Evaluation Framework, Design, and Reports

A training program is a success if it achieves timely results consistent with pre-established participant performance objectives related to wider organizational goals.

Much the same is true of evaluation methods. Evaluation only needs to provide sufficient information to assure that a training program is meeting its objectives—and that those objectives further attainment of organizational goals and objectives. Evaluation methods must provide results in time to inform decision makers as they consider choices for current and future training. After all, the purpose of training is to improve performance, and the purpose of evaluation is to improve training's effectiveness and efficiency.

Historical failures or inabilities to evaluate both training's costs and benefits have rendered training particularly vulnerable to cost-cutting pressures, and have inhibited its use as a lever for effecting strategic change. The fact that fewer than half of America's training programs are formally evaluated indicates implicit managerial trust that, somehow or other, training facilitates attainment of organizational goals. Yet, as one trainer put it, "The worst thing that ever happened to training is that it was taken on faith that it was good." As cost pressures increase, trainers must demonstrate training's value in more substantive ways if it is to gain its rightful place among investment alternatives.

Admittedly, measurement can never completely ascertain a training program's effectiveness or its efficiency in achieving beneficial effects. What worked at one time at one training location with a unique group of participants can't necessarily be transferred to another time, setting, and group and be expected to work as well. Still, evaluations build a case of support for training by providing an approximation of its value.

The Kirkpatrick Model

The evaluation framework that most training practitioners use is the Kirkpatrick Model. Although this model doesn't accommodate all the evaluation methodologies that training managers employ, it's the most widely known evaluation model and illustrates a commonly used set of levels or rigors of evaluation.

Almost universally, organizations evaluate their training programs by emphasizing one or more of the model's four levels. In summary, these levels are as follows:

- **Reaction.** How well did training participants like the program?
- **Learning.** What knowledge (principles, facts, and techniques) did participants gain from the program?
- **Behavior.** What positive changes in participants' job behaviors stemmed from the training program?
- **Results.** What were the training program's organizational effects in terms of reduced costs, improved quality of work, increased quantity of work, and so forth?

Participant reactions are easy to collect, but provide little substantive information about training's worth. At the other end of the scale, results-level information is more difficult to collect, but provides data to analyze for assessment of training's organizational impact. In general, the more data sources used to evaluate a training program, the more complete is the picture of its effectiveness.

The appropriate levels of evaluation data to gather and analyze depend on the evaluation "clients." As a rule, line managers have more interest in performance change and organizational results than in participant reaction and learning. A training department, on the other hand, would have an interest in collecting reaction and learning data to determine what components of training could be improved.

Currently, most employee training is evaluated at the reaction level. Evaluation at this level is associated with the terms "smile test" or "happiness test," because reaction information usually is obtained through a participant questionnaire adminstered near or at the end of a training program.

The fact that fewer than half of America's training programs are formally evaluated indicates implicit managerial trust that, somehow or another, training facilitates attainment.

Because reaction information doesn't reveal what participants have learned or whether what they've learned will transfer to their jobs, it isn't indicative of training's return on investment. This lack and frequent misapplications of reaction data have caused some evaluators to deride its collection. But most training practitioners believe that participants' favorable reactions are critical to training program success—because people learn better when they accept training willingly and react positively to the form it takes. The clients of training most likely to have an interest in reaction data are participants themselves, training department staff, and course instructors.

At the learning level of evaluation, tests are used to measure the knowledge, skills, or attitudes that participants acquired during training. These tests should reflect each training program's particular objectives. For instance, for an introductory skills course, a participant

may be considered successful if able to accomplish a given task at all, or the participant may be required to perform at a speed consistent with on-the-job requirements.

Measures of learning changes may be taken during a training program or at its conclusion. This level may indicate that a program's instructional methods are effective, but it doesn't show whether or how participants new learning will be applied on the job. The clients with the greatest interest in examining learning changes through training are participants, training department staff, and course instructors.

The third evaluation level deals with behavior or performance changes on the job. When learning doesn't transfer to the job, the two most likely reasons are that the work environment doesn't support the learned behavior or that a participant thinks the training was irrelevant. A participant's supervisor or colleagues (or the physical work setting) may discourage newly learned behavior, or the participant may believe that the new behavior, although encouraged by management, won't lead to personal benefit. That's why participants' managers should be consulted about training program development and should tell participants how the training will help them maintain or improve their positions.

Even with supervisory support and participants' commitment to transfer new learning to the job, there's no guarantee of increased productivity or ability to meet intended organizational economic objectives—because the changed performance may not be in keeping with desired organizational outcomes. That sorry situation occurs if a training program's objectives aren't properly established when the program is developed.

The ultimate measure of training program success relates results to organizational objectives. The results level of evaluation is the highest level of rigor and usually is expressed in organizational terms such as reduced costs. Organizational analysis, needs-identification forecasting, and needs analysis reveal areas where training can make a contribution to achievement of key organizational objectives.

It's difficult to isolate the beneficial organizational results of training. Nonetheless, training specialists need to work from an understanding of organization-level goals and objectives. In many cases, it is possible and feasible to link training contributions to organizational improvements. Doing so doesn't require absolute isolation of training's contributions. Rather, it requires indicators that demonstrate training's valuable role within the organization's systems.

A systems view

Taking a systems approach means looking at the interconnection of organizational parts and the relationship of the whole organization to its environment. Organizations are composed of functionally related components arranged in a hierarchy of subsystems, systems, and supersystems. Organizing subsystems into larger systems helps coordinate activities and processes in order to fulfill the overall mission of facilitating organizational goals. The translation of organizational goals into identified current objectives provides context, meaning, and direction for the entire organization.

Training is one operating system within an organization, and training should be evaluated as a system in support of other systems. In other words, training must contribute to achievement of the goals of the departments it serves and, through these, contribute to organizational objectives. This view of training necessitates the use of systematic means for performing training and linking its role to the goals of higher systems. Evaluation of training is the main method used to assess whether training is accomplishing desired effects of sufficient value.

Evaluation steps

According to Ratzlaff, there are three major evaluation steps:

- setting an evaluation's purpose
- selecting evaluation methods and design
- reporting evaluation findings.

Setting an evaluation's purpose. First, an evaluator must determine whether an evaluation will be formative or summative. Formative evaluations, sometimes called improvement-oriented evaluations, are used to decide whether a program should be modified. Summative evaluations are used to decide whether a training program should be maintained, expanded, contracted out, or eliminated. Formative evaluations are carried out before and during the first running of a training program. Summative evaluations collect information about a training program's value after it has run at least once.

An evaluation's purpose and scope are delineated by the questions or issues its clients want addressed. If clients aren't directive about what an evaluation should accomplish, it may take some discussion between a client and an evaluator to set evaluation objectives.

Continuing client involvement in planning provides opportunities to reassess a training program's objectives and, consequently, to reassess aspects of evaluation. Talking about relationships between training and intended outcomes can reveal whether training activities logically flow from objectives that support goals, and whether objectives relect realistic expectations.

An evaluator needs interpersonal skills to build rapport with clients and establish a climate of trust and candor. A free flow of information will be promoted by an evaluator's flexibility in considering possibilities for an evaluation's focus and design.

Laying the groundwork for a common understanding of a training program means asking clients questions. An evaluator should paraphrase and restate each client's views about a training program. That helps ensure that the evaluator has interpreted the client's views correctly; if not, it points out what needs clarification.

An evaluator must learn how the primary client—the evaluation's financial sponsor—expects to use evaluation information, because the answer affects what data will be collected and how findings will be presented.

The evaluator must also decide to what extent an evaluation can accommodate the questions of other stakeholders in the program—such clients as instructional designers, instructors, training participants, and their supervisors. It may not be possible to deal with all the issues raised.

The evaluator next develops tentative evaluation objectives and plans that lead to new questions. For a formative evaluation, there will be questions about how to staff the program and how to select participants (for instance, by mandatory or voluntary attendance, or on the basis of current or future needs). A summative evaluation might seek to characterize participants who benefited most from a program (for example, those who stayed for optional lab work or those who had a designated learning partner).

Selecting evaluation methods and design. As soon as the training program's nature (formative or summative) and tentative objectives are established, the evaluator begins to develop data-collection plans.

Data-collection strategy relates to how participants will be grouped for evaluation and when evaluation measures will be taken. An evaluator should always use more than one data source when assessing an aspect of a training program, because multiple sources give a rounded view and build stronger support for evaluation findings.

Once evaluation questions are outlined, the evaluator begins listing quantitative or qualitative data-collection techniques to use to answer each question. Whenever possible, to be economical, participants' learning and application of learning should be measured through data-collection techniques and tools already in use. Constructing measurement procedures and instruments is time-consuming and expensive, and existing procedures and standardized instruments for assessing participants' reactions or learning gains are often adaptable for many training programs.

Still, new data-collection procedures and instruments should be developed for a training program if the evaluation addresses unique questions, if no suitable standardized data-collection tool or technique is available, or if it's critical that program outcomes be measured precisely.

Evaluators should determine whether the primary client disapproves of a particular data-collection technique. For instance, if a client regards observation of training participants as obtrusive, then this technique should not be used. An evaluator must also select techniques that can answer evaluation questions within timeframes set by management. To do so, the evaluator must allow for time to train any others who'll be responsible for data collection. For example, line managers may need a training session or two if they'll be responsible for observing and assessing participants' learning.

Before proceeding, an evaluator may want discuss with the client the tradeoffs between implementing a practical but less rigorous evaluation design versus a more credible—but also more expensive—design. If necessary, the client may be persuaded to provide more time or funding or to narrow evaluation objectives.

At this point, an evaluator should review background information about the training program from the following items:

- Documents. For a formative evaluation these include front-end analyses prompting training and previous training project proposals, goal statements, budgets, materials, publicity, evaluations, memoranda by program staff. For summative evlaution, the evaluator will survey documents for the current program.
- Discussions with participant and training staff— instructors, instructional designers, previous evaluators, and training administrators. Consultation with program staff gains their insights and promotes their comfort and cooperation with the evaluation.
- Observations. After examining a variety of sources, an evaluator should write a detailed description of training program objectives and activities designed to achieve them. Once the client and evaluator reach agreement about the description, it should be distributed to program staff and training administrators.

In keeping with the program description, an evaluator should establish a schedule of evaluation activities to determine the order, length and, possibly, dates of evaluation activities. The schedule begins with a list of each major evaluation effort and its subactivities to each of which the evaluator adds starting and completion dates.

Throughout evaluation, an evaluator should note and record variances between expected and actual time devoted to each evaluation activity. This information facilitates the development of realistic scheduling for future evaluations or, in some cases, may be useful for specifying to the current client when and where unexpected obstacles are being encountered.

Reporting evaluation findings. An effective communication strategy disseminates information to the client and program staff through informal conversations, discussions, and meetings; formal meetings and presentations; and memoranda and written reports.

Depending on the scope, complexity, and formality of an evaluation and on clients' expectations, an evaluator may incorporate the evaluation schedule into a formal evaluation contract that states the following:

- the evaluation's purposes
- aspects of the training program to be examined
- data-collection techniques to be used
- a schedule of evaluation activities
- formats for reports of evaluation findings.

Even if there's no formal evaluation contract, timing and formats for the release of evaluation information should be part of the strategic planning for evaluation. Especially for a formative evaluation, frequent piecemeal release of findings to client and program staff is preferable to a lengthy final report because frequent communication allows for adjustments to the program or its evaluation.

Again, an evaluator's interpersonal skills are important because an evaluator needs to be tactful, informative, and non-threatening when communicating that a training program needs improvement.

The key evaluation document is an evaluation report of final evaluation findings that will affect clients' future actions. These reports tend to be longer for summative evaluations because, by nature, formative evaluations call for numerous interim reports. In any case, to be valuable, an evaluation report must be timely and clearly reported in four sections:

- ***Executive summary.*** This one-page synopsis of the report may be all that some clients read, so it should be particularly well reported.
- ***Introduction.*** This describes the training program's

objectives, which training program aspects were assessed, and what the evaluation's timeframe and budget were.

■ ***Data-collection techniques and evaluation design.*** This describes overall evaluation design, how training participants were selected, each data-collection tool or instrument, the evaluation questions each addressed, and the participants to whom each was applied or administered and the number of times administered.

■ ***Evaluation findings.*** This reiterates which training program aspects were investigated and provides the evaluation's major findings. This section should also describe constraints on the evaluation—because such major restrictions as time and budget can influence an evaluation's credibility.

If the evaluation is to recommend a program's cancellation or continuance, that recommendation appears here. For continuing programs, suggestions for revisions are offered, focusing on strengthening the programs for attaining better results.

Phases of evaluation

As already noted, training program evaluation doesn't just follow a training program's completion. Evaluation has three, roughly sequential phases:

■ It determines whether a training program as planned is likely to meet organizational goals.

■ It monitors training in process to ensure that training is being conducted as planned or undergoes necessary correction.

■ It determines whether a program as implemented met expectations.

Planned evaluation

Planned evaluations analyze the environment in which training is developed. This involves analysis of the organization's goals, resources, and people. Planned evaluations assess discrepancies between expected and actual performance, current employee skills, and training deficiencies.

Planned evaluation begins with the fundamental question: Will the organization benefit from providing training? The major components of a planned evaluation are organizational analysis, needs-identification forecasting, needs analysis, and marketing of training.

Organizational analysis

Trainers must be proactive rather than reactive in considering an organization's need for training. Instead of waiting to respond to training needs and opportunities that become clearly evident, trainers must anticipate performance problems and opportunities.

To be proactive, training departments translate organizational objectives into training objectives, then plan, deliver, and evaluate programs in terms of contribution to objectives. This means that preparation of training programs must begin with analysis of data collected from the entire organization.

It's impractical to conduct a full-scale organizational analysis at the beginning of each training program. But periodic organizational analysis gives guidance and a realistic context to training decisions. An annual analysis sponsored by the training department can:

■ help determine the state of the organization

■ help training department staff understand the organization and organizational systems they serve

■ serve as a base for future training programs by revealing training needs.

In the end, organizational analysis should:

■ describe immediate and future objectives of the organization and of its particular subdivisions

■ identify as realistically as possible the organization's economic, social, and political environment

■ discuss the organization's structure and resources, including equipment and facilities, monetary resources, and current human resource skills.

This analysis allows trainers to design training programs that meet organizational goals in the face of economic and political pressures and in the context of available resources.

Organizational analysis is often overlooked because of immediate pressure to conduct a particular training program for which the need is regarded as obvious. But without such analysis, trainers may believe a training program will fill a fundamental organizational need when, in fact, it would fill a only short-term or minimal need or treat only one symptom of a larger problem. If so, program staff may mistake their roles, misunderstand organizational direction, and actually impede accomplishment of goals. The result is frustration, waste, and trivialization of the training function.

So an organizational analysis should be as thorough as possible and based on information from a variety of sources. First, analysts should examine written organizational mission statements, strategic plans, statements of philosophy, and statements of objectives.

Consultations with executives or their managerial representatives can be an excellent source of information about the organization's immediate and long-term objectives. But it may be difficult for an analyst to gain access to upper managers or, in the short time available for consultation, it may be hard to arrive at focused objectives with a clear bearing on training.

Training managers can overcome these difficulties through the following activities:

■ devising and distributing short forms that allow managers to explain themselves easily and whenever their schedules permit

■ requesting appointments with selected managers and sending interview instruments to them in advance so they see what kind of information is needed and have time to consider their responses.

Needs-identification forecasting

A needs-identification forecast begins with scrutiny of key business indicators to determine whether an organization is achieving its goals or falling short of them.

As in organizational analysis, that involves examination of the organization's stated objectives, strategies for achieving those objectives, and measures of its success in attaining goals. But for needs-identification forecasting, analytical activity is directly aimed at discovering how organizational objectives may translate into training needs. So special attention should be given to human resource plans incorporated into goals, strategies, and objectives in order to analyze whether the plans

■ are feasible in terms of dedication of organizational resources to support them
■ relate to training programs that have run in the past or are likely to require development of new training programs.

If organizational goals and objectives aren't being met despite dedication of seemingly sufficient resources, human performance must be analyzed in terms of current and future business goals and objectives.

The data for this needs-identification forecast may be gathered from the following sources:

Organizational performance records related to
■ productivity measures quantifying output per units produced, tons manufactured, items assembled, money collected, items sold, forms processed, loans approved, inventory turnover, patients visited, applications processed, students graduated, tasks completed, output per hour, work backlog, number of incentive bonuses earned, or shipments made
■ quality measures such as scrap level, error rates, amount of waste, rework time, shortages, deviation from standards, product failures, inventory adjustments, and customer compliments or complaints
■ workforce performance and behavior indicators such as measures of absenteeism, tardiness, injuries, number of promotions or pay increases, training program attendance, requests for transfer, performance appraisals, and turnover rate
■ Safety and regulatory measures that count the number of employee accidents and injuries, Occupational Safety and Health Administration (OSHA) litigations, and reported safety violations.

Questionnaires for or interviews with managers and employees.

"Self-reports" of training deficiencies are the most obvious clues to training needs. If training is institutionalized as a benefit, employees will be more willing to report training deficiencies and seek assistance. But the analyst must determine whether a reported performance deficiency calls for additional training or for an alternative human resource management solution.

Observations of employees performing tasks.

Observers can form impressions by listening to employees talk about their organization, jobs, supervisors, and working environment. Observers should note any obvious interpersonal difficulties or communication breakdowns among supervisors and employees, employees' complaints or inattentiveness to work, or supervisors' poor communication of direction. If such problems appear, observers should seek the underlying reasons for these difficulties. Observers should also collect data on employees' overt performance and skills.

Plans for new business strategies.

Organizational changes—such as plans to expand or reduce the number of employees, products, or services, to merge or reorganize the company, or to introduce new operating methods or technology—often demand new or improved human resource capabilities.

If the completed needs-identification forecast shows variance between actual and expected performance, it may be possible to document dollar losses caused by performance deficiencies. With evidence of dollar losses, a training department can take a proactive approach. Dollar-loss estimates are the part of cost-benefit analysis that communicates to managers in the way they understand best.

Training needs analysis

If organizational analysis and needs-identification forecasting reveal a performance deficiency, that doesn't necessarily mean that a training problem exists. Training is an appropriate intervention only if employees lack the vocational preparation needed to perform their job tasks adequately.

As a guideline, Mager and Pipe suggest that trainers consider this scenario: if an employee is at gunpoint and cannot find the skill and motivation to perform a task effectively, the performance deficiency is likely to be a training problem. While Mager and Pipe don't literally recommend this drastic test, it dramatizes the importance of motivation. Employees may know how to perform a task well but, for various reasons, may not demonstrate it on the job.

Training needs analysis has two parts:
■ job or task analysis to describe the tasks performed on a job regardless of who performs them
■ personal analysis to describe the specific knowledge, skills, and attitudes that a particular person needs to develop in order to perform the job adequately.

Many methods are available for gathering information for training needs analysis. They vary in the time and resources they require and in their suitability for particular purposes. For instance, work participation helps an analyst take the novice's point of view, which suits analysis of entry-level training. But that method isn't appropriate for analyzing complex high-level tasks.

Frequently, more than one method is used because the strengths of one can offset the weakness of another. For example, to avoid disrupting an observation, an analyst may withhold questions while employees complete a series of tasks, and then conduct a brief group interview.

The most common information-gathering methods for needs analysis:
■ questionnaires and checklists, possibly open-ended
■ individual or group interviews with job holders, supervisors, or subject-matter experts
■ observations, perhaps including "critical incident" observation to determine what behaviors lead to excellent or unsatisfactory (as opposed to adequate or average) performance
■ work participation by the analyst.

Tying learning and performance objectives to organizational objectives

Learning objectives are evaluated by measures of training's impact on participants' knowledge or skills as evidenced in the training setting. Performance objectives, on the other hand, are evaluated by assessment of training's impact on participants' on-the-job performance and by organizational measures.

For example, if observation indicates that warehouse

employees frequently misplace inventory because of inability to read shipping manifests and crate labels, a training program's learning objectives will be measures of improved reading skills. Success in achieving the performance objective is evaluated by subsequent assessment of workers' performance in correctly unloading and storing inventory.

When learning and performances objectives are both achieved, a training program's development process is validated. If neither set of objectives is achieved or if learning objectives are met but performance objectives are not, training's design and development are suspect.

In structuring learning and performance objectives, a trainer should set goals for employee improvements attainable through training. Learning and performance objectives should clearly set out competencies and performance expected. In most cases, training shouldn't begin until most employees agree that learning and performance goals are reasonable and attainable.

Finally, the objective-setting process links learning objectives and performance objectives to organizational goals and objectives. Because organizational goals are broad, it may be difficult to connect them directly to specific job knowledge or skill. As a rule, they can be linked to team or departmental objectives and, through them, connected to institutional goals. That is useful because it encourages integrity in training design and development and because participants may find additional motivation if they understand how their immediate efforts fit into overall organizational success.

Marketing training to management

If preliminary organizational analysis, needs-identification forecasting, and needs analysis lead a training manager to conclude that training is necessary to counteract current or anticipated deficiencies, then the manager will prepare a recommendation to management proposing and advocating development of a training program.

To build support, a training advocate must present a strategy and action plan. This plan is the centerpiece of an in-house marketing effort aimed at internal decision makers. The presentation should be well thought out and comprehensive but concise. The manager should give a copy of the written plan to each person whose approval of the proposal is needed.

The plan should include the following:

- a one-to-two page summary of the most compelling data from the analysis effort
- a list of the positive effects expected if the training is conducted and a list of problems or risks associated with not providing the training
- a recommendation for training program development, including a general description of the proposed program's content, timeframe, design, staff responsibilities, resource constraints, and estimated cost.

There are colloquial expressions that categorize training plans depending on the resources they require. Mager and Pipe identified these expressions in their "shoulda, oughta, wanna" (should, ought, want to) principle:

- "Shoulda" refers to a training plan that will cover essential needs.
- "Oughta" refers to a plan for a program that would go beyond the bare basics and would be better to have if enough money and other resources are available.
- "Wanna" is the comprehensive, sophisticated training program the organization would have in ideal circumstances.

In a few organizations, upper management might easily afford funding for the "wanna" option and might even see it as essential. But, in many organizations, recommending the "oughta" option necessitates a "hard sell" to unsympathetic managers. Recommending a "shoulda" option usually looks good in short-term costs and may be easier to achieve, but it could prove more costly in the long run if supplementary training is required later.

For the initial presentation, a training manager probably won't have enough information to prepare a complete operational budget, but an estimated cost figure should be possible. Before implementation becomes a reality, decision makers need to see an operational budget with firm dollar figures and an exact time schedule. So, once enough information has been gathered, a training manager should present (in person or in writing) a second, more detailed training plan.

A later evaluation step is re-examination of training's preliminary cost-benefit analysis for comparison of actual and projected returns. Demonstration of positive results will increase the likelihood that training will be seen and used as a strategic tool for managing performance.

Training as a Performance Management Tool

Training is one among numerous tools for performance management. Gherson and Moore analyzed training a tool for strategic intervention and found that it has these advantages:

- Comparatively low cost. In general, the costs of training resources are spread over time and are less expensive than those of other interventions.
- Discretion. When training is part of competitive repositioning, training activity usually isn't visible to outside competitors.
- Ease of adjustment. Training can be altered relatively quickly if evaluations reveal that it's not meeting organizational needs.
- Potential for improving internal relations. Training is an investment in present staff capabilities. Recognition of that can raise employee morale.

Training also has disadvantages:

- Time. Training takes time. In technical jobs, for instance, training may be a relatively slow-response intervention compared to hiring new, already trained staff.
- Lost production. Employees aren't producing while they're in training or traveling to and from a training site.
- Potential for ineffectiveness. Employees may respond negatively to training they perceive as ineffective or unconnected to rewards or career development.

Just as involving managers in program design and evaluation plans encourages their sense of ownership, providing written and oral reports about potential economic improvement from training will encourage managerial appreciation of training's value.

Such a report should describe the following:

- performance problems to be alleviated or eliminated
- accounting model that will be used
- method used to estimate the dollar value of training benefits
- method used to estimate training's return on investment (ROI)
- potential benefits to the overall organization.

Process and implementation evaluation

Process evaluations monitor training programs to ensure that they follow a rigorous design, development, and implementation process. They gather information to answer whether a program will follow (or is following) a structured format.

If a program isn't proceeding exactly as planned, there may be good reasons (for example, a better workbook came on the market at the last minute) or bad ones (for example, most participants are finding the sequence of material confusing).

Process evaluations of training designs are future-oriented. They represent crucial questions that training designers must ask after they have decided to initiate training but before and as they actually do it. Process evaluations involve deciding how to produce the necessary learning and behavior changes in participants. Of course, training designs are compromises between theory and practicality, and require some degree of alteration as they are installed and operated. On-the-spot control and revisions adapt programs to the environment and make them work.

Most training programs have many discrepancies between design and implementation. As they are noticed and considered, modifications may be made. Changes often consist of implementing different or additional controls (such as simpler instructions, repeated learning activities, more frequent reviews, smaller task divisions, or more focused materials). In some cases, the design itself or expectations for participant performance must change.

The goal is to note and assess program progress, look for discrepancies, make revisions, try the revised version, and then reobserve and reassess to see whether acceptable progress is now being made.

Evaluations after implemention may be formative or summative. Summative evaluations stress that training programs that don't "pay off" may be abandoned or replaced by other human resource interventions. Spencer suggests that training managers turn this into an advantage. For example, if a training manager uses evaluation to avoid continuing a program past its prime, the manager can also make a case for transferring the old program's funds to a desirable program previously considered unaffordable.

Rigor and practicality in evaluation design

Evaluation designs differ in rigor and practicality. Rigor relates to the quality and quantity of information the evaluation produces and to how well the information traces participant and organizational changes to the training program. Practicality relates to lessening the time and expense or increasing ease of conducting and reporting an evaluation.

Unlimited resources would permit an evaluator to answer all questions that a client might have about a training program. An absence of constraints also would permit an evaluator to select the most rigorous evaluation design to ensure the validity of evaluation conclusions. Instead, the only certainty of an evaluation is that constraints, both economic and political, will restrict the ability to answer clients' questions.

Constraints also affect the rigor of evaluation design. An evaluator is responsible for selecting the most rigorous evaluation design possible under existing constraints and for lobbying management to lift undue, organizationally self-imposed constraints.

Rigorous evaluation designs have the following characteristics:

- They collect data from many, perhaps all, participants.
- They collect data more than once, possibly many times.
- They evaluate more at the organizational results level.
- They employ quantitative data-collection methods.
- They are more expensive.
- They are more time-consuming.
- They yield formal reports.
- They are used for making decisions about program continuation or cutback (that is, they are summative).
- They are used when training's success is critical for safety or strategic business purposes.

Less rigorous practical evaluation designs, in contrast, have these traits:

- They collect detailed data from a small sample of participants.
- They are usually conducted to determine participants' reactions and learning gains.
- They use qualitative data-collection methods.
- They are less expensive.
- They take less time.
- They use informal reporting procedures.
- They are used to identify program strengths and weaknesses and to recommend areas for program improvement (that is, they are formative).
- They are used when training success is desirable, but not critical.
- They are used when a rigorous evaluation design is unjustified or impossible.

It might seem that all training programs should use rigorous evaluation designs, but Brandenburg and Smith cite an evaluation study that dispels that notion.

New England Telephone's management commissioned a $40,000 evaluation to determine the effectiveness of a training program for technicians. The evaluation used a rigorous evaluation design and multiple measures of performance. The evaluation design permitted the evaluator to make strong recommendations in favor of the training program after a year. But the

report arrived too late for the client to make a decision based on program findings. By then, everyone familiar with the evaluation had left the department that had financed it. Despite the evaluation's rigor, resources were wasted because practical considerations about reporting the findings were neglected.

Two evaluation design concepts are often referred to in discussions of specific evaluation designs:

- pre- and post-testing
- control groups.

Pre- and post-testing. Measures collected before the implementation are called pre-tests; those collected afterward are called post-tests. Tests are commonly thought of as written examinations, but virtually any measure that's meaningful for an organization can serve as a test. Test data can be collected from training participants directly or from organizational data.

Pre-tests are usually taken once. It's advisable, though, to pre-test at least twice. Here's why: people are often placed in a training program because of extremely high or low pre-test data. For example, management trainees may be slated for an accelerated training program on the basis of their high performance-appraisal ratings. Conversely, people may be placed in a remedial training program on the basis of their poor records.

The danger of selecting training participants on the basis of one test is that extreme results tend to drift toward the average in subsequent tests, whether there's intervention or not. This is because an individual's test results often can't go any higher or lower. This tendency for extremes to drift back toward the middle can make accelerated programs look unduly unsuccessful and remedial programs look unduly successful in post-tests.

Control groups. Pre- and post-measurement tests alone don't always prove that training is responsible for positive change in participants, because outside factors may be responsible. Using a control group improves an evaluator's ability to determine whether changes are attributable to training. A control group consists of employees who receive the same treatment as the training group except that they don't receive training. Control groups can range from a few people to a department or division.

Evaluation Designs

These descriptions of evaluation designs are ordered roughly by the degree of experimental control or scientific rigor that they provide.

Pre- and post-test control group design. The evaluator assigns employees to either an instructional or a control group. Only the instructional group receives training. Data are gathered from both groups through pre- and post-tests. If, upon completion of the program, the instructional group shows greater post-test performance gains than the control group, training is held responsible.

But this design is effective only when the instructional and control groups share nearly equivalent characteristics. With an instructional group of experienced workers and a control group of entry-level workers, for example, it would be difficult to determine whether training or job experience was the cause of change. Random selection of employees for each group usually results in groups with similar characteristics. But in practice, it's almost impossible to assign employees randomly to training programs.

Many HRD specialists commend this design's ability to quantify and neutralize the effects of other factors. Others criticize it as an inappropriate attempt to inject scientific rigor despite the impossibility of keeping outside factors (such as management styles and peer pressure) constant. Another problem is that organizations may be unwilling to withhold training from some employees just because the evaluator needs a control group.

Multiple baseline design. This design may be used if a training program is to be introduced to different parts of the organization at different times. Data are collected on each group of participants before and after the training. The second group serves as a control group for the first group, the third is the control group for the second, and so on. If each group shows marked improvement in performance as compared to its control group, success is indicated.

This design is more cost-effective than the pre- and post-test control group design, because it eliminates the need for a rigidly maintained control group. But more time is required to introduce the training at different places. Besides, training is often location-specific, so a training program may have to be altered somewhat to meet the specific needs of each group.

Time series design. This design examines effectiveness through repeated measures of performance before and after program implementation. If the measures show improvement after training, training's effectiveness is supported. In this design, the training group serves as its own control group.

The time series design is more practical and cost-effective than, but not as rigorous as, the control group or multiple baseline designs because it provides less experimental control. It can't prove that outside events (such as a new product line or marketing program) occurring simultaneously with training aren't partly or entirely responsible for improvements. Still, if several measures are taken and performance improves only after the introduction of a training program, a case is built for the effectiveness of the program.

Single group pre- and post-test design. This design is widely used because it offers a measure of comparison and is inexpensive, but the absence of a control group makes it difficult to attribute changes to training. This design is criticized for its lack of rigor, but it can be useful.

For example, Aetna Life Insurance Company instituted a claims processing training program for new employees and used the single group pre- and post-test design to evaluate its effectiveness. A pre-test revealed that new employees had little knowledge of claims processing. A post-test given shortly after the training program revealed significant learning gains, and it was unlikely that anything other than training was responsible for performance improvement.

One-time case study. This design trains participants without any pre-test measure, but a post-training measure is taken during the program or shortly afterward. When this design is used, there usually has been little thought given to evaluation design.

Still, this is the most convenient evaluation design, because it requires only one measure. It's also the least rigorous, because it makes no comparisons. The best that can be said for this design is that it's more informative than taking no measure. This design is useful when constraints permit no preliminary data collection, if the primary evaluation client simply wants data collected to confirm or invalidate perceptions about a training program, or when participants have no prior background in the subject matter.

For example, when IBM offered a program on beginning Japanese to their international sales representatives, no pre-test was given because of their lack of exposure to the language. In such a case, it's appropriate to take performance measures only after participants' initial exposure to training content.

Data-collection tools

Data-collection tools can be categorized as either quantitative or qualitative instruments.

Quantitative instruments:

- performance records and tests
- standardized questionnaires and survey instruments
- personnel assessment instruments.

The candor and accuracy of questionnaire and survey responses can be strengthened by assuring respondents of anonymity

Quantitative data have the following characteristics:

- relatively easy to measure and assign dollar values to
- objectively based
- use a common measure of performance
- credible to management

Qualitative instruments:

- interviews
- observations
- focus group meetings
- case studies

These instruments may be supplemented by other tools. For example, an evaluator may use a checklist of behaviors to guide an observation or may videotape a focus group meeting for additional review.

Qualitative data have the following characteristics:

- difficulty in standardizing
- subjectivity
- behaviorally oriented
- less credible to management.

At first, quantitative data appear superior to qualitative data. Yet quantitative data are more influenced by outside factors than qualitative data. Also, the appropriateness of quantitative or qualitative data depends on an evaluation's purposes, evaluation clients' questions, and overall evaluation design.

Quantitative data collection is more suitable when the following circumstances hold:

- Evaluation is to determine whether a training program should be continued or expanded.
- Evaluation's purpose is to identify a training program's economic impact on the organization.
- A rigorous evaluation design is used.
- Standardized data about a training program are needed.
- The specific training is crucial to strategic businesses goals or safety.
- Formal evaluation reports are required.

Qualitative data is more suitable under these circumstances:

- The focus of the evaluation is to improve the program, to discover unanticipated consequences of the evaluation itself or to determine how a training program's success varies at different sites or among categories of participants.
- Quantitative information needs to be augmented to provide depth and detail about a program's success.
- Quantitative data are unavailable (for example, employee agreements may prohibit the collection of certain quantitative data).

To select the best design and data-collection methods for use in the real world of politics, personalities, and methodological imperfections, evaluators must match appropriate data-collection methods with evaluation purposes. So, in practice, quantitative and qualitative data are often gathered together.

The collection of information from multiple data sources or through several methods is called triangulation. Although more expensive and time-consuming, triangulation increases the probability of confirming training's responsibility for changes in employee performance and organizational measures.

Evaluation Practices

ASTD's research revealed that the actual practice of evaluation doesn't often follow the strict recommendations of evaluation literature. This is largely explained by the fact that many training practitioners haven't found the literature's advice applicable or useful for their organizations.

But, as well-known author and management consultant Thomas J. Peters has said, "What gets measured gets done. . . . Even imperfect measures provide an accurate strategic indication of progress, or lack thereof." So practitioners have employed various practical evaluations.

Here's an overview of current evaluation practices among organizational leaders in training, telling how and why they subscribe to their various practices. The evaluation techniques and practices explored don't meet traditional academic notions of rigor, but do provide valuable information, are reproducible, and can be quickly and easily conducted. Most of the training managers that par-