

Evaluating Public Sector Training Programs

This article discusses the importance of training program evaluation in increasing the effectiveness of agency training efforts. The article describes a framework available to the personnelist to plan, design, conduct and implement training program evaluations that more clearly identify the effectiveness of training programs.

By
Ronald R. Sims

The concept of training evaluation has received widespread recognition as beneficial, but the practice of evaluation has lagged behind (Bell and Kerr, 1987). Few reports of actual program evaluation have been published; compared to the number of training programs, few evaluations have been conducted. Lack of training program evaluation is even more evident in the public sector, and is possibly the least developed aspect of the training process in public agencies. Yet it is arguably the most important. Despite the hundreds of articles, books, and seminars devoted annually to the topic, training evaluation remains largely misunderstood, neglected, or misused. Too often, training is done without any thought of measuring and evaluating how well the training worked. Yet, the training process is not complete until and unless evaluation has taken place, for it is evaluation which informs training and gives it meaning.

If this is the case, why then is the evaluation of training in public agencies so underdeveloped? There are several reasons for this. Primary among them, as Brethower and Rummeler (1979) suggest, is that evaluation of training means different things to different people. There does not seem to be a consistent definition of what training evaluation is among personnelists. A second, more fundamental, reason, which persists even when there is consensus on the definition of training evaluation is the fact that the serious evaluation of training in public agencies is a difficult, tedious, time-consuming task which most personnelists would prefer not to engage in. A third reason for not conducting evaluations is that training administrators responsible for training simply tend to assume that training will work. A final reason is that a personnelist who champions a training program may feel threatened by the prospect of an objective evaluation of the program's effectiveness.

Recent research has shown that more than one-third of the members of the American Society of Training and Development responding to a survey reported that evaluation was the most difficult aspect of their jobs (Galagan, 1983). In another survey, Lusterman (1985) found that over two-fifths of the responding organizations reported significant changes when training effectiveness was evaluated. An even more recent survey of training directors found that 90% claimed that even though they believed

Ronald R. Sims is Associate Professor of Business Administration at the College of William and Mary. His research interests are primarily in the area of training and development and the application of experimental learning to school and work settings.

the evaluation of training to be important, they did not conduct evaluations because their organizations did not require them (Bell and Kerr, 1987). Unfortunately, because of the perceived difficulties and the inability to identify specific outcomes as a result of training, post-training evaluation and follow-up often are ignored (Rao and Abraham, 1986).

To demonstrate training's importance, personnelists must not only present excellent programs but also must demonstrate that their programs get results: improved job performance, more efficient use of resources, and satisfactory returns on the training dollars invested. It is the contention of this paper that personnelists can prove the *value of training* when they systematically plan and implement program evaluation. Without a systematic plan a lack of emphasis on the determination of the worth of a training program can mean danger in training efforts in the long run. That is, failure to systematically evaluate training leaves open the potential for growth in training without accountability. This may lead to the continuation or even proliferation of ineffective programs or, in times of budget cutbacks, the perception by top administrators that training programs are superfluous and should be cut. Again, if personnelists are to eliminate the current roller coaster approach to agency support for training, systematic evaluation must become a part of every training program—whether or not key agency stakeholders require it.

Training Evaluation: A Definition

Evaluation of training compares the post-training results to the objectives expected by administrators, trainers, and trainees (Mathis and Jackson, 1991). Hamblin (1970) defines evaluation of training as "any attempt to obtain information (feedback) on the effects of a training program, and to assess the value of the training in the light of that information". Put another way, evaluation tries to answer the question: did training work, and if not, why not? Hamblin further contends that the primary purpose of evaluation is to improve training by discovering which training processes are successful in achieving their objectives.

Similarly, Swierczek and Carmichael (1985) identify the goals of evaluation as:

1. To improve the training program.
2. To provide feedback to program planners, managers, and participants.
3. To assess employee skill levels.

In addition, personnelists evaluate training for professional reasons because evaluation is one way in which they can assess their own effectiveness as trainers. From an administrative standpoint, personnelists evaluate in order to justify the time and money spent on training. The evaluation of training is, therefore, an integral part of the personnelist's "bag of tricks".

Given the diversity of agency training needs, there is no single method most appropriate for evaluating training efforts. The circumstances dictating the need for training, different methods used in training, and the different purposes for evaluation all make plain the need for multiple approaches in training evaluation. Regardless of need, method, or purpose the personnelist must carry out a systematical identification and organization of important factors related to planning and executing the training evaluation process.

Having provided a definition of training evaluation, which will be revisited later in this paper, it is important to discuss in more detail training evaluation objectives benefits.

Training Evaluation Objectives and Benefits

The primary and overriding objectives of the evaluation of agency training programs should be to collect data that will serve as a valid basis for improving the training system and maintaining quality control over its components. It must be emphasized that *all* components of the system and their interaction are the objects of scrutiny and that personnelists should ensure that training programs are designed with *a priori* consideration given to evaluation. That is, public sector trainers should be committed to evaluating the effectiveness of their programs. Several potential benefits result from evaluating agency training programs:

1. Improved accountability and cost effectiveness for training programs which might result in an increase in resources;
2. Improved effectiveness (Are programs producing the results which they were intended?);
3. Improved efficiency (Are the programs producing the results for which they were intended with a minimum waste of resources?);
4. Greater credibility for the personnelists to include information on how to do a better job now or in future programs or to redesign current or future programs;

5. Stronger commitment to and understanding of training by key administrators so they can make up for deficiencies and confirm/disconfirm subjective feelings about the quality of agency training;
6. Formal corrective feedback system for developing strengths and weaknesses of training participants. Trainees that understand the experience more fully and are more committed to the program;
7. Managers better able to determine whether to send potential recruits to future training programs;
8. Quantifiable data for agency researchers and training program developers interested in training research;
9. Increased visibility and influence for public sector training program sponsors;
10. Increased knowledge and expertise in the development and implementation of training programs that produce the results for which they were intended.

This is not an exhaustive list of the objectives and benefits of a training program evaluation, however, personnelists who are responsible for training must continually ask themselves what are the objectives of evaluation; and what do they want to gain by conducting an evaluation?

A priori consideration of evaluation gives the personnelist at least five important advantages:

1. The ability to identify relevant audiences, interested in training evaluation, early in the process to ensure that evaluation feedback addresses their interests and information needs.
2. The development of an evaluation process that compliments the training program. Evaluative methods can be carefully incorporated to minimize any disruptive effects on the training program.
3. The ability to construct a research design that allows for valid conclusions about the program's effectiveness. This includes finding appropriate pre-measures, selecting appropriate groups or individuals to train, identifying comparison groups, and isolating extraneous variables prior to beginning training.
4. The ability to delineate material, data, and human resource requirements for evaluation and incorporating these as part of the

training program, not simply as an appendix to the training program.

5. The ability to modify the training program based on feedback gained through ongoing evaluation. Corrective feedback is crucial when modifying or upgrading subsequent stages of the training program.

Thus, personnelists committed to evaluation can enjoy benefits and advantages that have long been sacrificed in training designs without evaluation. Determination of the audiences of the evaluation can improve the likelihood of not falling into one of the many pitfalls which can effect the potential success of a program evaluation. In addition, although it may be impossible to design the "perfect" evaluation, major errors *can* also be avoided. The next section presents some of the evaluation pitfalls or mistakes that personnelists must be aware of in evaluating training efforts.

Pitfalls in Training Evaluation

Too often, training program evaluations have failed. Mainly these failures can be attributed to inadequate planning or designing, lack of objectivity, evaluation errors of one sort or another, improper interpretation of results and inappropriate use of results. Poor systems of training program evaluation produce anxiety, resentment, budget reductions, and efforts to sabotage the program. But what is of even greater importance, poor training evaluation programs do not provide firm data for improving and controlling the quality of the training system. Following are some common pitfalls or mistakes in training program evaluation (Tracey, 1971, 1984; Russ-Eft and Zenger, 1985; and Sims, 1990). Some of these pitfalls or mistakes can be easily overcome through good planning, while others are more difficult. However, personnelists must at least recognize the problems that occur when such mistakes or pitfalls occur.

Poor planning. To be effective, a training program evaluation must be carefully planned. Some of the common deficiencies in planning are these:

1. Failure to work out the details of the program, failure to include data-collection instruments, specific procedures to be followed, and the scheduling of surveys, interviews and observations.
2. Failure to train evaluators in the principles and techniques of evaluation, which includes the use of data-gathering instruments.

3. Failure to make clear to all concerned the purposes of the evaluation program and the uses to be made of evaluations and recommendations.

Lack of objectivity. Although it is impossible to guarantee that training program evaluations will be completely objective, there are some steps that can be taken to make certain they will be more objective.

1. Select evaluators who are capable of making objective judgments.
2. Train evaluators.
3. Design appropriate data-gathering instruments.
4. Look at all the components of the training situation as an integrated system.
5. Focus on important details—avoid “nit-picking.”

Rater errors. When scales are used to evaluate quality of performance

or materials, observers often differ in their ratings. These differences are rater errors, although this may not be the most accurate term to use for all these disparities. Some errors are caused by faults in the design of rating instruments; others, by the raters. Some errors occur only with certain groups of observers; and some occur only with individual observers. Other errors occur only when certain behaviors of individuals are rated. Some observers make errors when rating all persons; some when rating certain groups; and others when rating certain individuals. Some of the more typical rating error categories are: central tendency, halo effect, and recency.

Improper interpretation of data. The collection of data on the training program is one thing; interpreting that data is quite another. Here, the meaning and impact of the data are judged. If this step is not handled properly, the value of the information collected will be completely lost. Results from any evaluation must be interpreted cautiously, recognizing the extraneous variables that may have affected the findings. This is particularly true for those personnelists claiming to have identified productivity improvements resulting from a training program. Here are some of the main pitfalls in interpretation of data from training programs:

1. Assuming that consensus among one category of data on a single training system element guarantees a valid and accurate judgment.
2. Concluding that an observation or judgment made by only one observer or group of trainees for example, is inaccurate or invalid.

3. Taking comments or responses to open-ended questions at face value, and not considering the nuances of language and the problems of semantics.
4. Failing to take into consideration the perspective of the individuals providing the data.

Not reporting evaluation results in terms that are meaningful to the intended audience. Training evaluations in government agencies often yield results that are of little value to decisionmakers. This problem results because the evaluation collects the wrong information, uses technical jargon to describe the results or presents the results *after* critical decisions have been made. Evaluations of training programs conducted within agencies must focus on the direct outcomes of that training—behavior change. Personnelists must realize that the basic aim of any evaluation should be to ensure that relevant information is made available to decisionmakers (the audience) at proper times and in appropriate forms. By doing so, evaluation findings and generalizations may influence future decisions and policies.

Overgeneralization of findings. A problem related to the previous one is generalizing the findings of an evaluation in one agency to what might be expected in other agencies. Only by conducting repeated evaluations in many different agencies at many different locations can an accurate picture emerge. “Meta-analysis” (Glass, 1976) provides one means of examining different organizations. Meta-analysis is merely a grand term to describe the conduct of a summary analysis of a large collection of previous studies. Although the personnelist may need to be rigorous in their selection of high-quality studies and cautious in their interpretation of the results, such summaries can prove useful in identifying important trends in their evaluations.

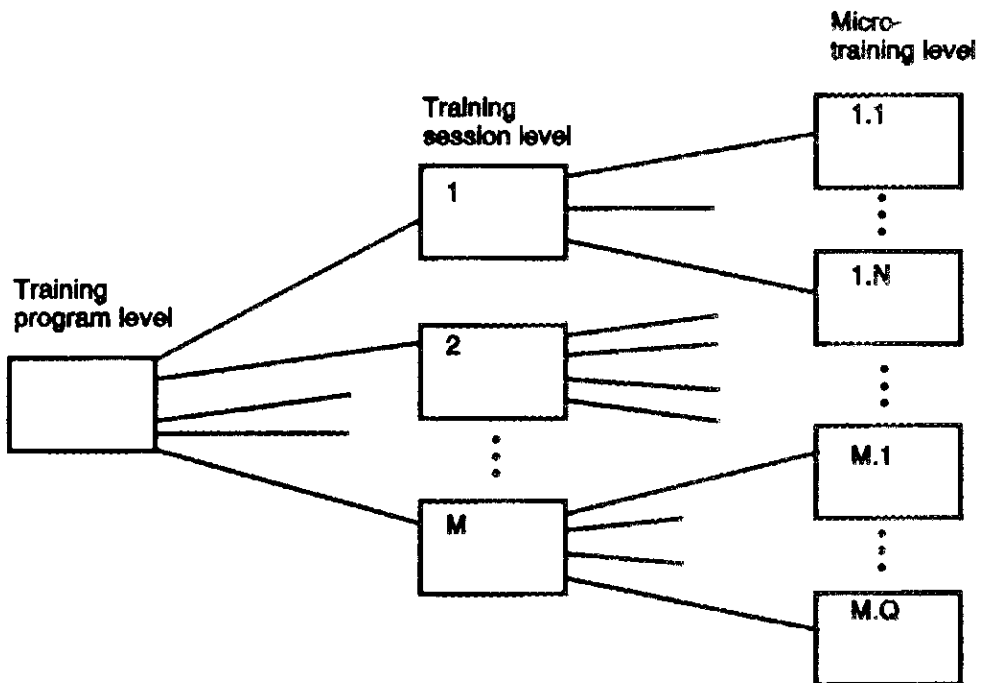
Inappropriate use of evaluation results. When tabulated, data collected during the training program evaluation have the aura of complete objectivity and truth. Sometimes the results of evaluation are used for purposes other than that originally intended. This is a major error. Some of the inappropriate uses to which evaluative data have been put are as follows:

1. Using data and reports on a single part of a training program to make decisions on the whole program.
2. Using data and reports designed for evaluating the whole training program as a basis for denying or granting funding for future training programs.

- Using otherwise unsupported and unvalidated data as a basis for causing significant changes to a training program or system to be made.

The personnelist should be aware of potential evaluation mistakes or pitfalls and keep in mind four general steps when organizing evaluations or new or existing training programs. First, determine why evaluation is being conducted. Second, identify resources needed to conduct the task. Third, design the evaluation process with particular emphasis on the role of the personnelist and other training participants. And fourth, implement evaluation even though this step most certainly will not always be smooth and efficient. Within this general framework, the evaluation of training by the personnelist can be thought of as being planned and executed at three separate, interacting levels adapted from Nicholas (1977) work in organization development and Sims (1990) in training: the training program level (TPL), the training session or component level (TSL), and the micro-training level (MTL) (Figure 1).

Figure 1 Interacting Levels of Training Evaluation



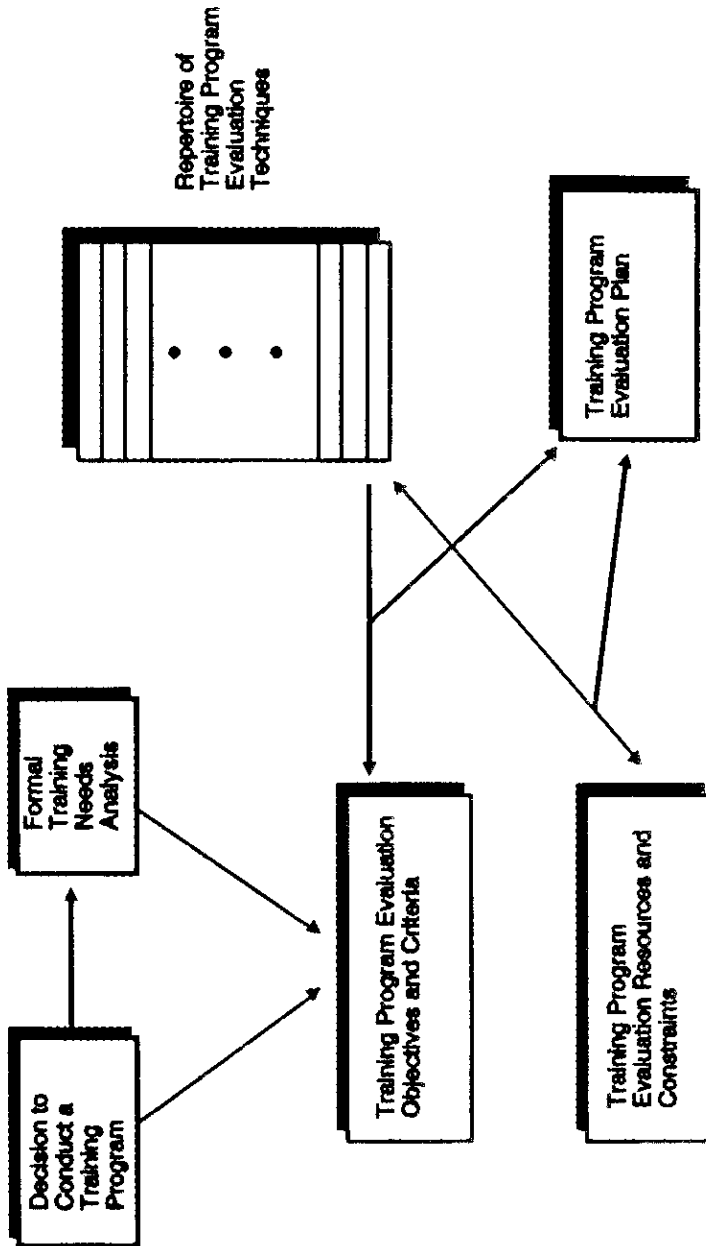
At the program level, planning involves establishing the broad strategy of the training program evaluation based on overall program goals. Personnelists and key agency decision makers work together to define the intent of training evaluation based upon perceived needs for the program or as determined by the personnelists in a formal needs analysis (diagnosis).

The training program evaluation plan requires specification, even if loosely, of the individual training sessions that will comprise it. This is necessary so that those responsible for implementing the evaluation program can establish a timetable of activities. The training evaluation materials must be obtained, and participants must roughly know how to schedule their time for the program evaluation. For each training component or session, evaluation objectives should be established to indicate how it will contribute to the overall training program evaluation. Once the individual training components have been loosely planned, it should be clear from statements of objectives of each session how the goals of the evaluation will be achieved.

During each component of the training program, the personnelist uses one or more "micro training levels." A micro training design level is a set of structured training activities used to achieve some training goal. The personnelist combines micro training designs such as those accepted in the training field and their own original designs to form sequences of activities to achieve program objectives. Each MTL has objectives which are compatible with objectives at the TSL and are the operants through which objectives at the TSL are achieved. Selection of MTLs and accompanying evaluations depend on the purposes they are to serve, and on the characteristics and constraints of the situation such as time limitations, costs, number of participants, and the level of training evaluation.

The considerations for developing evaluations of training programs should include some formal diagnosis of an agency's training needs and the training program that will be evaluated and are summarized in Figure 2. This diagnosis should specify training program evaluation objectives, evaluation criteria, and resources and constraints that will be encountered in planning and implementing the evaluation process. Wherever evaluation is called for, the personnelist must ensure that objectives are made in terms of clear statements that provide some amount of measurability. Criteria are those specific measures that establish whether or not objectives are met. Resources and constraints include not only money, personnel, equipment, time and space, but also, attitudes, norms and values of the agency toward training. From the possible training evaluation techniques, techniques are selected that will most likely achieve program objectives within given constraints and existing resources. This is a "systems analysis" approach applied to developing a training program evaluation plan. The result is an action plan indicating roughly the objectives

Figure 2 Considerations in Developing a Training Evaluation Program Plan



and procedures at all three training evaluation levels (program, session, and micro).

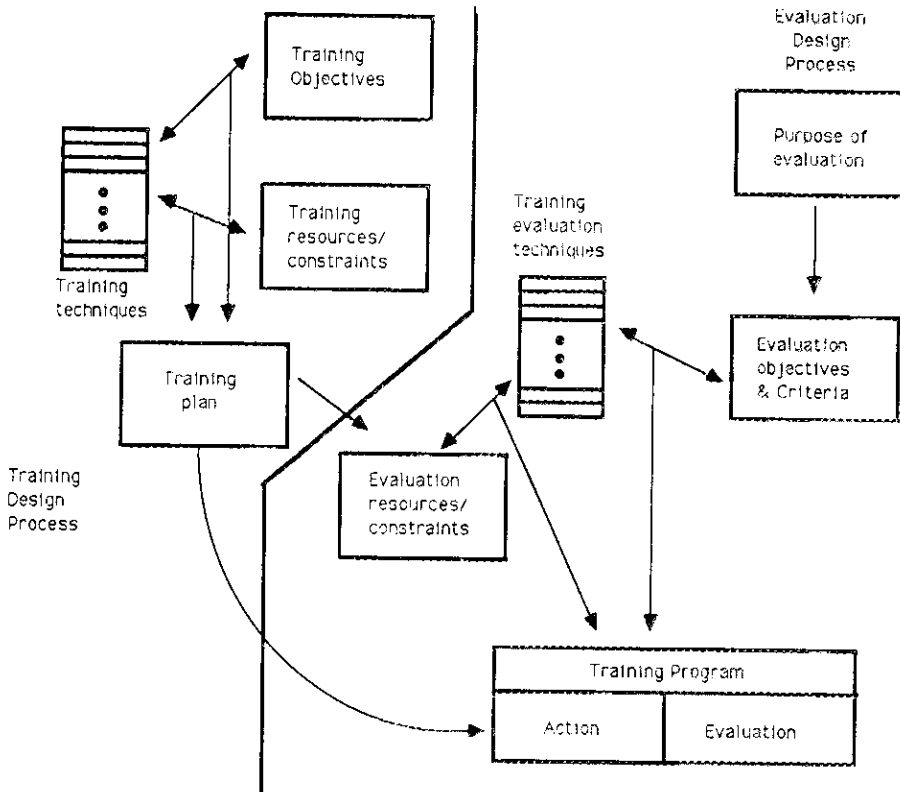
For training program evaluations that involve multiple components, such as indicated in Figure 1, specification of evaluation procedures for later sessions will most likely initially be vague; these are developed into more detail by the personnelist as their time of implementation draws closer. The important point is that thought is given by the personnelist as to how the micro designs and training sessions tie together to meet the overall objectives of the program. The pyramiding of objectives enables the personnelist to test assumptions concerning the ability of procedures at lower levels to meet objectives at higher levels (i.e., can the MTLs and TSLs evaluation activities reasonably be expected to achieve overall program objectives).

The Design of Training Program Evaluations

There is a need for more carefully designed training program evaluations and personnelists should be concerned with developing two major components of a training program. First, the training content or activities to be included in the program should be identified. Second, an outline or program for training evaluation should be developed. Simply put, the training program evaluation plan is created by (1) defining the agency's training needs, (2) deciding what has to be evaluated, (3) developing the training program with objectives and criteria clearly laid out to enable evaluation, and (4) developing an evaluation plan based on the objectives, criteria, and activities of the training program. Figure 3 presents a model for designing a training program with an accompanying evaluation program. The process on the right is the development of the evaluation part of the plan. As in the design of the action plan, the design of the evaluation part of the training evaluation is a systems analysis approach with antecedents of training program objectives, criteria, resources and constraints, and training evaluation techniques.

The left side presents steps necessary for designing the evaluation plan. Since the focus of this paper is training evaluation, the right side of the model is of immediate concern. For interested readers, training content development has been discussed more explicitly by others (see McGehee and Thayer, 1961; Goldstein, 1974; 1986, and 1989; Wexley and Latham, 1981).

Figure 3 Designing a Training Program Evaluation



Adapted from "A Systems Analysis Approach for Planning Evaluation of OD Interventions" by John M. Nicholas © 1977, pp. 358-362 Academy of Management Proceedings.:

Purpose of Training Evaluation

The first step in planning training evaluation is determining the purpose for evaluating the program; what do you want to know about the training program? Each kind of question necessitates consideration of how the evaluation program should be designed to provide answers. Stufflebeam et al. (1971) discuss three purposes for evaluation relevant to training program evaluation. First, evaluation can be used to identify differences in behavior. Individuals or groups may be compared to other individuals or groups, to ideals or standards (as in performance appraisal), or with themselves at different moments in time (a time-series study). This is a *comparative* evaluation. Evaluation can also investigate causes of behavior. The variables, within a training program, responsible for changes within individuals or groups can be identified through experimental manipulation. This is *explanatory* evaluation. Finally, a *predictive* analysis can evaluate how well training performance data correlate with other variables such as changes in individual job performance and/or system performance. The purpose for training evaluation will impact each step in developing both the content and evaluation components of the training program.

Criteria for Training Evaluation

The criteria used in a training evaluation program will depend on its purposes. Whatever the purposes of the evaluation, the personnelist can make use of five general categories, adapted from Suchman (1967):

1. Evaluation of *effort* assesses input, regardless of output. The questions "What did you do?" and "How well did you do it?" are addressed.
2. Evaluation of *performance* focuses on the results of the program. This requires clear statements of objectives; it asks "Did any change occur?" and "Were objectives achieved?"
3. Evaluation of *adequacy* determines how effective the program was. For example, trainees in a program may exhibit considerable practical application change, but the number of trainees may be inadequate in determining the benefits of the program for the agency.
4. Evaluation of *efficiency* is concerned with alternate ways of achieving the same ends that are more efficient in terms of time, money, human resources, materials, and convenience.

5. Evaluation of *process* focuses on the operation of a training program, with emphasis on how and why it works or does not work.

Besides Suchman, numerous criteria have been proposed (Schuler and Huber, 1990). The many options include changes in productivity, reported attitude on the job (e.g., satisfaction with supervisor, job satisfaction, stress, role conflict), changes in job knowledge, cost savings, benefit gains, and attitudes toward training (Burke and Day, 1986a and 1986b; Fisher and Weinberg, 1988; Fitzenz, 1988; Ford and Wroten, 1984; Kaman and Mohr, 1984; Kilpatrick, 1983; and Russ-Eft and Zenger, 1985). While several criteria have been tested through the years, at least four components commonly have been included in training evaluation: reaction to training, learning changes, transfer of training, results (Kilpatrick, 1983).

Employee reactions. Employee reactions to training are evaluated by conducting interviews or administering questionnaires to trainees. Here, personnelists are interested whether trainees liked the program, whether they thought the instruction was clear and helpful, and/or whether they believe that they learned the material. Even though positive employee reactions are necessary for training to be successful, positive employee reactions do not necessarily mean that training will lead to changes in knowledge or performance (Alliger and Janak, 1989).

In fact, trainee reactions constitute the lowest level of training evaluation (Birnbauer, 1987) and often can be misleading. Although reaction information can be useful for fine-tuning some aspects of the training program such as scheduling, mode of teaching, etc., data collected on the following three criteria generally provide stronger information for evaluating the program.

Employee learning. Instead of using employee reactions as the criterion in evaluating training performance, actual *employee learning* can usually be measured (Bell and Kerr, 1987). Employee learning criteria are used when personnelists wish to determine how well trainees acquire knowledge, skills, or abilities (KSAs) taught in training. Tests on the training material are commonly used for evaluating learning and can be given both before and after training to compare scores. Of course, learning enough to pass a test does not guarantee that the trainee can *do* anything with what was learned. However, learning tests should be considered since they are objective, reliable, and easy-to-administer tools for evaluating training effectiveness.

Application or transfer of training. Transfer of training concerns whether behavioral or performance changes taught in training are expressed on the job. Can trainees now do things they could not before (e.g., negotiate, conduct an appraisal interview)? Do they demonstrate new behaviors on the job? Has their performance improved? Data useful for

evaluating training transfer can be obtained through interviews of trainees and their co-workers and observations of job performance. Transfer of training is perhaps the most critical goal of training. Since training represents a step toward improving job performance and/or organizational effectiveness, a lack of transfer in KSAs to the job may indicate a poor training effort.

Results or bottom line measures. Results are evaluated by measuring the effect of training on the achievement of agency objectives. For example, "Did the agency actually save money following the training program?" or "How well were the training dollars invested?"

The difficulty with this measurement is pinpointing whether training caused the changes in results. Other factors may have had a major impact as well. Correlation does not imply causation. Still, the American Society for Training and Development (1988) notes that training must be treated as an investment, with similar potential payoffs as other organizational efforts. Cascio (1989) describes a training utility model that considers both the expected payoffs from training and an organization's capital investment decisions. He also provides computational formulas for examining important factors contributing to the cost-benefit ratio. Interested personnelists may do well to begin exploring training utility issues in Cascio's (1989) chapter.

Training Program Evaluation Resources and Constraints

When any training evaluation is undertaken a careful analysis of available resources and possible constraints should be included. Resources are the material and personnel needed for developing the training evaluation plan and subsequent implementation. Constraints are limitations or restrictions on what can be included in the evaluation plan or implementation.

Both are considerations in selecting among various methodologies and procedures to make up the evaluation plan. For example, time is a constraint for most training evaluations. The time available for collecting, analyzing data, and reporting results is limited. This influences not only procedures for data collection and analysis, but also the type of data collected. Time may not permit development of tailored survey instruments for collecting reaction data and personnelists may turn to alternatives such as off-the-shelf questionnaires. Such a decision may reduce evaluation time, however, the evaluator may be constrained to measure what the instrument purports to measure, rather than variables of interest. Some of the resources/constraints to training program evaluation that must be considered include:

1. *Funding*—the dollars allotted to cover training evaluation planning and implementation.
2. *Time*—limits imposed in developing and executing the evaluation. Time limits may be thought of as a sequence of “milestones,” such as completion of pretest and post-test data collection (for example, testing knowledge or skill usually by paper-and-pencil), completion of data analysis, dissemination of results to appropriate audiences, etc.
3. *Human resources*—trained personnel such as statisticians, computer specialists, research methodologists, and other personnelists.
4. *Organizational climate*—the trust and openness of administrators, employees, or trainees in providing and receiving evaluative feedback information.
5. *Availability of data*—availability and quality of organizational information such as records of individual, group, department and organization performance, reports, personnel records; availability of employees for providing new data through surveys, interviews and observation.
6. *Details of the training evaluation action plan*—objectives, timetable, procedures, participants, location; possible use of strategies which overlap evaluation strategies, such as survey feedback.
7. *Audiences*—kind and number of key players interested in the evaluation; information needs and interests.
8. *Technical ability and feasibility*—availability and feasibility of using standardized instruments, computerized analysis and storage of data; logistics in collecting and disseminating results; competencies and abilities of persons involved.
9. *Ethical concerns*—privacy considerations, employee and organizational confidentiality, obtrusiveness, or harmful aspects of data collection and reporting.

To a large extent, these are interdependent factors to which the personnelist must attend during the training program planning analysis.

Training Program Evaluation Design Options

Upon identifying important training outcome criteria, personnelists must select an evaluation design to measure changes in these variables. There are many ways to design and evaluate training programs to determine their effects. The three most common are shown in Figure 4. The level of rigor of the designs increases from *Post*, *Pre/Post*, to *Pre/Post Measure With a Control Group* (Mathis and Jackson, 1991).

Designing a good evaluation effort involves knowing when to collect evaluation measures and which groups to collect them from. Together, these factors determine the experimental design used to assess the impact of training. More specifically, the training evaluation design refers to the steps taken to ensure:

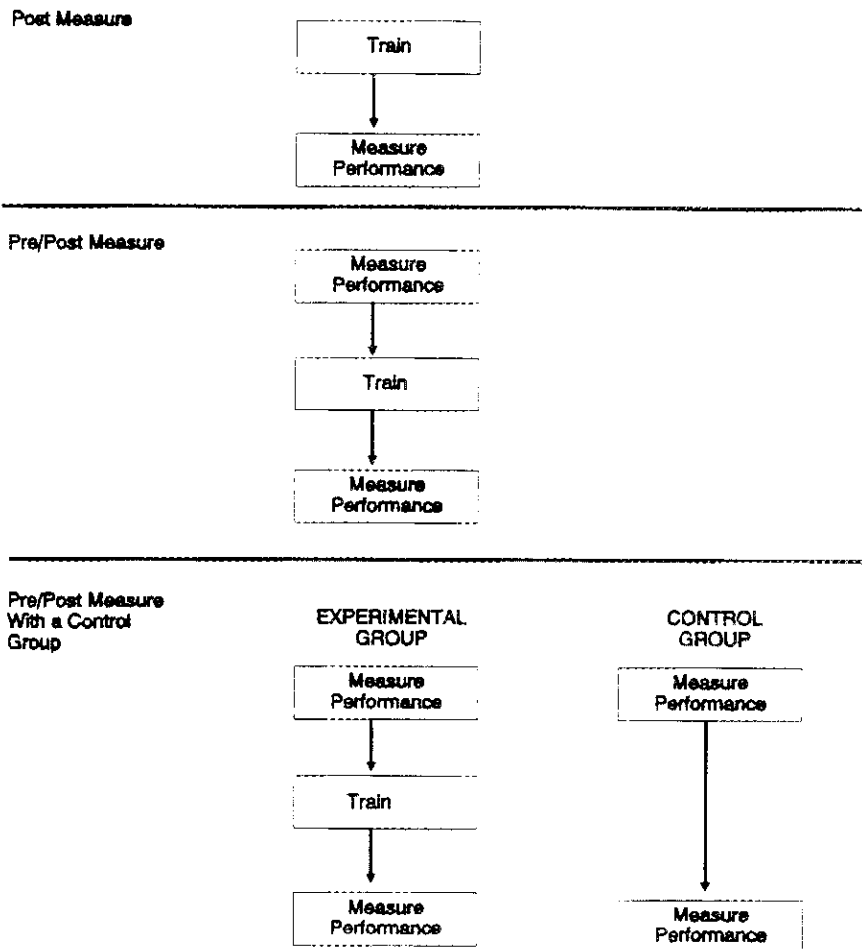
- a. that a change occurred (e.g., employee productivity increased, accidents declined, etc.);
- b. the change can be attributed to the training program; and
- c. that a similar change could be expected if the training were done again with other employees.

Of course, the ability to make such statements will depend on the experimental rigor incorporated in the training evaluation process. Conclusive statements about the effectiveness of training can be made only when the personnelist strictly adheres to experimental principles such as manipulation of variables, random assignment, control of extraneous and/or confounding variables, and equivalence of groups. Unfortunately, conducting experimentation in the field has proven to be a difficult, almost overwhelming, task. Many agencies generally demand that all employees in a department be trained, not only those randomly selected. It is also difficult for personnelists to control the many variables that can affect a worker's job behavior at a given time (e.g., interaction with co-workers, supervisor, personal relationships, promotions, etc.). However, previous training evaluations have been able to overcome these difficulties by using several highly effective designs for evaluating training (See Cook and Campbell, 1979; Kilpatrick, 1983; Smith and George, 1984; Wehrenberg, 1983; Hoffman, 1984 for more detailed discussions of the available designs depicted in Figure 4).

Regardless of the design selected, personnelists must adhere to certain basic experimental principles. First, when possible, both pretest and post-test data should be collected on relevant criteria. Second, selection of participants should be randomized when possible. If it is not possible, solid quasi-experimentation should be employed. Third, reliability of data col-

lected should be monitored throughout the evaluation. Fourth, when statistical analyses are performed, characteristics of the sample and data should conform to the assumptions of tests used. Finally, the evaluation process (i.e., training development, data collection and implementation) should be conducted in a manner to allow valid inferences about the effectiveness of training.

Figure 4 Training Evaluation Designs



The final success of the training program evaluation depends on how well the personnelist can overcome problems through a methodologically sound design and implementation scheme. The evaluation plan should be developed through selection of alternatives assessed against the objectives of the evaluation and existing constraints and resources. A familiarity with available resources, imposing constraints, and methodological alternatives will allow accurate, useful, and practical training program evaluations.

Multiple Levels of Training Program Evaluation

Thus far, this paper has described strategies for evaluating the overall or general outcomes of training. For purposes of discussion in this section training evaluation can be delineated at three levels: micro training evaluation (MTE), training session evaluation (TSE), and training program evaluation (TPE). MTE focuses on smaller aspects of the training program such as timing and scheduling of training activities, presentation of training information, the outcomes of individual exercises, and etc. For most training efforts, a rigorous MTE will not be conducted, since simple observation may be all that is needed to determine the effectiveness of such aspects of the program. However, factors such as scheduling and presentation method can greatly affect the other levels at which training evaluation can occur.

TSE focuses on the quality, processes, and outcomes of individual training components or sessions. Here, the personnelist wants to ensure that sessions have desired effects. Data collection can be as informal as interviewing trainees to assess their perceptions of the session, or as formal as testing trainees and interpreting scores to determine what was learned. Criteria such as learning and reaction are appropriately measured at this stage of the program. Evaluation at the training session level can provide valuable information useful for: 1) identifying areas where trainees may require additional training and/or 2) modifying existing training practices. Modifying the training based on employee reaction measures is fairly straightforward. A boring speaker can be replaced or a video rated as irrelevant dropped, depending on the feedback received.

At the training program evaluation level the criteria discussed earlier (employee learning, application of training, results, and employee reaction) are evaluated in relation to the entire agency. The methods and strategies described earlier would apply. Where MTE and TSE may be conducted in a matter of hours or days, TPE can continue for months or years following the training sessions to determine these effects. At this level, where evidence of outcomes or results is more mediated and distal than at lower levels, training evaluation is more difficult to carry out and usually is more complex in design.

The three levels of training evaluation are represented in Figure 5. This guide can serve as a "mental checklist" for personnelists when developing training programs to better understand the various components of the plan and how they relate to one another. A training program's objectives, criteria, procedures, and activities (methods) can be entered into columns 1 and 2. Column 1 contains the objectives of the training program and column 2 contains the methods that will be used to obtain training objectives. When evaluation is desired, the objectives and criteria for evaluation should be entered into column 3. Column 4 will contain the evaluation techniques used to accomplish the column 3 objectives. Column 4 includes the methodology and techniques for data collection, analysis, and reporting.

Personnelists should also derive *expected outcomes* of the program in terms of the objectives and criteria in columns 1 and 3. Specifying expectations is useful for two reasons: 1) it forces personnelists to judge whether the intended training program outcomes can reasonably be expected, and 2) it forces them to assess whether the training evaluation plan measures achievement of desired results. The final training program evaluation plan should enable personnelists to visualize the total training program process and anticipate further problems or design considerations. Thinking about expected outcomes also forces personnelists responsible for training to closely monitor steps in the training process to determine whether results are occurring and outcomes are feasible.

In conclusion, an agency training program should address the following questions to ensure adequate evaluation of training outcomes:

1. Does the evaluation design fit the objectives of the training program?
2. Does the design address important issues such as trainee needs and expectations, learning style, and trainee culture (expectations about authority, how hard to work, etc.)?
3. Does the evaluation method meet standards discussed by public sector trainers?
4. Does the structure provide a framework where emergent issues can be addressed? Can the design be modified to address trainees' felt needs without sacrificing objectives?
5. Can the design be carried out in the time allotted?
6. Does the design provide a mix of activities that appeal to different learning styles such as listening, discussing, interacting?

Figure 5 Training Plan and Evaluation

Level	Training Plan		Training Evaluation Plan	
	1 Objectives	2 Procedures	3a Objectives	4b Procedures
Program		c		
1		d		
2		d		
Training Session :		:		
M		d		
1.1		e		
:		:		
1.		e		
Micro-Training :		:		
M.1		e		
:		:		
M.Q		e		

- a. Evaluation objectives and criteria are specified only for those levels, sessions, and micro-designs where evaluation is to be done.
- b. Evaluation procedures are specified only for those levels, training sessions, and micro-training designs where evaluation is to be done.
- c. At the training program level, procedures roughly outline what is to take place at the session level: number of sessions, timetable, etc.
- d. At the training session level, procedures roughly outline what is to take place during each session: kinds of micro-training designs, participants, etc.
- e. At the micro-training design level, procedures define in detail step-by-step procedures, activities and timing for each.

Adapted from "A systems analysis approach for planning evaluations of OD Interventions" by John M. Nicholas, 1977, pg. 358-362, Academy of Management Proceedings.

7. Is the material logically and psychologically sequenced?
8. Is there redundancy in information presented in training? Should there be?
9. Does the evaluation design allow for on-going development of a learning climate?

Summary

Carefully designed and planned evaluations are keys to better agency training. However, one problem with many agency training programs is that they are not designed with *a priori* consideration given to evaluation. Evaluation should be an integral part of training and training program objectives and procedures should be compatible with evaluation purposes and procedures. In addition, personnelists need to realize that evaluations of training programs are conducted to answer certain questions, and that there may be different questions at different times in the history of a program. The information needs and interests of key agency members and interested parties should be identified early, rendering training program evaluation results credible.

Today, there is an increasing recognition that evaluation is critical to training. The "systems" approach to planning, designing, and implementing a training evaluation presented in this paper should improve the personnelists ability to conduct evaluation efforts. The personnelist must develop an orderly, goal-oriented, and systematic approach to training evaluation that considers a range of options useful for overcoming the resource shortages and organizational constraints often confronted in agency training.

In any training program instance the personnelist must first determine that it (training program) does have some beneficial effects. Several questions that should be answered include: What outcomes does it really have? How long-lasting are these effects? Which people and behaviors change? The final series of questions to be asked focuses on improving the product or program. They include: How can the delivery be improved? What improvements can be made in the content or the schedule? Answers to these questions can lead to improvements in the training program. These questions can only be answered by carefully planned, designed, and implemented evaluations. Thus, agency training programs will only progress if personnelists do the appropriate research and do it right.

While the problems facing personnelists in training evaluation allow for no easy solutions, evaluations can nonetheless be rewarding. Basically,

all parties gain through evaluating training—the agencies, programs, the training departments and the personnelists. Such assessments and evaluations yield important information, but only when care is taken in planning, designing, and conducting the evaluation. As Brooks (1977) writes, “if proper care is given to the manner in which findings are obtained and interpreted, the positive contributions [of the evaluator] will far outweigh the negative implications deriving from the imperfection of his tools” (p. 62). By continually measuring the effects of training, improving programs based on evaluation data, and keeping decision makers informed of and involved in the evaluation process, the full value of evaluation as a potentially powerful organizational tool can be realized by agencies.

References

- Alliger, G. M., and Janak, E. (1989). Kilpatrick's levels of training criteria: Thirty years later. *Personnel Psychology*, 42, 331-342.
- American Society for Training and Development. (1988). *Gaining the competitive edge*. Alexandria, VA: American Society for Training and Development.
- Bell, J.D., and Kerr, D.L. (1987). Measuring training results: Key to managerial commitment. *Training and Development Journal*, 41(1), 70-73.
- Birnbrauer, H. (1987). Evaluation techniques that work. *Training and Development Journal*, 41(1), 53-55.
- Brethower, S., and Rummel, G.A. (1979, May). Evaluating training. *Training and Development Journal*. 14-22.
- Brooks, M. (1971). The community action program as a setting for applied research. In F. Caro (Ed.) *Readings in evaluation research* (pp. 53-62). New York: Russell Sage Foundation.
- Burke, M. J., and Day, R. R. (1986a). A cumulative study of the effectiveness of managerial training. *Journal of Applied Psychology*, 71, 232-45.
- Burke, M. J., and Day, R. R. (1986b, August 21). Cost-effective training techniques. *Bulletin to Management*, p. 284.
- Cascio, W. F. (1989). Using utility analysis to assess training outcomes. In I. L. Goldstein (Ed.), *Training and development in organizations* (pp. 63-88). San Francisco, CA: Jossey-Bass Publishers.
- Cook, T. D., and Campbell, D. T. (1979). *Quasi-Experimentation: Design & analysis issues for field settings*. Boston, MA: Houghton Mifflin Company.
- Fisher, H. E., and Weinberg, R. (1988, January). Make training accountable: Assess its impact. *Personnel Journal*, 67, 73-75.
- Fitzens, J. (1988, March). Proving the value of training. *Personnel*, 65, 17-23.
- Ford, J. K., and Wroten, S. P. (1984, Winter). Introducing new methods for conducting training evaluation and for linking training evaluation to program redesign. *Personnel Psychology*, 651-66.

- Galagan, P. (1983). The numbers game: Putting value on HRD. *Training and Development Journal*, 37 (8), 48-51.
- Glass, G.V. (1976). Primary, secondary, and meta-analysis research. *Educational Researcher*, 5 (10), 3-8.
- Goldstein, I. L. (1974). *Training: Program development and evaluation*. Monterey, CA: Brooks/Cole.
- Goldstein, I. L. (1986). *Training in organizations: Needs assessment, development, and evaluation* (2nd ed.). Monterey, CA: Brooks/Cole.
- Goldstein, I. L. (Ed.) (1989). *Training and development in organizations*. San Francisco, CA: Jossey-Bass Publishers.
- Hamlin, A.C. (1970, November). Evaluatiotraining. *Industrial Training International*, 33.
- Hoffman, F. O. (1984, February). A responsive training department cuts costs. *Personnel Journal*, 63, 48-53.
- Kaman, V. S., and Mohr, J. D. (1984, October). Training needs assessment in the eighties: Five guideposts. *Personnel Administrator*, 29, 47-53.
- Kilpatrick, D. L. (1983, November). Four steps to measuring training effectiveness. *Personnel Administrator*, 28, 19-25.
- Lee, C. (1989, October). 1989 Industry Report. *Training*, 35, 49-64.
- Lusterman, S. (1985). *Trends in corporate education and training*. The Conference Board, Report No. 870.
- McGehee, W., and Thayer, P. W. (1961). *Training in business and industry*. New York: Wiley.
- Mathis, R. L. and Jackson, J. H. (1991). *Personnel/Human Resource Management*. St. Paul, MN.: West Publishing Co.
- Nicholas, J.M. (1977). A systems analysis approach for planning evaluations of OD interventions. *Academy of Management Proceedings*, 358-362.
- Rao, T.V., and Abraham, E. (1986). Human resource development. *Management Labour Studies*, 2, 73-85.
- Russ-Eft, F., and Zenger, J. H. (1985, April). Common mistakes in evaluating training effectiveness. *Personnel Administrator*, 30, 57-62.
- Schuler, R. S., and Huber, V. L. (1990). *Personnel and human resource management* (4th ed.). West Publishing Company. St. Paul, MN.
- Sims, R.R. (1990). *An experiential learning approach to employee training systems*. Westport, CT.: Quorum/Greenwood Press.
- Smith, H. W., and George, C. E. (1984, August). Evaluating internal advanced management programs. *Personnel Administrator*, 29, 118-31.
- Stufflebeam, D., Foley, W., Gepart, W., Guba, E., Hammond, R., Merriman, H., and Provus, M. (1971). *Educational evaluation and decision making*. Itasca, IL: Peacock Publishers.
- Suchman, E. (1967). *evaluative research*. New York: Russell Sage Foundation.
- Swiercaek, F. W. and Carmichael, L. (1985, January). The quantity and quality of evaluating training. *Training and Development Journal*, 95-99.

Tracey, W.R. (1984). *Designing training and development systems*. Washington, D.C.: American Management Association.

Tracey, W.R. (1971). *Designing training and development systems*. Washington, D.C.: American Management Association.

Wehrenberg, S. B. (1983). Evaluation of training: Part I. *Personnel Journal*, 62, 608-10.

Wehrenberg, S. B. (1983). Evaluation of training: Part II. *Personnel Journal*, 62, 698-702.

Wexley, K. N., and Latham, G. P. (1981). *Developing and training human resources in organizations*. Glenview, IL: Scott, Foresman.