3.5.1 Comparison of Training Techniques

Since so many different techniques are available, a training specialist must carefully evaluate the advantages and disadvantages of each technique to determine which is appropriate for a given situation. The selection of a training technique should be determined primarily by the objective of the training. For example, a lecture is ideal for disseminating a large amount of information to learners who are already motivated to receive it. But the lecture is not useful for changing attitudes or teaching new motor skills.

There are five major principles of learning.

1. *motivation:* active participation by the learner
2. *feedback:* knowledge of results
3. *stimuli:* meaningful organization of materials
4. *response:* practice and repetition
5. *stimulus response:* the most favorable conditions for transfer of training

An ideal training program should be consistent with each of these principles; however, all five principles may not be equally important, depending on the particular training activity. For example, all five principles may not be important in training a group of sales representatives about changes in the product mix for the coming year; a carefully organized lecture or video-recorded presentation of information may be adequate, and therefore, the four principles of active participation, knowledge of results, practice, and transfer of training would not be required.

PRINCIPLES OF LEARNING

	Motivation: *active participation of learner*	Reinforcement: *feedback or knowledge of results*	Stimulus: *meaningful organization of materials*	Response: *practice and repetition*	Stimulus-Response: *conditions most favorable for transfer*
ON-THE-JOB TECHNIQUES					
Job-instruction training	Yes	Sometimes	Yes	Yes	Yes
Apprentice training	Yes	Sometimes	Possibly	Sometimes	Yes
Internships & assistantships	Yes	Sometimes	Possibly	Sometimes	Yes
Job rotation	Yes	No	Possibly	Sometimes	Yes
Junior board	Yes	No	Possibly	Sometimes	Yes
Coaching	Yes	Yes	Sometimes	Sometimes	Yes
OFF-THE-JOB TECHNIQUES					
Independent study	Yes	No	Yes	Possibly	No
Corporate Universities	Yes	Sometimes	Sometimes	Sometimes	Sometimes

ON-THE-JOB TECHNIQUES

Job-instruction training	Yes	Sometimes	Yes	Yes	Yes
Apprentice training	Yes	Sometimes	Possibly	Sometimes	Yes
Internships & assistantships	Yes	Sometimes	Possibly	Sometimes	Yes
Job rotation	Yes	No	Possibly	Sometimes	Yes
Junior board	Yes	No	Possibly	Sometimes	Yes
Coaching	Yes	Yes	Sometimes	Sometimes	Yes

OFF-THE-JOB TECHNIQUES

Independent study	Yes	No	Yes	Possibly	No
Corporate Universities	Yes	Sometimes	Sometimes	Sometimes	Sometimes
Vestibule	Yes	Sometimes	Yes	Yes	Sometimes
Lecture	No	No	Yes	No	No
Case study	Yes	Sometimes	Sometimes	Sometimes	Sometimes
Conference or discussion	Yes	Sometimes	Sometimes	Sometimes	No
Role Playing	Yes	Sometimes	No	Sometimes	Sometimes
Simulation & business games	Yes	Sometimes	Sometimes	Sometimes	Sometimes
Programmed group exercise	Yes	Yes	Yes	Sometimes	Sometimes
Films & Television	No	No	Yes	No	No
Computer-based training	Yes	Yes	Yes	Sometimes	Sometimes
Teleconferencing	No	Sometimes	Sometimes	No	No

3.5.2 Criteria for Evaluating Training

The criteria for evaluating a training program should be determined by the training objectives. If a training program is designed to disseminate new information, then the effectiveness of the training should be determined by how well the information was disseminated. Most training programs attempt to accomplish several objectives, such as to change behavior, to provide new information, and to make an organization more effective. Consequently, most evaluations should utilize multiple criteria. Donald Kirkpatrick's model proposes four criteria for evaluating training programs: (1) reactions, (2) learning, (3) behavior, and (4) results. [1]

Reactions of Participants

How well did the trainees like the program? Since the trainees are the consumers of the product, how well they liked the program is important information for evaluating the program. Did they feel that the information was worthwhile? Was it presented in a way that was meaningful and interesting? Did they believe the program was well-prepared and carefully organized?

These reactions can be obtained by having participants complete a simple questionnaire at the end of a training program. Although care is needed in designing the questionnaire to minimize response biases, the process of assessing participant reactions is quite easy. In fact, the ease of assessing participant reactions probably explains why this is the only information that is typically obtained regarding the effectiveness of most programs.

Learning New Skills and Competencies

To what extent did the trainees learn and retain the information presented in the training program? Most training programs present knowledge that the participants are expected to learn. How well the participants learned and retained this information should be determined through the use of evaluations similar to the examinations schools and colleges use to see how much students have learned. These tests can be based on traditional evaluation methods: true-false questions, multiple-choice questions, and essay exams.

Although many trainers are apprehensive about administering an exam, the participants usually respond more favorably than the trainers anticipate. In fact, when participants know in advance that they will be examined on the content of a training program, their attentiveness and retention usually increase.

Behavioral Changes

To what extent did the behavior of the trainees change as a result of the program? For example, if a time management seminar is expected to help trainees use their time more effectively by setting priorities and by developing "to-do" lists, did the participants actually change their behavior as a result of the training?

Changes in behavior can be assessed by two major methods. First, individuals can assess changes in their own behavior with a simple self-report questionnaire. Second, changes in behavior can be assessed by the observation of others. Supervisors, for example, can be assessed by their superiors or their subordinates. Several useful profiles have been developed to evaluate changes in a supervisor as perceived by subordinates. [2]

Measurable Results

What final results were achieved from training—reduced costs, reduced turnover, improved productivity, or greater profitability? The primary reason for presenting most training programs is to improve organizational performance. The question, then, is whether the organization is actually performing more effectively or efficiently as a result of the training effort. After a training program on safety, is the frequency or severity of accidents declining? As a result of a new-employee orientation program, is the percentage of employees who leave within the first 60 days declining? As a result of a supervisory training program, are the departments of the trained supervisors obtaining higher levels of productivity, profitability, or morale?

Similarly, Jack and Patricia Phillips propose that organizations develop evaluation criteria from the following levels of training objectives:

1. reaction and perceived value
2. learning and confidence
3. application and implementation
4. impact and consequences
5. return on investment

This approach particularly emphasizes the importance of focusing on **application objectives**, which explain how the skills gained in a training program will be applied by the trainees on the job, and **impact objectives**, which describe how the application of the new skills by the trainees will impact the organization. [3]

The ultimate measure of results for a training activity is return on investment. How much did the training cost in both direct and indirect costs and what are the financial returns from it? This is called a **utility analysis** and a formula for measuring the utility of a training program has been proposed that combines the performance improvements generated by the training, the economic value of this improvement, the number of people trained, how long they remain with the company, and the direct and indirect costs of the training. [4]

Previous Next

3.5.3 Sources of Data

Training directors have a large variety of different variables they can use to evaluate training effectiveness. The ideal circumstance is to have data that specifically measure each variable being considered. For example, to know how a training program influences job satisfaction, direct measures of job satisfaction would provide the most specific data. Unfortunately, direct job satisfaction measures are often not available, so trainers may be forced to rely on indirect measures, such as absenteeism, turnover, or grievance rates. These variables are related to job satisfaction, but only indirectly.

Archive Data

Archive data refer to the kinds of information contained in the human resource files. To comply with federal legislation, companies are required to maintain extensive information in their files and this information can be effectively used to assess the value of training.

This information is referred to as **unobtrusive measures** because the nature of the data and the way they are obtained do not influence how employees behave. The advantages of using archive data are that they already exist without having to be collected and they may cover a fairly lengthy historical period, thereby providing a long-term perspective. Some of the types of information that are frequently available in the human resource files include accident records, affirmative action records, application forms, attendance records, attitude surveys, disciplinary records, exit interviews, grievance reports, insurance records, interview records, job specifications, medical records, payroll records, performance evaluations, salary-increase records, test scores, and turnover statistics.

Questionnaires and Surveys

A carefully developed questionnaire can be administered to a large sample of employees, and the information can be conveniently analyzed and interpreted. The popularity of questionnaire data for research purposes is largely attributed to the ease of administration and convenience of analysis. However, the disadvantages of questionnaires are that they are time consuming to develop and the data can only answer the specific questions in the questionnaire. If all the relevant issues are not covered in the questionnaire, the information obtained will be inadequate. Suggestions for designing questionnaires are presented in Unit 5: Employee and Labor Relations.

Performance Tests

Training that focuses on helping employees learn specific skills can be evaluated by asking the employees to demonstrate the skills in a controlled setting. A five-minute timed typing test, for example, is a useful performance test to check the progress of a typing trainee. The process of developing a performance test involves identifying a defined and central segment of a job that can be performed under controlled circumstances. Measures of both quantity and quality of performance should be recorded. Performance tests are especially relevant to jobs that produce a physical product that can be counted.

Interviews

Interviews are an effective way to learn how employees feel about training and what they think should be changed. The advantage of an interview is that it is a rich source of information in the sense that it provides information about a broad range of topics. The major disadvantage of interviews is that they are time consuming and only a limited number of people can be interviewed.

Simulations

Simulations are similar to performance tests, however, they are not necessarily composed of a segment of an actual job. Simulations provide an opportunity for trainees to practice what they have learned in a defined setting where their performance can be observed. Computer simulations allow trainees to diagnose problems and make business decisions in an environment that has been constructed by the trainer. Many decisions can be made in a short time and the implications of the decisions can be evaluated without having to suffer the actual consequences of a real-life setting.

Ratings/Checklists

Ratings and checklists are typically used to evaluate training that focuses on changing observable behaviors. Ratings and checklists can be constructed from the task analysis that was completed as part of the training needs assessment. Trainees are rated on each of the important behaviors they are expected to learn as a result of training. Checklists are used to record how often trainees make the appropriate responses they have learned in training.

Critical Incidents

Critical incidents are essay descriptions of the times when a trainee does something especially good or especially bad. These descriptions are typically accumulated in an employee's personal file. They are often used to evaluate an employee's performance, and they can also be used to evaluate the success of a training program.

Observations

Some training programs can only be assessed by observing the behavior of employees to discover whether they are behaving differently after training or if the training has solved the problems. The disadvantage of observations is that they are an intrusive type of measure. The process of observing an employee's performance often causes the employee to behave differently. (This phenomenon is called the "**Hawthorne Effect**" after the Hawthorne studies, 1924-1934.) Most employees tend to work faster when they know they are being observed and their productivity is being measured. However, other employees intentionally alter their performance to create meaningless data.

Performance Appraisals

Performance evaluations are supposed to measure performance; therefore, they should provide valuable information to assess the value of a training program. Those who have been trained should receive higher performance ratings.

Previous Next

3.5.4 Research Methods for Evaluation

Numerous research designs may be used to evaluate training programs; seven are discussed in this section. The first two represent poor research designs, even though they frequently are used to evaluate training programs. The next five are excellent designs that eliminate most of the competing explanations for the results.[1] These designs are illustrated below. The symbols used in these exhibits are those used in research designs for behavioral science studies. X refers to the experimental treatment, or, as used here, to the training and development program. O represents an observation, or here, an evaluation of the training program. The evaluation could use any or all of the criteria discussed previously: reactions, learning, behavior, and results.

1. Post-test-only design (case study)

X O

2. Pretest-post-test comparison

O X O

3. Pretest-post-test control group design

R O X O
R O O

4. Post-test-only control group design

R X O
R O

5. Solomon four-group design

R O X O
R O O
R X O
R O

6. Time-series design

O_1 O_2 O_3 O_4 X O_5 O_6 O_7 O_8

7. Separate sample pretest-post-test control group design

Group 1 R O X
R X O
Group 2 R O
R O

KEY

X = Experimental treatment or training program

O = Observation or measurement

R = Random assignment of participants to group/training condition

Experimental Designs and Use of Control Groups

The adequacy of an experimental design depends largely on its ability to control extraneous influences and allow a researcher to properly assess the effect of the training. An essential element in evaluating any training program is having a standard for comparing the results. The two most convenient ways of providing a basis for comparison are to either use a control group or to use the experimental group as its own control.

1. Post-test-only design (case study)

X O

1. **Post-test-only Design:** Many training programs are evaluated on the basis of a simple post-test-only design, also known as the case-study design, in which data are collected after the training program has been completed. In a simple post-test–only design, or after-only design, the training program is presented and then evaluated.

   This is a poor research design for evaluating a training program because no standards of comparison are available for interpreting the observations. An evaluation may show that the participants are exhibiting the proper kinds of behavior, but they may have been displaying the appropriate behavior before the training program was presented. When nothing is known about behavior before training, a case study cannot show that a training program created a change in behavior.

   Case-study designs are not entirely worthless. Although they are poor evaluation designs, they sometimes provide insightful information that helps in evaluating a program. The usefulness of a case study, however, depends on the skills of the observer. An astute observer may be able to evaluate what has occurred in a training program, assess the program's effectiveness, and prescribe changes that need to occur.

2. Pretest-post-test comparison

O X O

2. **Pretest–Post-test Comparisons:** To provide a comparison for the post-test observations, some evaluators conduct a pretest evaluation. Although this design compensates for one of the problems of the case-study design, it is still an inadequate research design.

   Improvements in performance or learning scores from the pretest to the post-test may or may not be attributed to the training program. For example, if a group of new trainees is exposed to one hour of training every day during their first week, improvements in their scores from the first of the week to the end of the week are typically attributed to the effects of training. Although the results may make the program appear successful, several competing explanations may also explain the change in the scores. Three of the most relevant competing explanations are referred to in experimental research as history, maturation, and the sensitizing effects of the pretest.

   History refers to the historical events that transpire between the pretest and the post-test. While the training program is being presented, many other things are also happening that may influence the scores regardless of what occurs within the training program.

   Maturation refers to the development and growth that occurs within the trainees—processes that occur regardless of the training program. A group of new employees may show a significant improvement in their scores at the end of training compared to their scores at the onset of training because of the time they have spent on the job rather than the time they have spent in training.

   Sensitizing effects of the pretest refers to the effects that pretesting might have on the trainees. Some attitudes and behaviors can be influenced simply by evaluating them. For example, studies in attitude change have found that asking people their opinions about racial prejudice influenced what they said on subsequent evaluations, even though nothing else happened between the pretest and the post-test.

   A lack of consistency or accuracy in the measuring instrument is referred to as **instrument decay**. Changes in the pretest and post-test scores are sometimes caused by measuring errors. Inaccuracy is especially troublesome when researchers use subjective, self-reported data and the three kinds of changes that can occur are called alpha, beta, and gamma. *Alpha change* occurs when the observed differences between the pretest and post-test measures represent a true change in the person providing the data as a result of the training. *Beta change* occurs when the instrument being used is recalibrated, such as when the person providing the data adopts a different reference group and evaluates everything more leniently or harshly than before. For example, if a training program caused trainees to change their reference group for evaluating themselves from average performers to excellent performers, then they might report lower scores when evaluating themselves even though they improved as a result of training. *Gamma change* occurs when the construct being evaluated is redefined, such as a sales training program that redefines the meaning of "quality customer service." Customer service scores could change because customer service now means something different.

   A method that has been proposed to improve the usefulness of the pretest–post-test comparison design is called the **internal referencing strategy (IRS)**. This method tries to detect the presence of competing explanations by using a combination of relevant and irrelevant questions to evaluate how well training increased learning. It is only useful for evaluating the criteria of learning. An effective training program should improve the knowledge of trainees, and they should be able to answer questions relevant to the content of the training. However, they would not be expected to answer irrelevant questions. These irrelevant questions provide an internal reference for assessing the benefit of the training. If there are changes in the scores of irrelevant questions then one would have good reason to be suspicious about interpreting changes in the scores of relevant questions. But, when the only improvements are on the scores of relevant questions, one has greater confidence in ascribing the improvement to the training.

3. Pretest-post-test control group design

R O X O

R O O

R O O

3. **Pretest–Post-test Control-Group Design:** This research design involves two groups: an experimental group and a control group. The R represents randomization, meaning that the participants are randomly assigned to the two groups. Since the individuals are randomly assigned to groups, the pretest observations for both groups should be essentially the same. If the training program produces a significant effect, the post-test observations for the experimental group should be significantly greater than the pretests of either group and the post-test of the control group.

   This research design is excellent for evaluating training programs and for determining whether the changes can be attributed to the training per se. The difficulty that this design typically presents for evaluating training programs is the problem of randomly assigning people to the two groups. This obstacle can be overcome, however, by randomly selecting certain individuals for training and by holding over the other individuals for a second training program. In many organizations, an entire staff cannot be trained at the same time, and therefore, having successive groups is sometimes a necessity.

4. Post-test-only control group design

R X O
R O

4. **Post-test-only Control-Group Design:** This is an excellent experimental design because it allows a researcher to eliminate most of the competing explanations for the experimental results. This design involves randomly dividing the sample of people into two groups and presenting the training to only one group. Neither of the groups receives a pretest. After one group is trained, both groups are evaluated to see if the training had a significant impact.

   This design is ideal when the training and the post-test measurement can be administered to employees as a single, natural package, or where a pretest would be awkward. Suppose you wanted to know whether a different set of training instructions would motivate employees to provide more complete data on a health survey. A random sample of employees could be identified and given a modified form with the revised instructions. If the data they provided were significantly more complete than the group that randomly received the original instructions, then a researcher could safely conclude that the revised instructions had a significant impact.

5. Solomon four-group design

R O X O
R O O
R X O
R O

5. **Solomon Four-Group Design:** The **Solomon four-group design** is the most highly recommended research design for evaluating training programs. It requires four groups to which the participants have been randomly assigned. Only the first two groups are pretested. One of the pretested groups and one of the non-pretested groups participate in the training program. All four groups are observed after the training of the two groups has been completed.

   This research design successfully eliminates alternative explanations for why the post-test scores of the experimental groups might be considerably higher than their pretest scores. If the scores are different, the improvement in performance can be attributed directly to the training program. Since this design appears complicated, many training specialists fail to seriously consider its usefulness. The difficulty, again, is in randomly assigning participants to the four groups. However, this problem can be conveniently overcome if the research design for evaluation is included in the initial design of the training program.

Quasi-Experimental Designs

Quasi-experimental designs are experiments that occur in a natural setting where researchers do not have complete control on the experimental setting. Quasi-experimental designs are recommended when better designs are not feasible, such as when subjects cannot be randomly assigned to groups or when you have difficulty scheduling when and to whom the experimental change, such as a training program, will be administered. Although quasi-experimental designs are not as conclusive as other experimental designs, they are, nevertheless, capable of indicating whether an experimental change, such as a training program, has a significant impact.

The time-series design and the separate sample pretest–post-test control group design are called **quasi-experimental designs** because they occur in a natural setting where you can introduce an experimental change and collect data, but you do not have complete control over the experimental setting.

5. **Solomon Four-Group Design:** The **Solomon four-group design** is the most highly recommended research design for evaluating training programs. It requires four groups to which the participants have been randomly assigned. Only the first two groups are pretested. One of the pretested groups and one of the non-pretested groups participate in the training program. All four groups are observed after the training of the two groups has been completed.

   This research design successfully eliminates alternative explanations for why the post-test scores of the experimental groups might be considerably higher than their pretest scores. If the scores are different, the improvement in performance can be attributed directly to the training program. Since this design appears complicated, many training specialists fail to seriously consider its usefulness. The difficulty, again, is in randomly assigning participants to the four groups. However, this problem can be conveniently overcome if the research design for evaluation is included in the initial design of the training program.

Quasi-Experimental Designs

Quasi-experimental designs are experiments that occur in a natural setting where researchers do not have complete control on the experimental setting. Quasi-experimental designs are recommended when better designs are not feasible, such as when subjects cannot be randomly assigned to groups or when you have difficulty scheduling when and to whom the experimental change, such as a training program, will be administered. Although quasi-experimental designs are not as conclusive as other experimental designs, they are, nevertheless, capable of indicating whether an experimental change, such as a training program, has a significant impact.

The time-series design and the separate sample pretest–post-test control group design are called **quasi-experimental designs** because they occur in a natural setting where you can introduce an experimental change and collect data, but you do not have complete control over the experimental setting.

6. Time-series design

$$O_1 \; O_2 \; O_3 \; O_4 \; X \; O_5 \; O_6 \; O_7 \; O_8$$

- **Time-series Design:** A time-series design, also called a **multiple-baseline design**, consists of periodically measuring something and introducing an experimental change during this series of measurements. If the measures that are taken after the experimental change are significantly different than the earlier measures, there is some reason to believe that the experimental change had a significant impact.

  An illustration of a time-series design would be using monthly incidence reports to examine the effectiveness of a safety training program. If the average incidence rates before the training are significantly higher than after training, there is good reason to believe the training was effective.

  The major disadvantage of a time-series design is that in the absence of a control group, it is not possible to know for certain that the change did not result from some outside historical factor that occurred at the same time. For example, the reduced incidence rates could have resulted from an OSHA inspection that occurred the same month the safety training was presented.

7. Separate sample pretest-post-test control group design

Group 1	R	O	X	
	R		X	O
Group 2	R	O		
	R			O

- **Separate-Sample Pretest–Post-test Control-Group Design:** Sometimes employees work together in groups and it is not possible to introduce an experimental change only to randomly selected members of a group. Employees can be measured or observed individually, but the experimental change must be administered to the entire group. In this situation, the separate sample pretest–post-test control group design is an effective and convenient research design. This design utilizes two existing groups that may not be equivalent. Only one group receives the experimental change. A randomly selected part of each group is pretested and the others are post-tested. If the pretest–post-test differences for the group that received the experimental treatment are different than the pretest–post-test change in the other group, there is some reason to believe that the experimental change had a significant effect. This design would be useful for studying such things as the effects of music on productivity, the effects of a safety-training film on attitudes toward OSHA, or the effects of a new communication system on productivity.

Previous Next