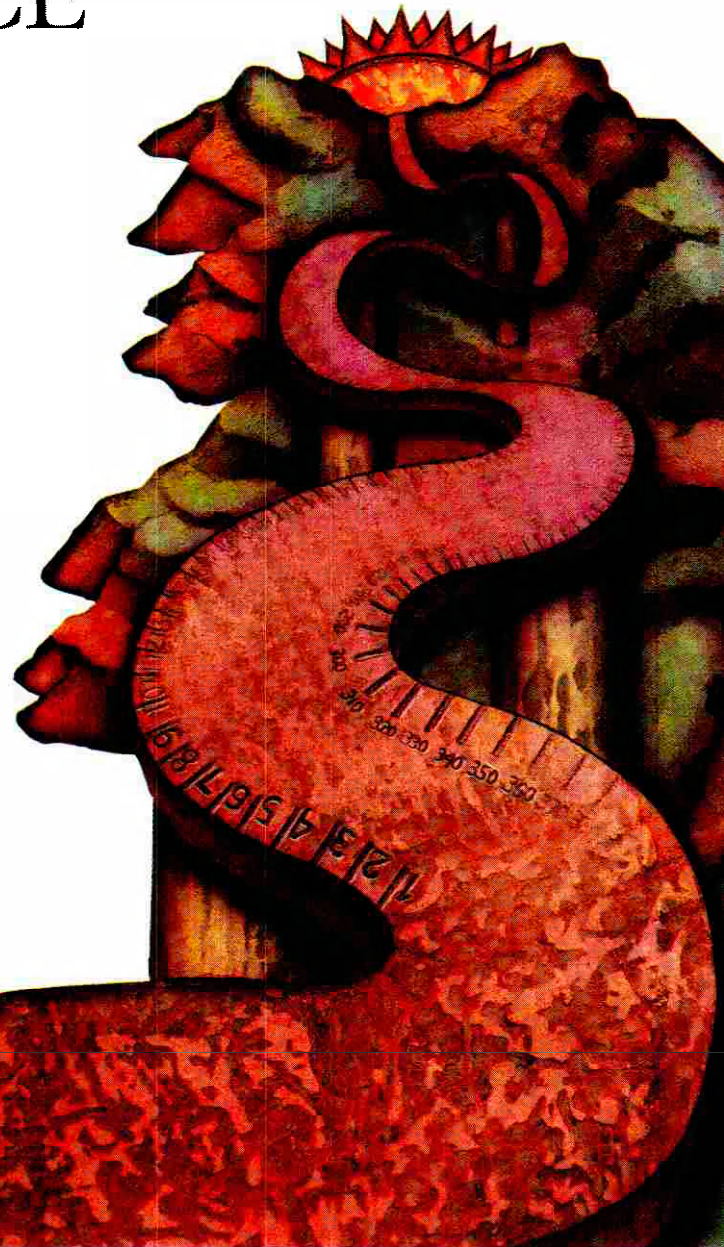
Other options

The simplicity and common sense of Kirkpatrick's model implies that conducting an evaluation is a standardized, prepackaged process. But other options are not spelled out in the model.

First, it's important to reexamine some faulty assumptions about four-level evaluations and evaluations in general.

Assumption: Evaluations are definitive. Most training evaluations are based on the philosophy that a single study can answer all questions about the effect of a training effort. Most evaluators aren't prepared for ambiguous findings. But in any evaluation, the degree of certainty regarding the results depends on such variables as the rigor (reliability) of the design, the measures, and the sampling strategy.

Most importantly, credibility depends on whether an evaluation's findings can be replicated. If only one evaluation is conducted, it probably can't stand on its own merit. Other studies are likely to point out flaws or alternative explanations for the findings.

Assumption: Evaluation equals effectiveness. Evaluation focuses on the learning aspect of training. It answers the question, "Have the requisite skills and knowledge appeared as a result of training?"—a level-2 evaluation in Kirkpatrick's model. An evaluation can become problematic when it also tries to measure effectiveness. Effectiveness focuses on whether the training has produced the intended outcomes (levels 3 and 4). To answer the effectiveness question, the evaluator must measure several organizational, individual, and training-related variables.

Evaluation and effectiveness are linked. But they shouldn't necessarily be arranged on a continuum, as they are in Kirkpatrick's model.

Assumption: Trainers are accountable for effectiveness. Many trainers who conduct evaluations don't have the skills, time, or resources to do an in-depth study. They may not be knowledgeable about the training topic. In such cases, does it make sense for trainers to be held responsible for the success of all training programs, especially those instituted by senior managers? Still, trainers often have the most to lose when results aren't positive.

Assumption: Level-4 evaluation is superior. Some studies show that Kirkpatrick's levels 1 through 4 may be correlated. But they measure different things.

So, why is level 4 often described as a higher level of evaluation? In fact, many level-4 evaluations are conducted only because they're viewed as the toughest assessment of training, even when the measures used have no real link to training. One shouldn't choose a level-4 evaluation and then try to tie it back to the training, just because the training involves some level-4 variables.

Suppose that the training is on interpersonal skills. Such variables as operational costs and equipment downtime may have some relation to employees' interpersonal skills. But those variables wouldn't be the best measures to use in the training evaluation.

Typically, trainers believe that level 4 is the pinnacle of training evaluation. But each level can provide equally valuable information, depending on the type of trainees being evaluated. Level 1 or 2 outcomes can provide some of the most useful information because those outcomes are often the easiest to measure and change.

Assumption: You just have to measure it. Many measures used to assess training are inappropriate and not sensitive enough to detect changes in trainees' behaviors. It's also difficult to know what questions to ask, how to phrase them, and to whom to direct them. And all measurement methods aren't equally reliable or valid.

Building onto the model

Instead of choosing a particular level and jumping into the evaluation process, you may want to make these add-ons to Kirkpatrick's model:

Consider the context. Understand how the training fits within the organization's operations and culture. Many things in addition to training can affect the work environment. In evaluating level-3 or level-4 changes, it's important also to measure contextual variables. For example, the lack of management support can undermine even the most effectively designed and delivered training program.

Establish a link. Draw a cause-and-effect path between training interventions and such outcomes as job behaviors and productivity. The path can show where and in what ways the training results in measurable changes.

Make appropriate choices. It's important to choose an appropriate evaluation design and appropriate measurements. Ask what you really need to know. Let the answer determine your approach. Choose one level at a time to evaluate specific results.

An appropriate initial evaluation might be a post-training assessment on changes in trainees' knowledge, using a paper-and-pencil test. Another approach would be to use a multi-rater tool to compare trainees to a control group.

Inventory your resources. No matter what level approach you use, make sure that you're realistic about what you can accomplish. Consider the costs, the amount of time you can spend, trainees' downtime (to fill out questionnaires or other evaluation instruments), and the expectations of your customers (trainees, senior managers, and others).

Set goals and long-range plans. Don't just do an evaluation; establish an evaluation program. Many organizations and their training programs grow and change in ways that make most evaluations obsolete in a few years. Establish a program that includes multiple evaluations at various levels on the effect of each training effort.

For example, in *In Action: Measuring Return on Investment* (American Society for Training and Development,

1994), editor Jack Phillips recommends evaluating different percentages of programs at the four levels. He says you could evaluate 100 percent of all programs at level 1, 70 percent at level 2, 50 percent at level 3, and 10 percent at level 4.

The training-impact tree

The first task in setting up a long-range evaluation program is to create a training-impact tree. It will help identify the variables that could affect a training intervention and help establish links between the training and organizational values and practices. (See the box "A Training-Impact Tree.")

But before you create the training-impact tree, establish your team. Assemble a group of people with diverse perspectives on the effects of training on the organization. Team members should be knowledgeable about the organization's values and practices, the scope and appropriateness of the training, and the factors that could affect training transfer. The team should spend at least one entire day developing the tree.

Step 1: Identify the organization's values and practices. Many organizations publish a list of values. But they don't always practice them. The team should make a list of the organization's main values and associated practices. An example of an organizational value is teamwork. It is manifested in such practices as formally recognizing team efforts, linking individual goals to group goals, and creating an environment of open communication throughout the organization.

Step 2: Identify skills, knowledge, and attitudes. Once the organization's values are linked to practices, it's easier to identify the type of training that will enable employees to perform effectively under current conditions. To that end, the team should tie each practice to a list of skills, points of knowledge, and feelings that people can be trained in and about.

University of Colorado professor Kurt Kraiger has written that a training evaluation should focus on three areas of learning: skills (technical and motor), "cognitions" (knowledge and thoughts), and feelings (attitudes and emotions). In other words, training affects what you do, how you think, and how you feel.

For example, if the team's goal is to achieve a free exchange of information within the organization, team members should identify the skills, cognitions, and feelings associated with open communication. A skill would be knowing how and when to share ideas without being asked. A cognition would be understanding how the individual members of a group can affect the group as a whole. A feeling would be team members' concern about the success of the team.

Step 3: Define the scope and purpose of the evaluation. In addition to helping identify how training fits within the organization, the training-impact tree also can help generate a list of questions to include on a training evaluation.

An evaluation should measure more than reaction, learning, behavior, and results. Those levels focus mostly on outcomes; they don't take into account the process leading to the results. For example, one can evaluate behavior changes without recognizing that they might depend on people's motivation, the degree of managers' support after training is completed, or the extent to which the training was appropriate for meeting needs.

Here are several areas beyond the four-level scope that can be evaluated:
▶ the quality, delivery, or retention of the training

A Training-Impact Tree

Here's an example of a training-impact tree. It can help identify the variables in an organization that might affect training outcomes.

Barriers to Training

▶ Organizational values aren't clearly communicated.
▶ Senior managers send mixed messages.
▶ Achieving goals requires competitive behaviors from employees.
▶ Managers aren't held accountable for their teams' success.
▶ Team members work on different floors or in different buildings.

Organizational Values

Example: teamwork
Supporting practices include the following:
▶ recognizing team efforts
▶ participating in cross-functional teams
▶ linking individual goals to group goals
▶ sharing responsibility and accountability for team outputs
▶ exchanging information openly and frequently
▶ establishing partnerships with other departments and teams.

Factors in Facilitating Training

▶ The organization has a good relationship with its unions.
▶ The organization has a stable profit margin.
▶ The organization's values are logical.
▶ The job descriptions are clear.
▶ The selection system includes effective teamwork as a criterion.
▶ Managers serve as positive role models.
▶ The performance-management system rewards team accomplishments.

Training

Cognitions (knowledge and thoughts)
▶ the role of team leaders
▶ the effect of individual members on a group as a whole
▶ the meaning of empowerment
▶ the way teams and groups work together
▶ the importance of sharing activities and tasks among group members
▶ the importance of diversity.

Skills
▶ expresses thoughts clearly
▶ listens and responds with empathy
▶ seeks input from others
▶ offers help when others need it
▶ makes sure all team members are heard and understood
▶ shares ideas without being asked
▶ regularly recognizes individual efforts in support of team goals.

Feelings
▶ the motivation to grow and develop
▶ a sense of team spirit
▶ self-efficacy
▶ a concern for group cohesion.

- how well the training cut deficiencies in a particular work group
- the usefulness of parallel training for managers and their staffs
- variables in the work environment that discourage or facilitate the effect of training.

Most evaluations can benefit from measuring organizational context. The training-impact tree can show such context by listing the barriers to training and the factors that facilitate training next to their associated values and practices. If the team isn't sure what contextual variables to consider, it can seek direction from trainees or from focus groups made up of senior managers, frontline leaders, customers, or people from all three groups.

In the article, "Individual and Situational Influences on the Development of Self-Efficacy: Implications for Training Effectiveness (*Personnel Psychology,* spring 1993), authors John Mathieu, Jennifer Martineau, and Scott Tannenbaum say that training doesn't occur in isolation from employees' job responsibilities and personal lives. Just providing time for employees to receive training doesn't ensure training effectiveness.

Organizations that use such standard reaction measures as smile-sheet responses to revise a program may be making fruitless "improvements." Trainees' negative reactions to the training may have nothing to do with the training itself. They may resent the time it takes or the fact that it is mandated. They may be distracted by personal problems. Or, they may not be interested in the topic.

Step 4: Identify data sources. The quality of the evaluation data depends heavily on the source. For example, self-assessment is rarely the best way to determine whether a person's behavior actually changed as a result of training. The criteria for choosing the best data sources include a source's objectivity, accessibility, and reliability. Data sources should be unbiased, should provide understandable information, should be easy to access, and should produce information that's immune to irrelevant influences.

It's best to collect data from a variety of sources. People are more likely to believe similar findings from different sources. Even disagreements among sources can provide valuable information by offering different perspectives.

Step 5: Choose the best method for collecting data. One of the most difficult aspects of developing an evaluation is selecting and implementing an appropriate design. There are books that can help. An evaluation design can use almost any traditional research method. But the choices may be limited due to practical considerations.

Ask yourself these questions:
- How frequently do you want to collect data? It's best to survey trainees immediately after training is completed. But the findings will be more conclusive if you also conduct a pretraining assessment and a long-term follow-up. It's fairly easy to detect changes measured over time.
- Would data from a control group strengthen the findings? If trainees show improvement over time, it might be due to factors other than the training. A control group (people who did not receive the training) can provide a basis for comparison.
- How many people should I collect data from? There are no rules. The criteria for determining the appropriate sample size include randomness, representativeness, the size of the trainee group, and the desired statistical results.
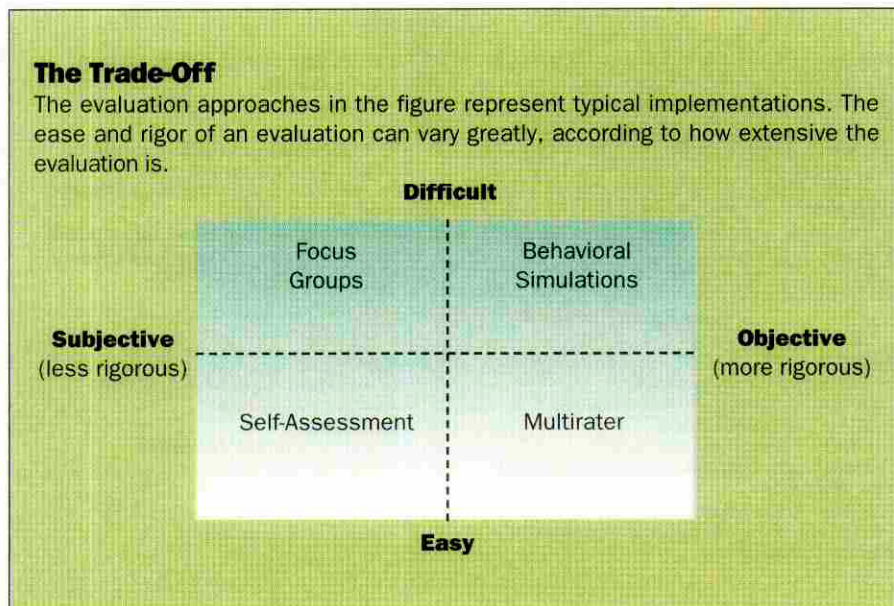
Ideally, the sample should be randomly drawn from the population being evaluated. But training rarely occurs randomly across an organization. Focusing on specific departments is all right, as long as they are similar to other departments in the organization.

If the population is fairly homogenous, fewer people will be needed in the sample. But the sample should always be representative of the target population. Every classification or demographic in the population will require more people in the sample.

Most statistical tests require at least 30 participants in each group being evaluated. For example, a training evaluation with a control group should involve 60 participants—30 in the control group and 30 trainees. Over time, there's usually a high rate of dropouts among participants. So for the benefit of follow-up assessments, it's best to pad the sample from the start.

Step 6: Select the best measurement approach. Typically, the easier the measurement, the less objective it is. To increase objectivity, use different measurement methods in the same evaluation or conduct several evaluations, each using different approaches—including self-assessment, multirater assessments (supervisors, peers, and subordinates rate trainees' performance), focus groups, and behavioral simulations (trainees use role play to practice new skills).

The assessment-center approach can yield valuable data, though it can also be difficult and time-consuming to implement. Alone, each of these

The Trade-Off
The evaluation approaches in the figure represent typical implementations. The ease and rigor of an evaluation can vary greatly, according to how extensive the evaluation is.

Difficult

| Focus Groups | Behavioral Simulations |

Subjective (less rigorous) — Objective (more rigorous)

| Self-Assessment | Multirater |

Easy

evaluation methods presents unique advantages and problems that might affect your conclusions about an evaluation. Combined, they represent a diverse and powerful approach for painting a complete picture.

Step 7: Gather and inventory your resources. Identify the people who will help conduct the evaluation. What are their skills? Which parts of the evaluation should they be responsible for? Do they have enough power and influence in the organization to implement the evaluation results?

The hardest part of establishing a long-range evaluation program can be gaining buy-in from people across the organization. Expect to do some internal selling of the evaluation. Try to create partnerships with your internal clients. You might want to develop a matrix listing the people who will serve as the doers, approvers, and reviewers. The matrix can show the points in the evaluation at which each person will be involved.

If you don't have the in-house resources to conduct the type of evaluation you need, you can hire an external consultant. Though many aspects of an evaluation can be conducted without outside help, more complex ones tend to require experienced evaluators. Be sure that the consultant understands the evaluation. Also make sure that he or she thinks in broader terms than just a four-level approach.

Many consultants charge $1,500 to $2,500 a day to conduct an evaluation. You can cut costs by administering the evaluation yourself. But you should expect to pay the consultant for several days of analysis and report generation. Typically, expect the consultant to spend four to five days planning, designing, and creating materials.

Training evaluations aren't one-time events. It's important to develop a schedule for periodic assessments. As you collect data over time, look for trends. Some evaluations may be "quick and dirty"; others may be more extensive. Remember: No one evaluation stands alone; almost all evaluations yield valuable information.
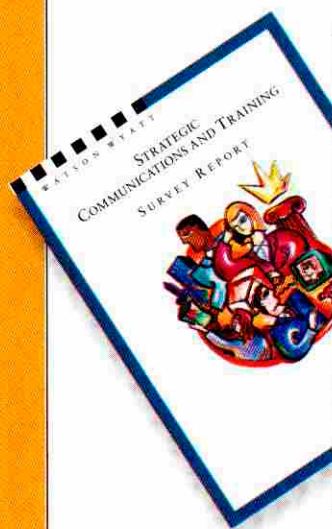
Once you've conducted several evaluations using various methods, you can fine-tune your approach and do fewer evaluations. The results can

lead the way toward the desired final destination: a strong training program. Let the program evaluations show where you've been and where you need to go. ■

___

Paul Bernthal is a research consultant with Development Dimensions International, 1225 Washington Pike, Bridgeville, PA 15017-2838. Phone