Very often, evaluation material is not used to full advantage—or at all.

Making Evaluation More Useful

BY ROBERT O. BRINKERHOFF

This article reflects the work and experience of the Evaluation Training Consortium (ETC) project. Since 1972, the ETC has been funded by the U.S. Office of Special Education (OSE); it aims to improve the evaluation of OSE supported personnel training programs. ETC has developed and disseminated techniques for evaluating training in the educational arena to thousands of professionals. In this article, Dr. Brinkerhoff, director of the ETC project, summarizes major concepts and approaches from this work, showing how they relate to evaluation in the HRD context. All have a common purpose: to make evaluation as potent and useful as possible in making training and development activities more effective.

One reason to evaluate HRD efforts is because HRD must account for itself and must be more productive and efficient. It must prove its utility, benefits and worth. Likewise, evaluation consumes valuable time and resources. Shouldn't it also have the same arguments for effectiveness, utility, benefits and worth?

Evaluation is defensible only to the extent that it can make current or future HRD efforts better. We must strive to make evaluation more useful.

HRD literature is full of excellent and thoughtful evaluation writing (Kirkpatrick 1975, Peterson 1979, among others). Much of the work and writing is aimed at *methods* of evaluation; how to make it more objective, accurate, precise and scientific. With some notable exceptions (Patton 1979, Alkin, et al., 1979), relatively little work has been devoted to making evaluation more useful and worthwhile. As a result, it is viewed as a problem instead of a *solution* to how training can be more efficient and effective.

Evaluation is an important part of any training and development effort. It is more than an assessment of outcomes or effects. Evaluation is systematic inquiry into training contexts, needs, plans, operation and effects. It should help collect information to decide what's needed, what's working and how to improve it and what's happened as a result.

Improving Evaluation

Given this broadened definition of evaluation, and given it would help improve HRD, here are guidelines for improving evaluation. These are not independent guidelines; they are interrelated by the reasons to conduct evaluation in the first place.

Evaluate the right things. Years ago, training evaluation was characterized by collecting participants' reactions (in rating form) to how the training was conducted, whether they liked it and so forth. Later, we were told to assess outcomes, to specify objectives quantitatively and to measure them. Most recently, we hear that examining immediate outcomes isn't sufficient. A good evaluation should get beyond the training and should look at the results: organizational payoff.

There is a new trend, however: It has become "right" to evaluate some things and "wrong" to evaluate others. This is unfortunate, for it increases the tendency to be trapped in self-serving evaluation. The rightness or wrongness of what is evaluated should be determined by the purpose for conducting the evaluation. Thus, if the purpose of an evaluation is to determine whether training participants "like," or are responsive to a certain training method, evaluating participant perceptions is more important than assessing on-the-job performance changes.

Judy Springer in her contribution to *ASTD Research Series No. 3* (Peterson, 1979) notes that many evaluations have measured the "wrong" things, because evaluation purposes were not closely allied with training purposes. If, for example, the purpose of a management training effort was to identify persons for promotion, an evaluation study examining increases in trainee skills as a group index, or reporting effects on the typical attendee, would be improperly focused.

To collect information on the "right" things, evaluation must first decide: Why evaluate? Once defensible purposes are defined, the search for what to evaluate will be easier.

Consider varied purposes for evaluation; evaluate to meet them. Evaluation often gets done for its own sake, as a categorical imperative ("of course we evaluate our training!"). Self-serving evaluation is likely, however, to accomplish very little. Above all, evaluation should be a useful activity, responsive to the needs and/or interests of particular audiences. It has many potential purposes:

- To determine the training "payoff," its impact on the organization;
- To make training more efficient;
- To be more accountable for the training function;
- To publicize the benefits of training;
- To determine the cost/benefit of training to the organization;
- To compare different training approaches;
- To determine whether training was delivered as planned;
- To find out which parts of the training effort worked best;
- To identify training problems and possible solutions;
- To determine needs for more or different training;
- To gain support for particular training methods or sessions;
- To determine whether a planned training approach will work.

No single approach to evaluation could serve all of these purposes. An evaluation focusing on pre- and post-changes in trainees, for example, would not help to identify problems encountered in the training delivery. An evaluation seeking to determine whether a training method had met its immediate intended objectives (e.g., a shift in knowledge) would not yield information as to the worth of the training in terms of organizational benefits.

There are many potentially effective evaluation models, approaches and ideas. The trick is to use or adopt one (or more) to meet a useful purpose.

Make sure evaluation purposes are clear and responsive to specific audiences. A useful way to consider an evaluation purpose is in regard to functions and audiences. Consider a purpose to be *someone* doing *something* with information from the evaluation. Purposes could be:

- To provide *HRD staff* with information about training design defects;
- To provide evidence to *management* that the workshop is effective;
- To give training information to the *company work-force;*
- To provide information to *workshop leaders* on their personal effectiveness.

Deciding what information should

be provided to whom helps assure that the evaluation will *have* a purpose. It also helps move along the evaluation planning process.

Audiences for the evaluation should help specify its goals. This will help assure that purposes will be responsive to needs. It also increases likelihood that gathered information will be used.

Broaden the focus of evaluation inquiry; relate it to reality. We must realize that training programs, like any other human activity, are fallible. Training designs are based on imperfect knowledge of learning theory, motivation and need. Educators deliver training in an environment where we exert only partial control. To approach evaluation as though training was a tightly controlled experiment (using a pre-post evaluation design) grossly limits the opportunity for evaluation to contribute to greater knowledge and development.

Evaluation should not limit itself to the intended outcomes of training. To be sure, evaluation should determine and account for training impact. It can also help to ensure and enhance training effectiveness by facilitating the growth and development of training efforts.

The focus for evaluation can be related to any or all of three major stages of HRD programming: *planning, delivering* and *recycling.* Evaluation related to planning can help determine:

- What are needs for training?
- What are the relative merits of alternative training strategies or programs?
- What are the strengths and weaknesses of past training approaches?
- What is the current status (e.g., perceptions, attitudes, knowledge levels, etc.) of potential trainees?
- How adequate is a given training plan; is it feasible?
- What is the current administrative support climate for HRD; how should training efforts be modified to help ensure success?

Evaluation related to delivery can help determine:

- Has the training plan been installed as it was designed?
- What problems emerge as training progresses?
- Is the training delivered consistently across several settings or instances?
- Is the training working as planned; are participants making use of materials and engaging in activities?
- What are the nature and likely causes of defects in the training?
- Has the training accomplished its intended objectives?
- What *other* (unplanned, suspected, etc.) results is the training producing?
- What were training costs?

Evaluation related to recycling can help determine:

- Have trainees' needs been met?
- Can trainees make use of training on-the-job?
- What use of training do trainees make on the job?
- Are reasons/problems for the training changing? getting better? getting worse?
- What new training needs are emerging?
- Is more, less or different training needed?
- Are the training results worth the costs?

Consider a broader range of information collection methods. We should avoid thinking that some kinds of information are for evaluation and other kinds are not. Any information that represents training or someone's reactions to it is potentially useful evaluation information.

The key to good information is usefulness in respect to a given evaluation purpose. The "goodness" of the information-collection method lies in whether it can *economically* collect *credible, convincing* and *timely* information relative to an evaluation purpose. The "test" of any information-collection procedure is to ask: "If we had the information this procedure will collect, can we make good use of it for this evaluation purpose?" Some guidelines here are:

- *Be creative.* Use a variety of information-collection procedures. Some options are: interviews, observations of participants, self-reports, tests, analysis of records and documents, analysis of work samples, surveys, panel reviews, "public" hearings, monitoring rates, trends, patterns, simulations, self-ratings, ratings by others, logs and diaries, case-studies.

• *Consider the argumentative and persuasive power of the information to be collected.* Consider this example. Assume we want to know whether participants valued handouts in a workshop. One piece of data reports 4.2 as the average rating on a 5-point scale (5 = high). Another piece of data is that I notice many (12 of 30) handouts were thrown in the trash. Which information is more credible and persuasive?

• *"Triangulate" information-collection procedures.* No single information source can satisfy an evaluation purpose. Several sources of information will give a richer understanding of the issue. Thus, if we want to know what use workshop "graduates" make of their acquired skills, we get a fuller picture from: 1) surveying them; 2) analyzing samples of their work; and 3) interviewing their supervisors.

• *Follow-up evaluation information by getting more information.* Rarely, if ever, will evaluation information be conclusive and certain. More often, it is equivocal and provocative. We make greater use of findings when we pursue further. Thus, initial findings are clues, which generate hunches, stimulate more clue-finding, etc.

• *Adapt information-collection procedures to the needs, value and belief systems of audiences.* Above all, evaluation information should be *actionable*—used by its intended recipients. Find out what evaluation information would be most credible to each audience. A summary of expert opinion on the benefits of a workshop may, for example, be less credible to line-workers than a summary of peer opinions.

Keep expectations for evaluation reasonable. In recent years, many evaluation approaches and models (such as Stufflebeam's CIPP) have been defined as "decision-making" models, intended to provide information upon which decisions can be based. This is certainly a useful way to consider evaluation, but at least three major problems stem from it:

• *Evaluation will never provide all necessary information to base a decision.* Likewise, there will always be additional bases for decision-making other than those within the scope of evaluation (political forces, personal values of the decision-maker, etc.). Evaluation can, and should, add information to the decision-making base.

• *Other functions exist for evaluation in addition to decision-making.* Evaluation that does not yield or influence a decision is not necessarily useless. The evaluator who expects every decision to employ all data is sure to be disappointed. Evaluation information is good for other purposes: to inform an audience; to gather support for decisions; to encourage staff to make decisions; and to refine and redirect further evaluation activity.

Evaluation to serve only decision-making may miss opportunities for greater impact. Evaluation bound to a deductive set of decision-questions (e.g., did the training achieve objectives?) is likely to overlook and ignore other questions (e.g., what else has the training achieved?). It may become too rigid to respond to the emerging problems and changes encountered in a training program.

• *Consumers of evaluation information are often misled as to the certainty of evaluation findings.* Evaluation results are always equivocal and subject to varying interpretation. We can expect them to be relevant, specific, persuasive and credible.

Ernest House, in a well-conceived essay (1977), discusses the argumentative logic of evaluation. He points out that when evaluation is expected to adhere to principles of sound argumentation (coherence, relevance, credibility) versus the principles of rigid scientific inquiry, it becomes less limiting, broader in scope and function and more useful.

• *Don't just report evaluation findings.* Educate consumers to use them. Evaluation is most successful when evaluators and evaluation consumers discuss results. When findings stimulate discussion about their meaning and applicability, they will most likely be used. Reporting functions might include meetings, discussions, presentations, etc.

There is another suggestion relative to evaluation reporting: *education*. Sometimes, an evaluation will collect certain information, often statistical data, with the warning, "I know this doesn't really mean anything, but 'they' wanted hard data, so here it is!" When this happens, evaluation should resist appeasing someone's appetite for so-called "hard-data." This information may have little or no use. To not resist is to run the greater, long-term risk of loss of credibility. The proper solution is to recognize the legitimacy of the need and patiently explain why the called-for data won't help. Then, seek and provide information that *will* meet the need.

Increasing Utility

Very often, evaluation information is not used to full advantage, or at all. This results from several factors: how evaluation is done, the purposes it is addressed to and the expectations held for it. By concentrating too much on certain kinds of training variables (such as outcomes), and by limiting its methods to supposed experimental/scientific procedures, evaluation will not be helpful. Evaluators need to broaden the scope of application and methods, and more closely involve their audiences in evaluation design, conduct and application. If they do, evaluation usefulness will increase.

Bibliography

Alkin, Marvin C., R. Daillak and P. White. *Using Evaluations*. Beverly Hills, Sage, 1979.

House, Ernest. *The Logic of Evaluative Argument*. Center for the Study of Evaluation, UCLA, 1977.

Kirkpatrick, Donald L., ed. *Evaluating Training Programs*. Madison, ASTD, 1975.

Peterson, Richard O., ed. *Determining the Payoff of Management Training (Research Series #3)*. Madison, ASTD, 1978.

Robert O. Brinkerhoff is associate professor and director of the Evaluation Consortium, Western Michigan University, Kalamazoo, Mich.