Chapter VII

Evaluation Of Training
Even though Evaluation is listed as the last phase of the Instructional Systems Design (ISD) model, evaluation actually takes place during all of the phases. Evaluation is used during the ISD process to evaluate the training program itself. This type of evaluation is referred to as internal evaluation. Internal evaluations are done during the ISD process and are associated with the analysis, design, development and implementation stages. Each of these four separate steps is tied directly to each other through the elements of the evaluation process. Needs assessment and evaluation of training especially work hand-in-hand. They work together as parts of a continuous feedback cycle to help training planners determine content. Evaluation is also done at the end of the training to determine whether the learners have mastered the objectives of the training program. This type of evaluation is referred to as external evaluation.

7.1 Evaluation in practice

Change is everywhere. It is inescapable. Hardly a day goes by without news of the new world economy or the shift from a production to a service orientation. Indeed, change is necessary to survive in an uncertain world. Also, today’s business environment is highly competitive. Because of the sweeping effects of change and competition, a great deal of interest has been placed on higher education and lifelong learning. Consequently, business is turning to training in order to cut costs and increase productivity among employees.

However, in the rush to train and educate people, many organizations have failed to treat the evaluation of such training as a priority. At best, the evaluation of training has been a perfunctory task with little analysis and usefulness. Yet evaluating the effectiveness of costly training efforts is paramount to the success of any program.

There is ample evidence that evaluation continues to be one of the most vexing problems facing the training fraternity. Catanello and Kirkpatrick’s 1968 survey of 110 industrial organizations evaluating training revealed that very few were assessing anything other than trainee reactions.

---

Looking at similar data and the emphasis in much of the literature, one wonders if there has been much change in 20 years. Galagan(1983) and Del Gaizo(1984) both refer to a survey of Training and Development Journal readers in which 30% of the respondents identified evaluation of training as the most difficult part of their job. Easterby-Smith and Tanton (1985) report on their British survey involving HRD practitioners in fifteen organizations. In virtually every case the only form of evaluation being done was end-of-course trainee reactions, and the data so obtained seldom used.3

A 1985 survey by Marguerite Foxon of a sample of Public Service and private company trainers in Sydney to determine both their attitude to evaluation and what was being carried out by them in practice, also has similar findings.4 All expressed a firm belief in the principle of evaluation, and all administered end-of-course forms of varying degrees of complexity to gauge trainee reactions to the instructors, content, and facilities. But 75% admitted that was as far as their evaluation went, mainly because they did not know what else to do. As Easterby-Smith and Tanton (1985) observe, much current practice is only a ritual, and in many cases the evaluation that counts is done before the course is ever given; post-course data merely confirm prior judgments that the training is satisfactory.

In the minds of many practitioners evaluation is viewed as a problem rather than a solution, and an end rather than a means. Where evaluation of programs is being undertaken it is often a ‘seat of the pants’ approach and very limited in its scope. Overawed by quantitative measurement techniques, and lacking both the budget and the time as well as the required expertise for comprehensive evaluations, trainers often revert to checking in the only way they know - post-course reactions - to reassure themselves the training is satisfactory.

If the literature is a reflection of general practice, it can be assumed that many practitioners do not understand what the term evaluation encompasses, what its essential features are, and what purpose it should serve. Many

practitioners regard the development and delivery of training courses as their primary concern, and evaluation something of an afterthought. Adopting the premise that no news is good news, many practitioners still avoid the evaluation issue. Preferring to “remain in the dark”, and worried that evaluation will only confirm their worst fears (since they have no other alternative to offer management if the current program is shown to be educationally ineffective), they choose to settle for a non-threatening survey of trainee reactions.

Many training professionals agree that evaluation is important to successful training, but few conduct complete and thorough evaluations. Some trainers gather data for evaluation but do not analyze those data for trends or use them to improve existing training programs. Such an oversight can be costly, especially in light of the billions of dollars that have been spent and will continue to be spent annually on training efforts as a result of the demographic, economic, and technological changes.

Typically evaluation is an afterthought or not done at all. Too often, training departments have little or no idea how their training relates to the business objectives of the company. This could be due partially to trainers’ lack of measurement and evaluation skills, which result in measurements that are not valid, reliable or even useful to the management of the company.¹

There are very few subjects that cause more fear and perspiration among training professionals than quantifying where and how training is contributing to the organization.² This is often a problem because the costs of training are expressed in dollars, but the benefits are often subjective and difficult to quantify. Another difficulty is the Training Manager’s reliance on others for accurate information. Line and functional managers are busy people and often don’t have the time to keep thorough records that adequately quantify training’s benefits. Other problems encountered in calculating return on investment include the fact that costs can be seen up front and benefits accrue slowly over time. Also, there is a mistaken belief among some people that training does not represent a profit center and


records don't need to be kept. Nevertheless, to justify a training program's existence, its effectiveness must be continuously calculated.

Although for many years trainers have attempted to evaluate their programs, until quite recently, there has not been a bona fide effort to use valid and reliable methods to conduct such evaluations.

But if the results of a recent survey by ASTD are to be believed, the scenario seems to be changing for the better. The National HRD Executive Survey 1997 conducted by ASTD involving panel of nearly 300 HRD executives and managers reveals a fair amount of measurement and evaluation activity in organizations today. It reports that 90 percent of organizations evaluate at least some of their training courses.¹

The following quotation from Daniel M. Goodacre is most appropriate as a conclusion here:²

"Managers needless to say, expect their manufacturing and sales departments to yield a good return and will go to great lengths to find out whether they have done so. When it comes to training, however, they may expect the return – but rarely do they make a like effort to measure the actual results. Fortunately for those in charge of training programs, this philanthropic attitude has come to be taken for granted. There is certainly no guarantee, however, that it will continue, and training directors may be well advised to take the initiative and evaluate their programs before the day of reckoning arrives."

7.2 Towards a definition

A variety of definitions can be found in the literature, many of them stipulative, and the inconsistencies in the use of the terminology has "muddied the waters" of training evaluation a great deal, affecting the success of evaluation efforts.³

The majority of writers tend to view it as the gathering of information in order to make a value judgment about the program, such as necessary

changes or the possible cessation of the program. Williams (1976) defines evaluation as the assessment of value or worth.\(^1\) Harper & Bell (1982) refer to the planned collection, collation and analysis of information to enable judgments about value and worth.\(^2\) However, as Williams (1976) observes, value is a rather vague concept, and this has contributed to the different interpretations of the term evaluation.

Some definitions (Goldstein, 1978; Siedman, 1979; Snyder et al., 1980) focus on the determination of program effectiveness. Several definitions emphasize evaluation as a basis on which to determine program improvements (Rackham, 1973; Smith, 1980; Morris, 1984; Foxon, 1986; Tyson & Birnbrauer, 1985).\(^3\)

Evaluation is also confused by some with the terms measurement and assessment. Evaluation involves description and judgment; measurement and/or assessment provides the data on which to base the evaluation. This confusion of terms is most obvious when considering the use of “evaluation” and “validation”. While most American writers do not see validation as separate from evaluation, there are still British writers who appear to draw the distinction (Hawes & Bailey, 1985; Rae, 1985). Rae regards assessment as the measuring of the practical results of the training in the work environment; this, with validation of the training and training method, comprises evaluation. It must therefore be borne in mind that the terms “validation” and “evaluation”, often used in HRD literature, do not always mean one and the same thing.\(^4\)

The literature reveals a broad range of definitions and considerable confusion in the use of associated terms, and it would seem that HRD practitioners have yet to give serious consideration to what the term evaluation actually means.

---


7.3 Purpose of evaluation

As well as the lack of agreed-on definition of evaluation, there is an equally broad range of opinions as to the purpose of evaluation. More than 20% of the writers neither describe nor imply a purpose for the evaluation. Where purposes are outlined, they provide some telling insights. For example, 15% see the purpose of evaluation as justifying the training department’s existence and providing evidence of cost benefit to the organization. The majority of these articles surfaced in the period 1980-83, and clearly reflect the preoccupation of many practitioners with keeping their jobs during an economic downturn and resultant HRD budget cuts.¹

Despite the regular reference in the literature to Kirkpatrick’s four stage model, only a small percentage consider the purpose of evaluation specifically in these terms.

Bramley and Newby (1984) identify five main purposes of evaluation:²

1. Feedback - linking learning outcomes to objectives, and providing a form of quality control
2. Control - making links from training to organizational activities, and to consider cost effectiveness
3. Research - determining relationships between learning, training, transfer of training to the job
4. Intervention - in which the results of the evaluation influence the context in which it is occurring
5. Power games - manipulating evaluative data for organizational politics

According to Clive Shepard, there are many good reasons for measuring the success of a training program:³

- To validate training as a business tool: Training is one of many actions that an organization can take to improve its performance and profitability. Only if training is properly evaluated can it be compared

---

³ Shepard, Clive, “Evaluating online learning”, www.fustrakconsulting.com
against these other methods and you can expect it, therefore, to be selected either in preference to or in combination with other methods.

- **To justify the costs incurred in training**: We all know that when money is tight, training budgets are amongst the first to be sacrificed. Only by thorough, quantitative analysis can training departments make the case necessary to resist these cuts.

- **To help improve the design of training**: Training programs should be continuously improved to provide better value and increased benefits for an organization. Without formal evaluation, the basis for changes can only be subjective.

- **To help in selecting training methods**: These days there are many alternative approaches available to training departments, including a variety of classroom, on-job and self-study methods. Using comparative evaluation techniques, organizations can make rational decisions about the methods to employ.

According to Phillips (1983), the purposes and uses of evaluation are to improve the Human Resources Development process and to decide whether or not to continue this process. He states that evaluation should:

1. Determine whether or not a program is accomplishing its objectives
2. Identify the strengths and weaknesses in a Human Resources Program.
3. Determine the cost/benefit ratio of an HRD program
4. Decide who should participate in future programs
5. Identify which participants benefited most or least from the program
6. Reinforce major points made to the participants
7. Gather data to assist in marketing future programs
8. Determine if the program was appropriate (Phillips, 1983)

Training and development staff are becoming more and more accountable for the effectiveness of their programs. With more emphasis on return on investment, companies are asking what is the value of training. When times were good, and money plentiful, training and development programs flourished. Many companies were happy to train employees, because it both strengthened the organization and served as a retention tool. Since 1998, money spent on learning and development as a retention tool – from leadership-development training to management-skills seminars – has increased 15 percent, according to Hackett Benchmarking and Research, a

---

1 Evaluation of Training By Jennifer Stone and Virginia Watson

354
firm that tracks best practices, in HR, finance, and other areas of “knowledge work”. But can that level of investment continue during the economic downturn? Jack J. Phillips, an expert in measurement evaluation, and return on investment says that HR has to demonstrate the link between training and business results.¹

7.4 Principles of evaluation

The basic approach to evaluation should be to determine the extent to which the training program has met the objectives identified prior to training. Planning for the evaluation should begin at the same time that planning for the training program begins. If the goals of the program are clearly stated as specific objectives, the appropriate evaluation method can be implemented at the same time as the program.²

Evaluation of the training program must be based on the following principles:³

1. Evaluation specialist must be clear about the goals and purposes of evaluation
2. Evaluation must be continuous
3. Evaluation must be specific
4. Evaluation must provide the means and focus for trainers to be able to appraise themselves, their practices, and their products.
5. Evaluation must be based on objective methods and standards
6. Realistic target dates must be set for each phase of the evaluation process. A sense of urgency must be developed, but deadlines that are unreasonably high will result in poor evaluation.

Decision points in planning training evaluation⁴

Before beginning an evaluation several decisions needs to be taken:

1. Should an evaluation be done – is it worth the time and effort?
2. What is the purpose of evaluation?

¹ Lachnit, Carroll, “Training proves its worth”, www.workforce.com
3. What will be measured?
4. How comprehensive will the evaluation be?
5. Who has the authority and responsibility?
6. What are the sources of data?
7. How will the data be collected and compiled?
8. How will data be analyzed and reported?

These decision points are intended to increase awareness of and interest in the evaluation of training, to improve planning skills and to encourage more systematic evaluation of training.

Areas of evaluation

According to Peter Warr and Rad Kham, the evaluation exercise should be carried out covering the following aspects of a training program:¹

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Area of evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Context</td>
<td>Training infrastructure facilities</td>
</tr>
<tr>
<td>Input</td>
<td>Course design, objectives, course contents, handouts, objectives of specific session</td>
</tr>
<tr>
<td>Process</td>
<td>Methodology, adopted learning climate, instructor who delivered the topic</td>
</tr>
<tr>
<td>Outcome</td>
<td>Increase in knowledge/skill and attitudinal change.</td>
</tr>
</tbody>
</table>

Table: 7.1 Areas of evaluation

7.5 How is evaluation done?

As with definitions and purposes, there is great variety in the evaluation models and techniques proposed. In some cases it is very difficult to separate the techniques from the ‘model’- the writers are actually presenting an evaluation approach using a specific technique rather than a model.

If the literature reviewed is a reliable guide, Kirkpatrick’s four-stage model of evaluation is the one most widely known and used by trainers. Perhaps

this is because it is one of the few training-specific models, and is also easily understood.

7.6 Kirkpatrick’s Model

Donald Kirkpatrick set forth his four-level approach to the evaluation of training in a series of articles appearing in the journal of what was then known as the American Society of Training Directors. The first of these four seminal articles was published in November of 1959. The remaining three articles were published in the succeeding three months, with the fourth and final article appearing in February of 1960.

It is one of the most comprehensive and widely referenced models of evaluation. Since the introduction of Kirkpatrick’s model, a lot of things have happened in writing about and teaching evaluation. But the content has remained basically the same. Despite the fact that the Kirkpatrick model is now nearly 40 years old, its elegant simplicity has caused it to be the most widely used methods of evaluating training programs. ASTD’s survey, which reports feedback from almost 300 HRD executives and managers, revealed that 67% of organizations that conduct evaluations use the Kirkpatrick model.\(^1\)

The model is simple and practical. Its chief purpose is to clarify the meaning of evaluation and offer guidelines on how to get started and proceed. The model maintains that there are four levels to measure the quality or effectiveness of a training course. Kirkpatrick suggests that these four levels of evaluation may form a hierarchy. Accordingly, change farther up the hierarchy of outcomes is unlikely unless change has occurred lower in the hierarchy. That is, if no learning has occurred, it is unlikely that on-the-job behaviour will change. If behaviour does not change, it is unlikely that measurable improvements in results will be observed.

Each level has its advantages and disadvantages. It is important to plan the evaluation process, as the training is being planned. It is important to consider all levels at the outset, even though only one or two levels may be used ultimately. If the time, money, and expertise are available, it’s

---

important to proceed through all four levels without skipping any.\textsuperscript{1} Moving down the column, the table presents these levels, in order, from simple and inexpensive to complex and costly.

<table>
<thead>
<tr>
<th>Levels</th>
<th>Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 1</td>
<td>Reaction</td>
</tr>
<tr>
<td></td>
<td>Does the trainee like the course? (Trainee reaction to the course)</td>
</tr>
<tr>
<td>Level 2</td>
<td>Learning</td>
</tr>
<tr>
<td></td>
<td>Did trainees learn what was based on the course objectives? (achievement of knowledge, skills or attitudes)</td>
</tr>
<tr>
<td>Level 3</td>
<td>Behavior</td>
</tr>
<tr>
<td></td>
<td>Are the learners applying what they learned? (Trainee behavior changes on the job)</td>
</tr>
<tr>
<td>Level 4</td>
<td>Results</td>
</tr>
<tr>
<td></td>
<td>Did the training produce measurable results? (Ties training to the company’s bottom line)</td>
</tr>
</tbody>
</table>

Table: 7.2 The four levels of evaluation

7.6.1 Level 1: Reaction Evaluation

Reaction is the term that Kirkpatrick uses to refer to how well the participants liked a particular training program. Evaluation of participants’ reactions consists of measuring their feelings; it does not include a measure of actual learning.

Level 1 (Reaction) is the most commonly used method of evaluation, probably because it is the easiest type of evaluation to administer and evaluate. This level produces what some people dub the “smile sheet”, which measures how well the students like the training.

\textsuperscript{1} Kirkpatrick, Donald L., “Great Ideas Revisited”, Training & Development, Jan. ’96, Vol.50, No.1, pp 54-59
It is important to determine how people feel about the programs they attend. Reactions are important because, if learners react negatively to your courses, they are less likely to transfer what they learned to their work and more likely to give bad reports to their peers, leading in turn to lower numbers.\(^1\) Decisions by top management are frequently made on the basis of one or two comments they receive from people who have attended the programs. Also trainees who enjoy a training program are more likely to obtain maximum benefit from it.\(^2\) Although the participants’ enthusiasm doesn’t link directly to return on investment, it’s important because reaction is a leading indicator of information retention.\(^3\) Kirkpatrick contends that although the evaluation of reactions is an easy measurement, many trainers do not follow these five essential steps for accurate measurement:\(^4\)

1. Determine what information is desired.
2. Devise a written “comment sheet” that includes items determined in the previous step.
3. Design the sheet so that reactions can be easily tabulated and manipulated by statistical means.
4. Make the sheets anonymous.
5. Encourage the participants to make additional comments not elicited by questions on the sheet.

It is suggested that the comment sheets be given to the trainees before the program is over so that the suggestions can be used in improving the later sections of the training programs.\(^5\)

The participants should be encouraged to make open-ended comments on the back of the questionnaire.\(^6\) Specific open-ended questions can be included in the questionnaire to provide information thought to be useful common questions are:

- What are the three best things about the course?
- What are the three worst things about this course?

---

\(^1\) Shepherd, Clive, “Evaluating online learning”, [www.firstakeconsulting.com](http://www.firstakeconsulting.com)


What three changes should be made to this course?
What aspects of the process helped you to learn?
What aspects hindered your learning?

A number of times comment sheets that are used, ask the trainees to write in their answers to questions. Using a form of this kind, it becomes very difficult to summarize comments and to determine patterns of reaction. Although Kirkpatrick suggests that participants should feel free and be encouraged to make additional comments, he also contends that this type of qualitative data is extremely difficult to analyze. Thus, it is difficult to discern any patterns or trends in order to revise the training program. It has been emphasized by Kirkpatrick that the form should be designed so that tabulations can be readily made. (Refer Annexures 7.1, 7.2, 7.3)

To determine what training-evaluation tools were being used by industry, Fisher and Weinberg (1988) of Bell Communications Research, Incorporated (Bellcore) conducted a phone survey in March of 1986. The data indicated that the typical instrument to gather information regarding reactions was a “short, quickly constructed, open-ended questionnaire”. This “happy sheet”, as Fisher and Weinberg refer to it, provided subjective impressions and no data that could withstand statistical analysis or measures for reliability.

Fisher and Weinberg warn that while this questionnaire does provide a “general estimate of a particular course’s success based upon the views of the participants”, the data may be somewhat inaccurate because participants have a tendency to report what a trainer wants to hear. Also, some questionnaires have poorly constructed questions or items that predispose participants to respond in predicted ways.

Dixon (1987) claims that “the use of participant reaction forms can cause more problems than benefits for the training function of an organization”. This statement is especially true when participant reactions are the only evaluation method used. Dixon contends that three major problems result from the use of reaction forms:

---

1. The expectation that training must be entertaining. Because reaction sheets measure how the participants felt about the training, the trainer may tend to emphasize participant enjoyment during the training rather than substantive information. As a trainer is often rewarded with high marks when the participants enjoy themselves, this relationship between evaluation and participant enjoyment can become a vicious cycle. The trainer’s ratings are also a major factor in the rewards that the trainer receives from management or the client organization: renewal of a contract or a promotion. Obviously, under these circumstances the use of a reaction sheet can lead to a conflict of interest.

2. Faulty instructional design. The term “faulty instructional design” refers to a questionnaire design that asks for information that participants cannot legitimately provide. As Dixon states, the art of questionnaire design is to ask questions for which a participant can give informed responses.

3. The perception that learning is passive rather than active. This perception refers to the common belief that it is the trainer’s responsibility to ensure that participant learning occurs. Measuring how well this responsibility has been met with a reaction sheet is problematic, as a reaction sheet asks questions about the trainer’s performance and the course design without asking about the participants’ efforts to learn. Dixon emphasizes that evaluation and learning are not complete unless both functions have been measured. Ultimately, it is the responsibility of the trainer to provide information and the responsibility of the participant and the trainer to process the information. Reaction sheets rarely take into account the participant’s role as part of the training program.

7.6.2 Level 2: Learning Evaluation

It is important to recognize that favorable reaction to a training program does not assure learning. Learning, in terms of new or improved skills, knowledge and attitudes, is the primary aim of a training event. According to Kirkpatrick, the second level of analysis in the evaluation process is that of learning. Kirkpatrick defines learning as the “principles, facts and techniques that were understood and absorbed by the participants” and identifies the following guidelines or standards for evaluation in terms of learning:

---

1. Each participant’s learning should be measured by quantitative means.
2. A pretest and posttest should be administered so that any learning can be attributed to the training program.
3. The learning should be measured by objective means.
4. When feasible, a control group should be used so that comparisons can be made with the actual training group.
5. When feasible, the evaluation results should undergo statistical analysis so that learning can be viewed in terms of correlation and/or levels of confidence.

Obviously, evaluation of learning is much more difficult to measure than reaction. According to Kirkpatrick’s guidelines, a knowledge of statistical procedures is essential for accurate and meaningful measurement.

According to Carnevale and Schulz (1990), the measurement tools used to evaluate learning should reflect each training program’s particular objectives.1 Also, measures of learning changes may be taken during or at the end of a training session.

Endres and Kleiner (1990) state that pretests and posttests are necessary when evaluating the amount of learning that has taken place.2 Without a point of comparison, the measurement of learning at the end of the training program will not reveal exactly how much knowledge has been obtained from the training experience. Although paper-and-pencil tests are the most frequently used tools to measure knowledge, there are other means for gathering this kind of data.

For instance, when simulations, role-plays, or demonstrations are used to measure knowledge, the trainer can use before-and-after situations in which participants can demonstrate or perform the knowledge and techniques that they have learned. This information is consistent with Kirkpatrick’s research on the measurement of learning. In fact, like Endres and Kleiner, Kirkpatrick maintains that simulations and demonstrations can closely approximate the participants’ work environment and can help them relate the learning in meaningful ways, especially when specific job skills are the focus of the training.

---

It is relatively easy to measure the learning that takes place in training programs that are teaching skills. Programs like job instruction training, work simplification, interviewing skills, induction skills, reading improvement, effective speaking, and effective writing, would fall under this category. Classroom activities such as demonstrations, individual performance of the skill being taught, and discussions following a role-playing situation can be used as evaluation techniques. For example, in an effective speaking program, each trainee is normally required to give a number of talks and an alert trainer can evaluate the amount of learning that is taking place by observing the trainee’s successive performances.

So in these kinds of situations, an evaluation of the learning can be built into the program. If it is organized and implemented properly, the trainer can obtain a fairly objective measure of the amount of learning that has taken place. He can setup before and after situations in which each trainee demonstrates whether or not he knows the principles and techniques being taught. In every program, therefore, the trainer should plan systematic classroom evaluation to measure learning.

Where principles and facts are taught rather than techniques, it is more difficult to evaluate learning. The most common technique is the paper-and-pencil test. In some cases, standardized tests can be purchased to measure learning. In other cases the trainer may construct his own. One of the standardized tests used is the one developed by Kirkpatrick and Plany, called the Supervisory Inventory on Human Relations. (Refer Annexure 7.4). There are also standardized tests available in such areas as creativity and economics.

Kirkpatrick has suggested the following guidelines for the use of such tests:

1. The tests should be given to all trainees prior to the program.
2. If possible, it should be given to a control group, which is comparable to the experimental group.
3. These pretests should be analyzed in terms of two approaches. In the first place, the total score of each person should be tabulated. Secondly, the responses to each item of the inventory should be tabulated in terms of right and wrong answers. This second tabulation not only enables a training man to evaluate the program but also gives

---

him some tips on the knowledge and understanding of the group prior to the program. This means that in the classroom he can stress those items most frequently misunderstood.

4. After the program is over, the same test or its equivalent should be given to the trainees and also to the control group. A comparison of the pretest and posttest scores and responses to individual items can then be made. A statistical analysis of these data will reveal the effectiveness of the program in terms of training.

Tests can also be developed to suit the company’s specific requirements and there are many examples trainers and others responsible for the program have developed their own tests to measure learning in their programs. (Refer Annexure 7.5).

In Morris A. Savitt’s article called “Is Management Training Worthwhile?” he described a program that he evaluated. He devised a questionnaire which was given at the beginning of the program to determine how much knowledge of management principles and practices the trainees had at the beginning. At the end of the 10-week program, the same questionnaire was administered to test the progress made during the course. This is an example of a questionnaire tailored to a specific program.¹

It is easy to see that it is much more difficult to measure learning than it is to measure reactions to a program. A great deal of work is required in planning the evaluation procedure, in analyzing the data that are obtained, and in interpreting the results.

7.6.3 Transfer-of-Learning

Although the primary purpose of training is to bring about learning, not much money would be made available for training if that were as far as it went. Sponsors of training need to know that what is learned as a result of training will be applied back on the job with some effect, hopefully positive. Before we look at level 3 evaluations it is necessary to have an understanding of learning transfer to the job.

There may be a big difference between knowing principles and techniques and using them on the job. In the article, “Human Relations Skills Can Be


364
sharpened”, Robert Katz says, that if a person is going to change his job behaviour, five basic requirements must exist:¹

- He must want to improve
- He must recognize his own weaknesses
- He must work in a permissive climate
- He must have some help from someone who is interested and skilled
- He must have an opportunity to try out the new ideas.

These are actually the barriers to the transition of learning to changes in behaviour on the job.

According to Clive Shepherd, there are many reasons why learning does not get applied:²

- It is forgotten
- It is discouraged
- It is not reinforced
- There is no opportunity to apply it
- When applied, it doesn’t work

If any of these are the case, then you need to know about it. You may be able to revise the course or introduce new follow-up procedures that will help cure the problem.

Nanda (1988) also looked at the transference of supervisory skills following training programs and found that most supervisory-training programs are knowledge based.³ However, to be of value to the trainee and the organization, that knowledge must result in a change of attitude, followed by a change in the supervisor’s behavior. Unfortunately, the impact of most supervisory-training programs does not go beyond knowledge and awareness. One factor that often inhibits transference of learning is the organizational climate, which may be inconsistent with what is taught in the training program. This inconsistency often renders such training programs entirely ineffective. As Nanda says, “perhaps changes in attitude among top managers are key to the skill development of supervisors.”

² Shepherd, Clive, “Evaluating online learning”, www.fastrakeconsulting.com
Kelly (1982) starts with the assumption that typically only 10 percent of a company’s training transfers skills to the job.\(^1\) What happens to the other 90 percent of training? She suggests that 40 percent is lost because the training function is often isolated or peripheral: “Therefore, management, who views anyone paid to do a peripheral job as a peripheral person, will not bring that person’s ideas into the workplace”. An additional 40 percent, she suggests, is lost because most trainers or management educators do not build transfer into the training programs. Finally, 10 percent may be lost when the course designer does not deliver the training.

For skills to be transferred to the job, Kelly believes that they must be built into the training “before the first specific behavioral objective is chosen, before the first course activity is imagined or before a packaged product is selected”. In other words, the course should be designed with the specific intent of transfer to the actual job situation.

To ensure transfer of training a plan for transfer should be built into the training program from the outset. The work environment should provide positive incentives to apply the skills gained in training. The initial performance targets should be based on the training needs identified in the assessment phase. Training topics should be specific that are relevant and job related. The work-group manager or supervisor should deliver the training whenever possible. It should be ensured that practice during the training sessions clearly matches the on-the-job situation. The training should not be considered to be complete until transference has been evaluated.

It is interesting to note that if transfer of learning is considered at all, this consideration usually occurs after the training has been designed or even delivered. However, most of the guidelines suggested above should be followed during the design phase.

7.6.4 Level 3: Behaviour Evaluation

At level 3, we are measuring the extent to which new knowledge, skills and attitudes are translated into new job behaviour - in other words, the extent to which learning is being applied.

Evaluation of training programs in terms of on-the-job behaviour is more difficult than the reaction and learning evaluations, and a more scientific approach is needed, and many factors must be considered.

In the HRD literature there are relatively few examples of studies that have specifically attempted to assess the transfer of training skills or knowledge to the job. Even Kirkpatrick warns that “evaluation of training programs in terms of on the job behavior is more difficult than the reaction and learning evaluations...” As a result, much training is delivered without a plan for measuring the transfer of training.

A couple of the best evaluation studies are briefly described below:¹

The Blocker Study. A different approach was used in the study conducted in an insurance company having approximately 600 employees. Fifteen supervisors who took a course on democratic leadership were analyzed during the three-month period following the course. Eight of the supervisors were classified as authoritarian based on their behavior prior to the program.

During the three-month period immediately following the program, the changes in behaviour of the supervisors were analyzed through a study of their interview records. They used standard printed forms which made provision for recording the reason for the interview attitude of the employee, comments of the supervisor, and action taken, if any. Each supervisor was required to make a complete record of each interview. They did not know that these records were to be used for an evaluation study. There were a total of 376 interviews with 186 employees.

The interview records were classified as authoritarian or democratic. The changes in interview approach and techniques were studied during the three-month period following the course to determine if on-the-job behaviour of the supervisors changed.

The Sorensen Study. The most comprehensive research on evaluation in terms of on-the-job behaviour is Olav Sorensen’s study, “Observed changes inquiry”, designed to answer the following questions:

Have graduates of General Electric’s advanced management courses been observed to have changed in their manner of managing?
What inferences may be made from similarities and differences of changes observed in graduates and nongraduates?

First, the managers (graduates and nongraduates alike) were asked to indicate changes they had observed in their own manner of managing. Next, subordinates were asked to describe the changes they had observed in the managers during the same period. Then, managers’ peers were asked to describe changes in the managers’ behaviours. Last the superiors of the control and experimental groups were asked to describe the managers’ changes in behaviours. Sorensen did not use a before-and-after measure, instead, he asked each participant to indicate what changes, if any took place.

Kirkpatrick suggests a framework for evaluating training programs in terms of behavioral changes. ¹

1. A systematic appraisal should be made of on-the-job performance on a before-and-after basis.
2. The appraisal of performance should be made by one or more of the following parties (the more the better): The participant; The participant’s superior(s); The participant’s subordinates; and/or The participant’s peers or other people who are familiar with the participant’s performance.
3. A statistical analysis should be made to compare before-and-after performance and to relate changes to the training program.
4. The post-training appraisal should be made three months or more after the training so that the participants have an opportunity to practice what they have learned. Subsequent appraisals may add validity to the study.
5. A control group (of people who did not receive the training) should be used.

Endres and Kleiner (1990) use Kirkpatrick’s model in suggesting an approach to evaluating the effectiveness of management training.² They caution against relying on in-house performance-appraisal systems as the

primary measure of transfer of learning, as it is difficult to separate the
effects of training efforts from those of other factors. Instead, they suggest
setting initial performance objectives and monitoring accomplishment of
those objectives after training. They offer an example in which participants
write personal and professional objectives at the end of the training
experience. These objectives are then sent to the participants approximately
a week after the training. Two months later they are sent again, and the
participants are asked to comment on their performance against these
objectives. A certificate of completion for the training is issued only after
each participant’s feedback is secured.

Like Kirkpatrick, Endres and Kleiner suggest multidimensional on-the-job
evaluations, including feedback from the participant, his or her subordinates,
and peers. By using all three forms of feedback, they say, “the built-in biases
of the evaluator can be reduced as the number of evaluators having different
perspectives is increased”. Finally, they remind evaluators that other factors
can impact the effectiveness of management training and development,
including the manager, the trainer, the organization, and the environment. As
they state, “All four are complex creatures”.

Measuring changes in behaviour resulting from training programs involves a
very complicated procedure. But it is worthwhile and necessary if training
programs are going to increase in effectiveness and their benefits made clear
to top management.

7.6.5 Level 4: Results Evaluation

The real benefits of training cannot be measured in terms of learner
reactions, nor the amount of learning that has been achieved; not even the
extent to which behaviour may have changed. The real benefits come from
improved performance, which can be stated in terms of results such as,
reduced turnover, reduced costs, labour savings, improved efficiency,
reduction in grievances, increase in quality and quantity of production,
productivity increases or improved morale.

Kirkpatrick’s fourth level of evaluation is results or impact on the
organization. According to Kirkpatrick, attempting to measure results is not
for the fainthearted! Although measuring training programs in terms of
results may be the best way to measure effectiveness, Kirkpatrick himself
points out, “there are ... so many complicating factors that it is extremely
difficult if not impossible to evaluate certain kinds of programs in terms of results.” The separation of variables to measure how much of the improvement is due to training is extremely difficult. Instead of offering a specific formula, Kirkpatrick simply reports anecdotal efforts to measure results. He does applaud attempts by researchers such as Likert to use qualitative data in measuring results, but he laments the fact that current research techniques are essentially inadequate and that progress in this area is slow.

In an article in Harvard Business Review, Rensis Likert, shows that changes in productivity can be measured on a before-and-after basis.¹ A group of supervisors was trained in using democratic leadership in which decision-making involved the use of a participative technique. Another group of supervisors was trained to make their own decisions and not to ask subordinates for suggestions.

Such factors as productivity, loyalty, attitudes, interest and work environment were measured. Both training programs resulted in positive changes in productivity. But the participative approach resulted in better feelings, attitudes, and other human relations factors. Likert concludes, “Industry needs more adequate measures of organizational performance than it is now getting.”

Zenger and Hargis (1982) recommend experimental-research designs using pretesting and posttesting of experimentally trained groups with untrained control groups.² However, outside an ideal laboratory environment, this approach is not without its challenges.

In an update to Zenger and Hargis’ 1982 article, Kelly, Orgel, and Baer (1984) recommend quasi-experimental designs based on samples and groups that exist naturally in the work environment.³ An example would be two similar departments, one that receives training and one that does not. Rather than evaluating performance differences statistically and presenting those statistics-which, according to them, few people really understand-they suggest demonstrating results visually through graphic presentations.

Ban and Faerman (1990) report on their attempt to measure both skill transference and results following an intensive, twenty-four-day advanced supervisory-training program. They had hoped to study impact with an experimental design by surveying a control group of managers who had not participated in the training program. However, they had to abandon this part of their study because of logistical problems. They conclude "the literature on training evaluation may be too optimistic in recommending experimental or quasi-experimental design for many field situations".

Similarly, Trapnell (1984) remarks that "impact evaluation is not a science" because of the number of variables other than training that may affect long-term results. Despite this comment, though, Trapnell encourages the use of available secondary data, such as savings resulting from reductions in downtime, accident rates, absenteeism, customer returns, assembly-line rejects, staff turnover, and employee grievances.

Given the difficulty of results evaluations and the relative lack of objective, valid tools to use, are they worth pursuing? McEvoy and Buller (1990) suggest not only that it would be wise to think twice about pursuing such evaluations but also that not all training is results oriented. They describe their attempts to conduct a comprehensive, four-step evaluation of their training for developing executive leadership, which is similar to Outward Bound. They found that training is often used for purposes other than achieving a measurable impact on the performance of an individual employee or the organization. For example, sometimes training is seen as a prerequisite for performance that has already been judged successful or as a cultural "rite of passage" that all those hoping to advance must complete. In these cases the value of the training is more symbolic than technical.

These studies suggest that evaluation training on the basis of results or organizational impact may not be the ultimate measure. In the years since Kirkpatrick proposed his model, little has been added in the way of specific, valid tools to objectively measure training impact. It would be a good idea to conduct further studies in a greater variety of occupational settings to

---

determine reasonable, more precise estimates of performance differences between trained and untrained employees.

It is recommended that level 1 evaluation should be done for all courses. Level 2 evaluations should be done for any courses in which the trainees need to retain a set of knowledge or apply a specific skill. Level 3 evaluations are necessary in cases in which the course objective(s) is to change behavior on the job. A Level 4 evaluation should be done in those cases in which the results represent a top priority to the company; the evaluation should be able to be realistically linked to hard financial information.\(^1\)

With all the effort involved, however, it would be impractical for most companies to conduct Levels 3 and Level 4 evaluation on every single course. Recommendations include concentrating on the most expensive programs, the strategic value of a course, or courses that have high priority to upper management.

In fact, many authors emphasize the importance of considering early in the design process how each level of evaluation will be addressed. The ability to track and report regularly on the effectiveness of training programs beyond participant reaction (that is, to documentation of learning, behavioral changes, and transfer of learning) can be critical to the success of a training program. It can also cement organizational recognition of the value of training and can help to ensure continued support.

7.7 Disagreement with the model

Says Paul Bernthal, manager of research at Development Dimensions International, “Kirkpatrick’s classic model has weathered well. But it has also limited our thinking regarding evaluation and possibly hindered our ability to conduct meaningful evaluations. The simplicity and common sense of Kirkpatrick’s model imply that conducting an evaluation is a standardized, prepackaged process. But other options are not spelled out in the model.”\(^2\)

\(^2\) Abernathy, Donna J., “Thinking Outside the Evaluation Box”, www.astd.org
Although measuring training programs in terms of results may be the best way to measure effectiveness, Kirkpatrick himself points out the complications, especially for level-4 evaluations. A number of experts don’t believe that level-4 is applicable to soft-skills training, since, there are too many variables that can impact performance, other than training itself.

Although the Kirkpatrick model has served trainers well in terms of evaluating whether learners liked their instruction, whether they learned something from it, and whether it had some effect for the company, evaluation experts are now pointing out that the four-level approach has weaknesses. Mainly, it can’t be used to determine the cost-benefit ratio of training (ROI), and it can’t be used diagnostically, i.e., when a training program doesn’t deliver the expected results.

7.8 A New Model

Not everyone agrees that the Kirkpatrick model should be used for evaluations. Elwood Holton, writing in HRD Quarterly (1996), goes so far as to say it isn’t even a model, but rather merely a taxonomy.¹ The biggest problem, he says, is in trying to use the four levels to determine where a problem exists with a given educational program.

Holton proposes a new model for evaluation of training that, unlike Kirkpatrick’s four-level system, will “account for the impact of the primary intervening variables such as motivation to learn, trainability, job attitudes, personal characteristics, and transfer of training conditions.” The important differences between this and the Kirkpatrick system are:

1. Absence of reactions (level one) as a primary outcome.
2. Individual performance is used instead of behavior.
3. The inclusion of primary and secondary influences on outcomes.

Three primary learning outcome measures are proposed:

1. Learning: achievement of the learning outcomes desired in the intervention.
2. Individual performance: change in individual performance as a result of the learning being applied on the job.

¹ [http://131.96.116.76/insit/vmvw/evaluation/new.htm]
3. Organizational results: consequences of the change in individual performance.

And, according to the model, the three primary influences on learning are:

1. Trainee reactions
2. Motivation to learn
3. Ability

This model has been proposed but needs to be tested, Holton says. A simpler model may emerge from such testing; for example, perhaps measuring only primary intervening variables will be sufficient, or perhaps only a few key variables within each category should be measured.
7.9 Return on Investment

Forty years ago, Donald Kirkpatrick developed the four levels of evaluation that have become the preferred framework for practitioners. Today’s human resource development staff must conduct, in addition to the established evaluations, another measurement at the fifth level: return-on-investment or the financial implications of training. The issue of ROI in training and development has become a critical challenge for HRD professionals. In the past decade, the interest in ROI has mushroomed, leaving most major organizations scrambling for ways to tackle the issue.

The fifth level has been defined by Phillips, (1997): Return on investment (ROI) is a measure of the monetary benefits obtained by an organization over a specified time period in return for a given investment in a training program. Looking at it another way, ROI is the extent to which the benefits (outputs) of training exceed the costs (inputs).\(^1\)

ROI has become a hot topic for some good reasons.\(^2\)

**Reason 1.** In most industrialized nations HRD budgets have continued to grow year after year, and, as expenditures grow, accountability becomes a more critical attention issue. An increasing budget draws the attention of internal critics, often forcing the development of the ROI evaluation.

**Reason 2.** Total quality management and continuous process improvement have heightened attention on measurement issues. Organizations now measure processes and outputs that previously were not measured, monitored, or reported. This measurement focus has placed increased pressure on the training and HRD function to develop measures of its outputs and successes.

**Reason 3.** The reengineering and restructuring experience and the threat of outsourcing have caused many HRD executives to focus directly on bottom-line issues. Because of this scrutiny, many training and development functions have been reengineered to link programs to business needs and to enhance efficiencies. Change processes have caused HRD executives to examine evaluation issues and measure the contribution of specific programs. The threat of outsourcing has forced some HRD managers to

---

\(^1\) Shepherd, Clive, “Evaluating online learning”, [www.fustraconsulting.com](http://www.fustraconsulting.com)

\(^2\) “Why ROI?” CLIP & SAVE JULY 2000, American Society for Training & Development (ASTD), [www.astd.org](http://www.astd.org)
align programs to organizational objectives and success measures more closely so management can understand HRD’s contribution.

Reason 4. The business management mindset of current training and HRD managers leads them to place more emphasis on economic issues within the training function. Today’s training manager is more aware of the bottom-line issues in the organization and is more knowledgeable of the operational and financial areas. This new “enlightened” manager often takes a business approach to training and development, and the ROI issue is part of this process.

Reason 5. Accountability has been a persistent trend for all functions in organizations. Each support function is attempting to show its worth by capturing the value that it adds to the organization. From an accountability perspective, the HRD function should be no different from the other functions: It must show its contribution.

Reason 6. Top executives now are demanding ROI calculations in organizations where they were not required previously. For years, training and HRD professionals have convinced top executives that training cannot be measured, at least to the level desired by the executives. Yet, many senior managers are learning that training can and is measured in many organizations, thanks in part to articles in publications aimed at them. Now aware that training can be done, top executives are demanding the same accountability from their HRD departments. In some extremes, these executives are asking their HRD departments to show the return or face significant budget cuts; others are just asking for results.

For many years, measuring the return on investment for training and development has been a critical issue – on meeting agendas, in the literature, and on the minds of top executives. Although interest has heightened and some progress has been made, the topic still challenges even the most sophisticated and progressive HR departments. Some HR professionals argue that measuring ROI for training is not possible; others quietly and deliberately develop ROI measures. But overall most practitioners acknowledge that they must show a return on investment in training so that they can maintain training funds and enhance HR’s status.
7.10 The ROI Model

Phillips (1996) suggests that we measure these benefits through the use of a modified version of Kirkpatrick’s (1994) evaluation model.\(^1\) Philips’ model adds a fifth level to Kirkpatrick’s that includes measuring the return on investment. Critical to this method is the concept of “chain of effect” implied in this five level model. Each level of measurement is dependent on the previous level and each level is linked to the others. Without this link it is difficult to isolate the effect of training or to conclude with any degree of confidence that training is responsible for any improvements in performance. Thus a corporation cannot measure return of investment (the fifth level) of training without taking accurate measures at the other four levels.\(^2\)

Here’s the basic formula for calculating ROI:\(^3\)

- Collect level 4 evaluation data. Ask did on-the-job application produce measurable results?
- Isolate the effects of training from other factors that may have contributed to the results
- Convert the results to monetary benefits.
- Total the costs of training

The non-monetary benefits can be presented as additional – though intangible – evidence of the program’s success.

The model in the figure below is a framework for developing ROI.\(^4\) It tracks the steps in measuring ROI – from collecting post program data to calculating the actual return. The model assumes that training costs will be compared with monetary benefits and that training programs will also have intangible, but reportable benefits.

---

7.10.1 Isolating the effects of training

It’s a common scenario. After a major training program, there’s a boost in trainees’ work performance. Clearly, the two events are linked. But then a manager asks the dreaded question: “How much of the improvement was caused by the training?”

This familiar inquiry is rarely answered with much accuracy or credibility. Performance improvements may be linked to training, but usually non-
training factors have also contributed. As most HR practitioners know, it can be difficult to show a cause-and-effect relationship between training and performance.

Here are several ways to isolate training’s effect on performance: ¹

Use of control groups. A highly credible approach for isolating the effect of training is the use of control groups in an experimental training design. The experimental group receives training; the control group does not. Participants in both groups should be similar demographically, selected at random, and subjected to the same environmental influences. It isn’t necessary to take pre-program measurements of the two groups. Rather, measurements taken after training show the difference in performance between the two groups that can be attributed directly to training.

For example, here are some statistics for a particular performance indicator: ²

<table>
<thead>
<tr>
<th></th>
<th>Before</th>
<th>After</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group receiving training</td>
<td>10</td>
<td>16</td>
<td>+60%</td>
</tr>
<tr>
<td>Control group</td>
<td>10</td>
<td>12</td>
<td>+20%</td>
</tr>
<tr>
<td>Effect of training</td>
<td></td>
<td></td>
<td>+40%</td>
</tr>
</tbody>
</table>

It can be seen from this table that performance has risen by 20% even without training. The increase in performance that can be attributed to the training is therefore 40%.

A disadvantage of the use of control groups is a misperception that the training staff is turning the workplace into a lab. To avoid this negative image, some organizations conduct a pilot of the training program using pilot participants as the experimental group and non-participants as a control group. In fact, the non-participants aren’t even informed of their status as “the control group.” Sometimes, management may not want to take the time to experiment; it may just want to make sure employees get the training. But using control groups is worthwhile when training programs are costly and linked with organizational objectives.

Forecasting. This approach is more analytical and mathematical than the trend line. Instead of drawing a straight line, a linear equation is used to calculate a value of the anticipated performance improvement. A linear model (such as \( y = ax + b \)) is appropriate when only one variable influences the results. When several variables intervene, it’s necessary to use sophisticated statistical models. Without them, forecasting is difficult to implement. Still, it can be an accurate predictor of performance variables without training, if the appropriate data and models are available.

Participant estimation. This approach involves asking participants to determine how much performance improvement is due to training. Their actions have produced the improvements, so they should have some idea of how much improvement is because they applied what they learned in training. Management tends to find such reports credible because participants are at the center of the improvement.

Participants’ input can be obtained by asking the following questions:

- What percent of the improvement can be attributed to the application of skills, techniques, or knowledge gained in the training?
- What is the basis for your estimation?
- What degree of confidence do you have in your estimation?
- What other individuals or groups could make an estimation?
- What other factors do you think contributed to the improvement?

To be conservative, it’s recommended to factor in a confidence level. For example, if a participant estimates that 50 percent of an improvement is due to training but is only 70 percent confident about that estimate multiply the confidence percentage by the improvement percentage and divide by 100, for a confidence level of 35 percent. Then multiply that figure by the amount of the improvement in order to isolate the portion attributable to training. To calculate ROI, convert that portion to a monetary value. To enhance this approach, management can approve participants’ estimations. For example, in performance-management training at Yellow Freight Systems, participants estimated the amount of savings that could be attributed to the program. Managers reviewed and approved the estimates confirming participants’ estimations.

One disadvantage to this approach is obvious: It’s just an estimate. The input data may be unreliable. Some participants aren’t comfortable providing
estimates; some may not be able to estimate improvements because they
don’t know which factors contributed. The advantages are that it’s
inexpensive, timesaving, and easily understood by most participants and
others who review evaluation data. And the estimates do originate from a
credible source - the people who actually produced the improvement.

Supervisor estimation. Participants’ supervisors may provide input in lieu of,
or in addition to, participants’ estimations. In some settings, participants’
supervisors may be more familiar with other factors that could have
produced the improvements. It’s recommended to ask supervisors the same
questions asked of participants. Supervisor estimation should be treated
the same way as participant estimation in summarizing and analyzing the data.
The evaluator may not know which estimates to use. A conservative
approach is to use the lowest value and include an appropriate explanation.
Or, the evaluator can recognize that each source has its own perspective and
average the two, placing equal weight on each group’s input.

This approach has the same disadvantages as participant estimation.
Because it’s subjective, it may be viewed skeptically by management,
supervisors may be reluctant to participate. Or, they may be incapable of
providing accurate estimates. The advantages are also the same: It’s simple,
inexpensive, and fairly credible because the information comes from the
“horse’s mouth” - in this case, the supervisors of people who received the
training. Credibility rises when supervisors’ estimates are combined with
participants’ estimates and when a confidence level is factored in.

A restaurant chain implemented a training program on performance
management for manager-trainees. Trainees learned how to establish
measurable goals for staff, how to provide performance feedback, how to
measure progress toward goals, and how to take action for ensuring that
goals are met. Trainees developed action plans for improvement, using the
skills taught in the training. The top managers learned how to convert
measurable improvements into economic values. They decided employees
could focus on any improvement areas (such as inventory, food spoilage, or
employee turnover) on the conditions that they use the new skills taught in
training and that improvements be converted to either cost savings or profits.

As part of a follow-up evaluation, trainees’ action plans were documented to
show results in quantitative terms converted to monetary values. Trainees
were asked to estimate (conservatively) the percent of improvement that
resulted from the application of skills either acquired or enhanced in training. Each improvement was calculated using an annual monetary value. To implement the improvements, trainees worked closely with the restaurant managers, who estimated for the trainees the percent of improvement (outlined on their action plans) that could be attributed to training.

Management estimation. Top-level managers can provide estimates on the percent of improvement they attribute to training. At Litton Guidance and Control Systems, management applied a subjective figure (60 percent) to improvements due to training, after considering other contributing factors such as changes in processes, procedures, and technology. The upshot was that training received credit for 60 percent of the improvements in quality and productivity. Clearly, this approach can be highly subjective. But then, the input is from people who provide the training funds.

Customer input. Input can be directly elicited from customers. Ask them why they chose a particular product or service. Ask them to explain how their reactions to a product or service were influenced by employees who were using the knowledge and skills taught in training. This approach focuses directly on what training programs are designed to improve.

For example, following a bank’s teller-training program, customers indicated in market-research data, a 5 percent reduction in customer dissatisfaction with teller knowledge.

Expert estimation. Experts such as independent consultants and industry sources must be carefully selected regarding their knowledge of a particular process, program, or situation. For example, an expert in quality can provide fairly reliable estimates of how much quality improvement can be attributed to training and how much can be attributed to other factors associated with a TQM effort.

The approach may lack credibility because the estimates come from external sources. Still, it’s a quick source of input from a reputable source. Right or wrong, management can sometimes place more confidence in external expert than internal staff.

Subordinate input. In some situations, participants’ subordinates can provide input on training given to supervisors and other managers on implementing work-unit changes or developing new skills in dealing with employees. Subordinates usually can’t estimate how much of an improvement is
attributable to training, but they can provide input about specific changes that have occurred since the supervisor received training. And they also can help determine the extent to which other factors have changed.

Subordinate input is usually obtained through surveys or interviews. When the survey responses show significant changes in supervisors’ behavior after training and no significant change in the general work climate, improvements in work performance can be attributed to changes in supervisors’ behavior.

Typically, subordinates are aware of the factors that have caused changes at work, and they can provide reliable input about the magnitude of such changes. When combined with other approaches, subordinate input is even more credible.

Other factors. In some situations, it’s feasible to calculate the effect of factors other than training that may have contributed to some improvement and then to conclude that training accounts for the rest.

For example, a consumer-lending program for a large bank experienced a significant increase in the number of loans after training was provided to loan officers. In addition to the effect of training, other factors included falling interest rates and loan officers’ growing confidence in their knowledge and expertise.

This approach is appropriate when other factors are easily identified and when the necessary mechanisms for calculating their effect are in place. In some cases, it’s just as difficult to estimate the effect of factors other than training. This approach is highly credible when the methods used to isolate the effect of other factors are credible.

But which one? With 10 approaches available, it can be difficult to select the most appropriate one. It’s important to consider the following criteria:

- Feasibility
- Accuracy
- Credibility
- Costs
- Time - including that of participants, managers, and others.
Generally, two approaches are better than one. In using multiple sources, it’s recommended to combine the inputs. This conservative approach builds acceptance. The target audience should receive explanations of the approach and the subjective factors.

7.10.2 Converting the results into monetary benefits

It is useful to divide training results into hard data and soft data. Hard data are the traditional measures of organizational performance. They are objective, easy to measure, and easy to convert to monetary values. Management tends to find hard data highly credible. Hard data is available in most types of organizations, including manufacturing, service, non-profit, government, and educational. Hard data present the following areas of a work process:

- Output
- Quality
- Time
- Cost

For example, a government office that approves applications for visas typically collects data in all four areas to measure overall performance: output (the number of applications processed), quality (the number of errors in processing each application), time (the time it takes to process and approve an application), and cost (for processing each application).

Soft data are needed on training programs that focus on developing such “soft” skills as communication. Typically, soft data – such as employee absenteeism and turnover – are subjective because they have to do with behaviour they are difficult to measure and convert to monetary values. And when compared with hard data, soft data are usually found to be less credible as a performance measure.

---

Here are some examples of hard and soft data:

<table>
<thead>
<tr>
<th>Hard</th>
<th>Soft</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Output</strong></td>
<td><strong>Work habits</strong></td>
</tr>
<tr>
<td>Units produced</td>
<td>Employee absenteeism</td>
</tr>
<tr>
<td>Items assembled or sold</td>
<td>Tardiness</td>
</tr>
<tr>
<td>Forms processed</td>
<td>Visits to the dispensary</td>
</tr>
<tr>
<td>Tasks completed</td>
<td>Safety-rule violations</td>
</tr>
<tr>
<td><strong>Quality</strong></td>
<td><strong>Work climate</strong></td>
</tr>
<tr>
<td>Scrap</td>
<td>Employee grievances</td>
</tr>
<tr>
<td>Waste</td>
<td>Employee turnover</td>
</tr>
<tr>
<td>Rework</td>
<td>Discrimination charges</td>
</tr>
<tr>
<td>Product defects or rejects</td>
<td>Job satisfaction</td>
</tr>
<tr>
<td><strong>Time</strong></td>
<td><strong>Attitudes</strong></td>
</tr>
<tr>
<td>Equipment downtime</td>
<td>Employee loyalty</td>
</tr>
<tr>
<td>Employee overtime</td>
<td>Employees’ self confidence</td>
</tr>
<tr>
<td>Time to complete projects</td>
<td>Employees’ perceptions of job responsibilities</td>
</tr>
<tr>
<td>Training time</td>
<td>Perceived changes in performance</td>
</tr>
<tr>
<td><strong>Cost</strong></td>
<td><strong>New skills</strong></td>
</tr>
<tr>
<td>Overheads</td>
<td>Decisions made</td>
</tr>
<tr>
<td>Variable costs</td>
<td>Problems solved</td>
</tr>
<tr>
<td>Accident costs</td>
<td>Conflicts avoided</td>
</tr>
<tr>
<td>Sales expenses</td>
<td>Frequency in use of new skills</td>
</tr>
<tr>
<td><strong>Development and advancement</strong></td>
<td></td>
</tr>
<tr>
<td>Number of promotions or pay increases</td>
<td></td>
</tr>
<tr>
<td>Number of training programs attended</td>
<td></td>
</tr>
<tr>
<td>Requests for transfer</td>
<td>Performance-appraisal ratings</td>
</tr>
<tr>
<td><strong>Initiative</strong></td>
<td></td>
</tr>
<tr>
<td>Implementation of new ideas</td>
<td></td>
</tr>
<tr>
<td>Successful completion of projects</td>
<td></td>
</tr>
<tr>
<td>Number of employee suggestions.</td>
<td></td>
</tr>
</tbody>
</table>

**Table: 7.3** Hard and soft data
The conversion

Here are five steps for converting either hard or soft data to monetary values:

➢ *Step 1: Focus on a single unit.* For hard data, identify a particular unit of improvement in output (such as product, services, and sales), quality (often measured in terms of errors, rework, and product defects or rejects), or time (to complete a project or respond to a customer order). A single unit of soft data could be one employee grievance, one case of employee turnover, or a one-point change in customer-service index.

➢ *Step 2: Determine a value for each unit.* Place a value on the unit identified in step 1. That’s easy for measures of production, quality, time and cost. Most organizations record the value of one unit of production or the cost of a product defect. But the cost of one employee absence, for example, is difficult to pinpoint.

➢ *Step 3: Calculate the change in performance.* Determine the performance change after factoring out other potential influences on the training results. This change is the output performance, measured as hard or soft data, that is directly attributable to training.

➢ *Step 4: Obtain an annual amount.* The industry standard for an annual performance change is equal to the change in performance data during one year. Actual benefits may vary during the course of a year or extend past one year.

➢ *Step 5: Determine the annual value.* The annual value of improvement equals the annual performance change, multiplied by the unit value. Compare the product of this equation to the cost of the program, using this formula: ROI = net annual value of improvement – program cost.

There are several other ways to convert monetary values. Some are appropriate for a specific type of data or data category; others are appropriate for any type of data. Here are some options.

*Converting output to contribution.* When a training program has produced a change in output, the value of the increased output can be determined from accounting or operational records. In for-profit organizations, this value reflects the “profit contribution” of an additional unit of product or service. In not-for-profit organizations, the contribution of value may show in the savings from producing an additional unit of output for the same input.
The calculations for measuring such contributions depend on the organization and its records. Most monitor performance output. If such data aren’t available, managers may use marginal-cost statements and sensitivity analyses to pinpoint the values associated with changes in output.

For example, a bank’s sales seminar for customer-loan officers resulted in an increase in the volume of loans (output). To measure the training’s return on investment, it was necessary to determine the value (profit contribution) of one additional consumer loan — an easy item to calculate from the bank’s records.

The first step was determining the yield, also available from bank records. The next step was calculating the average spread between the cost of funds and the yield received on a loan. For example, the bank could obtain funds from depositors at 5.5 percent on average, minus the cost of making a loan, including the advertising and employees’ salaries.

Calculating the cost of quality. The cost of quality is an important measure in most manufacturing and service firms. Because many training programs are designed to improve quality, the HR department must place a tangible value on quality improvement. For some quality measures, that’s easy. For example, if quality is measured as product-defect rate, the value of an improvement is shown in eliminating the cost to repair or replace a defective product.

The most obvious cost of poor quality is the waste generated by mistakes: defective products, spoiled raw materials, and discarded paperwork. Such waste is translatable to monetary values. In addition, employee errors can cause expensive rework. The most costly rework is when a product is delivered to a customer and returned for repair. Maintaining a staff to perform rework is added overhead. In most manufacturing plants, the cost of rework is from 15 to 70 percent of productivity. In most banks, about 35 percent of operating costs are due to rework.

Perhaps the highest cost of poor quality is customer dissatisfaction. It can lead to loss of business. Customer dissatisfaction is difficult to quantify. Typically, sales-and-marketing managers and marketing surveys are the best sources for measuring the effects of customer dissatisfaction.

Converting employees’ time. Many training programs focus on reducing employees’ work time. Employee time is money, including wages and
benefits. A training program may enable a team to perform tasks in less time or with fewer members; time management can help individual employees to save time. The value of the time saved is an important measure of a program’s success, and conversion is relatively easy. The most obvious timesaving is the reduced labor costs of performing work. The monetary savings equal the hours saved, multiplied by the per hour labor cost.

For example, after attending a time-management training program, participants estimated that they now save an average of 74 minutes per day, worth $31.25 per day or $7,500 per year in labor. This timesaving is based on the participants’ average salary, plus benefits.

Generally the average wage (with a percent added for employee benefits) is sufficient for most ROI calculations. But some employees’ time is worth more. Some experts recommend that “employee maintenance” costs other than employee benefits be figured out in the average labor cost per employee, including such items as office space, furniture, telephone facilities, computers, calculators, and administrative support. Then the average wage rate may rise. The most conservative approach is to use salary plus employee benefits.

In addition to reduced labor cost, other benefits can result from a time savings, including improved service, avoided penalties, and added opportunities for profit. Timesavings are realized only when the amount of time saved translates to a cost reduction or profit contribution. The time saved must be used productively.

Using historic costs. Sometimes a company’s records will show the cost and value of one unit of improvement. Its necessary to identify the appropriate records and tabulate the actual costs of items in question. Historic data are usually available for hard data and some selected soft data.

For example, a training program for improving safety performance used various measures for all safety related items, including the accident frequency rate and the total cost of workers’ compensation. By examining the company’s records and using a year of data, the HR department was able to calculate the average cost of each safety measure.

Using internal and external experts. When converting soft data without historic records, it’s recommended to consider input from experts on the processes involved. They can provide the cost (or value) of one unit of
improvement. They tend to be close to the situation and to have earned management’s respect. When internal experts aren’t available, external experts can fill the gap. Most experts use their own approaches, so it’s best to explain specifically what’s needed. They should understand the processes and be willing to provide estimates with explanations.

In one organization, a training program for reducing the number of employee grievances ended in soft data, to be monitored by the organization. Except for one instance of reimbursed back pay, the organization had no records on the costs of grievances (such as the cost of external assistance or the time involved in working with a complainant). An expert had to estimate—in this case, the manager of the labor relations department. He based his estimate on his perception of the average settlement when a grievance is lost, including such costs as arbitration and legal fees. He also factored in an estimated amount of time spent by supervisors, staff, and employees associated with the grievance. This internal estimate, though imprecise, was appropriate for the analysis. And management found it credible.

Using data from external studies. For some soft data, it may be appropriate to use research to estimate the value. It’s fortunate that many databases contain studies on the costs of various items related to training, including employee turnover, absenteeism, and grievances, as well as safety and customer satisfaction. Ideally the data should come from a similar setting in the same industry.

For example, the evaluation of an HR program for reducing the turnover of branch managers in a financial-services company included the costs of employee turnover, including recruitment, orientation, and training for a new manager, as well as the costs of the severance an unemployment pay for an exiting manager. Many HR practitioners don’t want to calculate the cost of turnover, particularly when it’s needed just for a one-time event, such as training evaluation. In the example, the cost was determined (based on industry standards) to be about one-and-a-half the average annual salary of an employee, adjusted for the average base salary of a branch manager.

Using participants’ estimates. Sometime the people closest to an improvement can provide the most reliable estimates on its value. Training participants can estimate the value of a soft data improvement they have made by applying the skills they learned in training.
For example, to calculate ROI on a supervisory training program on lowering the rate of employee absenteeism, it was necessary to determine the average value of one absence, without the benefit of historic records. During the training, participants estimated the cost of an absence, based on their personal experience. Next, supervisors were asked to estimate the average cost on absence in their work units, based on how an employee’s absence is compensated. Then all of the estimated values were averaged.

*Using supervisor’s estimates.* Participants’ supervisors are another source for determining the value of a unit of improvement due to training. For example, after completion of a training program for managers at a company, participants estimated the value of the improvements directly related to the training. Their managers also provided estimates after reviewing the process by which the participants had created their estimates. Then, the managers either confirmed or adjusted participants’ values.

*Using senior managers’ estimates.* Senior managers can place a value on an improvement, based on their perception of its worth, when its too difficult to calculate the value or when other sources of estimates are unavailable or unreliable.

*Using HR’s estimates.* This approach may be perceived as biases. After all, the HR department will determine the basis for its claim for improvements due to training. For example, in a training program for dispatchers at an oil company, the HR department estimated the cost of one employee absence to be $200. Then it used that value to calculate the savings due to training on reducing absenteeism rate.

When reporting training results, credibility is always an issue. Its crucial that the data be accurate and that the conversion process be believable. Many HR practitioners are more comfortable reporting that training resulted in a reduction in employee absenteeism from 6 to 4 percent, without placing a monetary value on the improvement. They assume that the people receiving the information will assign their own values. Unfortunately, those people may know little about the cost of absenteeism. Or, they may underestimate the actual value. That’s why accurate ROI is important.

Less than precise estimates, assumptions, forecasts, and external data may make some HR practitioners hesitant to conduct conversion. But they can raise credibility by following these guidelines:
• Take a conservative approach when making estimates and assumptions
• Use the most credible and reliable sources for estimates
• Explain the approaches and assumptions used in conversion
• When results appear overstated, consider adjusting the numbers to achieve more realistic values
• Use hard data whenever possible.

With soft data, senior management may adjust the results so that they are more linear and concrete. Or, they may adjust the results to reflect the time value of money because most investments in training are made at one time and the return is realized at a later time. Such adjustments are usually negligible compared with the benefits.

Many organizations are trying to become more aggressive in determining the monetary benefits of training. They are no longer satisfied just to report business results. Instead, they are converting business results to monetary values and comparing them with the cost of training to obtain the true return on investment.

7.10.3 Forecasting and measuring costs

ROI relates to a specified period of time, typically a year or two years. First all of the costs associated with the particular training program over this period are measured.¹

Design and development costs. The first category of cost to be considered is the design and development of the training program, whether this comprises classroom events, self-study materials, simple coaching sessions or some combination. You will need to consider:

• Internal days of design and development
• Costs of external designers and developers
• Other direct design and development costs (purchase of copyrights, travel, expenses, etc.)
• Outright purchase of off-the-shelf materials

¹ Shepherd, Clive, “Evaluating online learning”, www.fastrakconsulting.com
Promotional costs. Most organizations devote effort to promoting their training programs. This second category takes promotional costs into account:

- Internal days of promotional activity
- Costs of external agencies
- Other direct costs of promotion (posters, brochures, etc.)

Administration costs. An allowance must be made for the time taken by the training department in administrating the training program. This will typically be a factor of the number of students:

- Hours of administration required per student
- Direct administration costs per student (joining materials, registration fees, etc.)

Faculty costs. The next category of costs relates to the delivery of the training, whether this is mediated by faculty (tutors, instructors, coaches, etc.) or is self-administered (workbooks, CBT, online training, etc.). Let’s start with the information needed to calculate faculty costs:

- The number of students who will be going through the program
- Hours of group training (whether classroom-based or delivered in real time, online)
- Hours of one-to-one training (typically face-to-face, but could conceivably be conducted by telephone, video conferencing link or in real-time, online)
- Hours of self-study training
- Additional faculty hours (preparation time, the time needed to review or mark submitted work or the time needed to correspond by email or bulletin boards with online students)
- Faculty expenses (travel, accommodation, subsistence, etc.).

Materials. Then there’s the cost of materials:

- Cost per student of training materials (books, manuals, consumables, etc.)
- License cost per student for use off-the-shelf materials
Facilities. You will also need to allow for the cost of your training facilities, whether these are internal or external. Make sure to include the rental or notional internal cost of the following:

- Training rooms
- Open learning / self-study rooms
- Equipment used

Student costs. Probably the most significant delivery cost relates to the students themselves. It is only necessary to charge a student’s cost against the program if training is undertaken in time that would otherwise be productive and paid for, so you only need to estimate the amount of travel and training that is undertaken in productive work time, i.e. not in slack time, breaks or outside work hours.

When an employee goes through a training program in work time, the organization is not only having to pay that person’s payroll costs, they are also losing the opportunity for that person to add value to the organization. When a salesperson is on a course, they are not bringing in new business. Similarly, a production line worker is not creating products, a researcher is not developing new ideas and an accountant is not finding ways to save money.

If an employee can be easily replaced while they are undergoing training, then there is no lost opportunity – the cost is simply the employee’s payroll costs. In many cases, however, it is simply not practical to obtain a suitable replacement, so the output that the employee would have generated in the time that they are receiving training will be lost. In this case, the true cost of the employee being trained is the lost opportunity – the ‘opportunity cost’. They are greater than an employee’s payroll costs and need to be considered in any serious evaluation of costs.

Finally, don’t forget to include any direct student expenses - travel, accommodation and subsistence.

Evaluation costs. You also need to make an allowance for the time spent evaluating the training, whether this is an ROI analysis or some other method.
7.10.4 Calculation of ROI

On the assumption that benefits will continue to accrue some time after the training, then the period that you specify is critical to the ROI figure you will obtain. You may like to specify a period that fits in well with your organization’s planning cycle – perhaps a year or two years. On the other hand, you may wish to calculate the period to correspond to the lifetime of the benefit, in which case you will need to know how long the average student stays in a position in which they can continue to apply the knowledge and skills being taught.¹

It is relatively simple to calculate return on investment:

\[
\% \text{ ROI} = \left( \frac{\text{benefits}}{\text{costs}} \right) \times 100
\]

Let’s imagine you have been running an online management-training program and want to calculate the ROI over the first year. You measure the costs as $100,000 and the benefits as $130,000. Your ROI is 130,000 / 100,000 x 100, or 130%.

Payback period. Another way of looking at ROI, is to calculate how many months it will take before the benefits of the training match the costs and the training pays for itself. This is called the payback period:

\[
\text{Payback period} = \frac{\text{costs}}{\text{monthly benefits}}
\]

Payback period is a powerful measure. If the figure is relatively low – perhaps only a few months – then management will be that much more encouraged to make the training investment. As a measure, it also has the advantage of not requiring an arbitrary benefit period to be specified.

7.11 Recommendations by ASTD

Currently, it’s difficult to pinpoint the state of ROI within the field. Many HR managers are reluctant to disclose internal practices. And many say that little progress has been made, even in the most progressive organizations. It is also difficult to find case studies that show specifically what organizations have done in measuring ROI.

Recognizing this void the ASTD has compiled and published case studies with real life examples for measuring the ROI in training in *Measuring Return on Investment* (ASTD, 1994). The case studies in the ASTD project represent a wide range of settings, strategies, and approaches in manufacturing, service, and government organizations. The training audiences varied from all employees to managers only to specialists only. Though most of the programs focused on training and development, others included such areas as total quality, performance management, and employee selection. The cases provide a rich source of information on the strategies and thought processes of some of the best practitioners, consultants, and researchers in the field. The companies’ returns on investment ranged from 150 to 2,000 percent.

Several common approaches have emerged. They could be considered best practices or just recommendations. Whichever, they seem to have worked well for the companies in the case studies.¹

✓ *Set targets for each evaluation level.* Some organizations set a target for each level of evaluation, a target being the percentage of HR programs that will be measured at that level. For example, many organizations require 100 percent of their training programs to be evaluated at level 1 because it’s fairly easy to measure participants’ reactions. Level 2 (learning) is also relatively easy to measure. Typically the target range is 40 to 70 percent, depending on the type of program. Level 3 evaluation (on-the-job application) involves more time and expense to conduct follow-up evaluations so targets tend to be lower at 30 to 50 percent. Level 4 (business results) and level 5 (ROI) evaluations require significant resources and budgets so their targets tend to be small: 10 percent for level 4 and five percent for level 5.

Establishing evaluation targets has several advantages. One, it provides measurable goals for assessing the progress of all training or a particular segment. It focuses attention on accountability and communicates a strong message to the HR staff about the need for measurement and evaluation.

✓ *Evaluate at the micro level.* Measurement and evaluation usually focus on a single program or a few tightly integrated programs. ROI measurement is

more effective when applied to one program that can be linked to a direct payoff. When all of the courses in a series must be completed before their common objectives are met, it may be appropriate to wait to evaluate the series as a whole. The decision to evaluate several courses in a series should take into account the training goals, timing of the courses, and cohesiveness of the series.

It can be difficult to evaluate a series conducted over a long period of time. A cause-and-effect relationship becomes more confusing and complex. Also, it is hard to evaluate an entire function, such as management development, career development, executive education, or technical training.

✔ Use a variety of methods. The companies in the case studies use a variety of approaches to collect evaluation data. They don’t latch on to one or two practices and stay with them regardless. They recognize that every program, setting, and situation is different. They know that techniques such as interviews, focus groups, and questionnaires work well in some situations and that action plans, contracts, and performance monitoring are needed in others. These companies use internally developed criteria to match a particular data-collection method with the training program.

✔ Isolate the effects of training. A critical aspect of the evaluation process is trying to isolate the effect of the training from other factors occurring during the same period that could affect business results. Most of the time, training can take only partial credit for improvements in on-the-job performance. When planning to measure ROI, the case study organizations go beyond a standard control group analysis to use one or more techniques for isolating extraneous factors.

✔ Use sampling wisely. It’s rare for organizations to use statistical sampling in selecting a sample of training programs in which to measure ROI. For most, the result would be too many calculations. For the sake of practicality, many organizations decide to evaluate just one or two sessions of their most popular training programs. Others select one program from each major training segment. It’s recommended that organizations calculating ROI for the first time select only one course to measure, as a learning experience. If sampling is used, it’s important to be statistically sound. But it’s more important to consider the trade off between the available resources and what kind of ROI calculations management is willing to accept. The sample number depends on the following variables:
  - The HR staff’s expertise on evaluation
- The type of training programs being evaluated
- The recourses allocated for evaluation
- The degree of support from management for training and development
- The organization’s commitment to measurement and evaluation
- The amount of pressure from others to show ROI calculations.

Other variables endemic to the particular organization may apply.

✓ *Convert program results to monetary values.* The organizations in the case studies seek a specific return on investment. Consequently, data on business results must be converted to monetary benefits. These companies aren’t content to show just that training can result in such improvements as increased productivity and decreased employee turnover. They take the process a step further by converting such improvements to monetary units so that the improvements can be compared to costs and further developed into an ROI calculation.

For such hard data items as productivity, quality, and time, the conversion to monetary units is relatively easy; soft-data items such as customer satisfaction, employee turnover, and job satisfaction aren’t so easy to convert. Still, there are techniques for making the conversions with a reasonable amount of accuracy, and several strategies are used in these case studies.

There are still some **unanswered questions and problems** about measuring ROI.

- *Cost standards.* The methods used to monitor cost vary widely. What one organization considers a cost of training, another does not. The HR field needs standard cost data. It’s becoming increasingly difficult to compare costs from one program to another. Most efforts to solve the problem have failed. In the interim, it’s necessary to describe the cost components that make up the total cost of any effort to measure ROI.

- *Evaluation design.* Many organizations don’t design their evaluations to isolate the effect of training. Control groups are rarely used even though they can be used effectively without the disruption, problems, and inconvenience usually feared by practitioners. Though a control

---

group approach is preferable, other evaluation designs such as trend line analysis, forecasting and estimations can be useful.

- **Standard methodology.** Evaluation techniques vary, though there are only so many ways that data can be collected and analyzed. Often data collection methods are used without regard to their advantages or disadvantages. The different levels and terminology adds to the confusion. As professionals, we need to standardize and publicize evaluation methods.

- **Statistics.** Many Hr practitioners avoid statistics. But statistical analysis can provide a sound basis for conclusion about training results. And though many top managers don’t understand statistical analysis, they need to feel confident that any conclusions about training results are supported by appropriate methodology.

- In several of the case studies and other evaluation projects, the power of statistics is largely ignored. Sample sizes are so small that the results can’t be considered supportable, at least statistically.

- **Converting data to monetary values.** Because of the subjective nature of the process many HR programs aren’t converted to monetary units. Yet, this conversion is an essential step in ROI calculations in which monetary benefits must be compared with costs. It should be a fundamental requirement for some level 4 evaluations.

Although progress has been made in the implementation of ROI, there are other barriers that inhibit the implementation of the concept. Some of these barriers are realistic; others are actually myths based on false perceptions: ¹

**Costs and time.** The ROI process will add costs and time to the evaluation of programs, although the amount will not be excessive. A comprehensive ROI process will probably not cost more than 4-5 percent of the overall training and HRD budget. The additional investment in ROI would perhaps be offset by the additional results achieved from these programs and the elimination of unproductive or unprofitable programs.

**Lack of skills and orientation for staff.** The typical training and development program does not focus on results: it focuses more on learning outcomes.

¹ "Barriers to ROI Implementation", Info-line 9805
Consequently, a tremendous barrier to implementation is the change for the overall orientation, attitude, and skills of the HRD staff.

_Faulty needs assessment._ Some training programs have been implemented for the wrong reasons (such as an effort to chase a popular fad or trend in the industry). Thus, an ROI calculation for an unnecessary program will likely yield a negative value.

_Fear._ A concern may exist about the consequences of negative ROI. The ROI process also stirs up traditional fear of change.

_Discipline and planning._ A successful ROI implementation requires much planning and a discipline approach to keep the process on track. Implementation schedules, evaluation targets, ROI analysis plans, measurement and evaluation policies, and follow-up schedules are required. The HRD staff may not have enough discipline and determination to stay on course. This becomes a barrier, particularly when there are no immediate pressures to measure return.

The search for best practices has revealed some important concerns. It’s almost universally agreed that more attention regarding ROI is needed. But only a few successful examples of ROI calculation exist. The process isn’t as difficult as it may seem. The approaches and techniques can be useful in a variety of settings. Practitioners and researchers must continue to refine the techniques and show successful applications.
7.12 Evaluation designs

Designing a good evaluation effort involves knowing when to collect evaluation measures and which groups to collect them from. Together these factors define the experimental design used to assess the impact of training.

Because reaction measures simply assess whether or not participants like the training and think it will be useful, these measures are collected during or immediately after the training. Ideally, an additional reaction questionnaire should be sent to the participants several months after the training, to see if they still believe that the training has been of use in their jobs.

The purpose of learning, behaviour, and trying results measures is quite different from that of reactions measures, in that the trainer is to discover whether or not a change has occurred in the variable being measured – that is, are trained employees behaving differently now than they were before the training? If there has been a change, the trainer will want to know whether or not it can be attributed to the training program – that is, did the training program bring about the change?

There are two basic strategies for determining whether a change has occurred. The first is to compare the trainees after the training to the way they were before the training. At the very least, this comparison involves the collection of evaluation measures at two points in time. The second strategy is to compare the learning, behaviour, or results of the trained group to the learning, behaviour, or results of a group that has not been trained but is otherwise identical to the trained group. The strongest evaluation designs draw on both these strategies. There are many complex and highly effective designs for evaluating training.¹

Simple Designs

Train → Measure

One-shot posttest-only design

Measure → Train → Measure

One-group pretest – posttest design

Stronger Designs

Measure → Measure → Measure → Train → Measure → Measure → Measure → Measure

Multiple baseline design

Group 1

Measure → Train → Measure

Group 2

Measure → No Training → Measure

Pretest – posttest control – group design

Figure: 7.2 Designs for Evaluating Training
One-shot posttest-only design. In the one-shot posttest-only design, training evaluation measures are collected only from the trained group, after the training has been conducted. Because there is no pretraining measure and no untrained group for purposes of comparison, there is no way to determine whether or not a change has occurred, or whether any change has been caused by training. However, if the aim is to determine whether a desired standard of performance has been reached, this simple design might produce useful data. For instance, one might be able to verify that 90 percent of the trainees passed the learning test at the end of training, or that after the training program, customer complaints averaged only one per thousand.

One-group pretest-posttest design. Another very simple design is the one-group pretest-posttest design, in which the training group is assessed both before and after the training. For instance, the productivity of the trained group might be found 5 percent higher than it was before training. Although this design does allow a trainer to determine if there has been a change in learning, behaviour, or results, it does not enable the trainer to conclude with absolute certainty that the training brought about the change. A change from one time period to another can be caused by anything that occurs between measurements, not just by the training. For instance, there might be a new supervisor, revised work methods, a change in the quality of raw materials, an increase in workload, employee turnover, a change in the pay system, or union activity. Any of these or similar factors could affect behaviour or results measures and cause the training evaluation to be misleading. If non-training concurrent events seem unlikely to account for a change from pre-to post training, then the trainer can gain some useful information about the effectiveness of the training from this simple design.

Multiple-baseline design. An improved evaluation design that avoids some of the above-mentioned problems is the multiple baseline design. In this design, the trainer measures the group several times both before and after training. The trainer probably should not use an obtrusive measure, such as a questionnaire or learning test. Trainees could improve over time just because they are gaining practice with the measure. Objective measures of behaviour or results are less obtrusive, and they are easy to collect repeatedly.

The multiple baseline design allows the trainer to observe trends in performance and to see if there is a change in the trend immediately after the training. For instance, a trainer may find that employees are slowly improving with experience over time but that a big jump in performance
occurs after the training. The average results of a training group on a series of evaluations might be 10, 11, and 12 on pretests and then 15, 17, and 18 after the training. This design enables the trainer to detect the training effects over and above simple experience effects and also helps to rule out coincidental factors as explanations for any changes occurring immediately after the training. It would be an extreme coincidence if any other event occurred exactly and only at the same time as the training. Thus if this design is used, the trainer can be more certain that the training caused the observed change. This design is the best one to use if all employees are to be trained simultaneously, leaving none to serve as control group.

*Pretest-posttest control-group design.* An even better design uses a control group of employees who are very similar to the training group except that they do not receive the training (at least not yet). In the pretest-posttest control-group design, both the group to receive training and the control group are measured at least once before and once after the training. This design allows the trainer to draw quite firm conclusions about (1) whether any change has occurred and (2) if it has, whether the change has resulted from the training.

The trainer might normally expect to find that the trained group improves from pretest to posttest, whereas the control group stays the same. However, other patterns of results can also be interpreted under this design. For instance, if the training group is the same after the training as before, but the control group is worse on the posttest than on the pretest, then the training was probably effective in preventing a decline in performance that would otherwise have occurred.

### 7.13 Evaluation Instruments

The literature is cluttered with suggested evaluation techniques ranging from simple questionnaires to complex statistical procedures. Often the one technique is presented under several different names, such as pre & post testing which is variously referred to as pre-then-post testing (Mezoff, 1981), the 3-Test Approach (Rae, 1983), and Time Series Analysis (Bakken and Bernstein, 1982).\(^1\)

---

The nature and type of organization exerts a subtle influence (possibly control?) over the scope and methods of evaluation, and the conduct of evaluation is also dependent on whether internal or external evaluators are used. The obvious constraint determining the type of evaluation chosen is the availability of resources. This includes time, money, and personnel, as well as the evaluator’s own expertise. Possibly the latter is the major constraint. Lange (1974) expresses similar concerns, stating, “too many bad evaluations are being presented ... evaluation is a good concept based on solid theoretical thinking. But its practice is not well developed”.

The literature review confirms the belief of Morris (1984) that evaluation is regarded by most practitioners as desirable in principle, difficult in practice. It also highlights the lack of well-written and documented articles for practitioners to learn from.

Some of the evaluation instruments which have not been discussed earlier in the chapter are discussed here:

7.13.1 Observations

A good trainer or tutor will be able to detect the reactions of learners by observing their behaviour and their comments. This method can generate useful feedback, but not in an objective or structured fashion.

The work behavior of trainees is observed before, during, and after training. A trained observer watches and records the behavior. Sometimes the behavior is videotaped or audio taped to be able to study later. This method provides evaluation of both verbal and nonverbal behavior.

The major disadvantages of observations include modified behavior of the participants as a result of the observer’s presence, poorly trained observers who collect unreliable data, and observations which are expensive and time-consuming. The impact of the observer can be minimized by carefully choosing observers, giving observers standard forms to fill out, and trying to minimize the presence of the observer. (Marrelli, 1993) This approach works best when participants are unaware that they are being observed. For

---

2 Shepherd, Clive, “Evaluating online learning”, www.fasttrackconsulting.com
example, to measure changes in customer service, an organization can hire shoppers to observe sales people.

The observers could include instructors, supervisors, co-workers or professional observers. They can watch out for the way in which learning is being applied. They can also provide positive reinforcement where the results are successful and constructive feedback and encouragement where they are not.

Observation of the trainees at the workplace must be structured, and the evaluation process must record it, which can give a proof of improved performance.¹

7.13.2 Surveys and questionnaires

These can capture data on participants’ accomplishments and behavioural changes after training. Responses can be collected from all participants or a sampling. Surveys and questionnaires are inexpensive, as well as easy to implement and tabulate.²

A questionnaire or survey is a printed or computerized form using questions such as multiple-choice, ranking scale, rating scale, or open-ended. Questionnaires/surveys can be given to both the learner (Level 1, Level 2) or could be given to supervisors to access on-the-job behavior changes (Level 3). They are most appropriate for level 3 (on-the-job application) and level 4 (business results).³

Postal questionnaires. As with other methods, there are advantages and disadvantages to the questionnaire. The advantages revolve around the fact that there is a high degree of objectivity. First, the people who fill in the questionnaire are less likely to be influenced by either a positive or a negative group dynamic, and you tend to get a truer impression of what people really think. Second, afterwards, when you are called to account for


405
your decision, you may find it useful to have the completed questionnaires there as proof of the soundness of the basis of your decision.¹

The disadvantage of a postal questionnaire is quite simply that people tend not to fill them in, in which case you can end up without the information you need. To make the task easier, you can arrange for this questionnaire to be completed face-to-face, so that you will have no trouble getting a good rate of return.

The questionnaires must contain a section where the respondents can say what they like in the way they like. It relieves the feeling of being straitjacketed which objective questions can give them.

‘Happy Sheets’. The most popular form of evaluations the issue of a questionnaire at the end of the program. The much-maligned end-of-event questionnaires have a lot to offer. First, it may the learners’ only chance to comment on environmental issues, by which we mean things like the beds (on residential courses), the blinds – or lack of them (in conference rooms) and the noise level (if working through a detailed text in a busy office). These factors all impinge directly on the learning and it is worthwhile your knowing about them. There is no point in looking for a design flaw in your training if all that was wrong was some stale sandwiches.²

‘Happy sheets’ also give the learners an opportunity to reflect on whether they enjoyed the training or not; and as enjoyment is such a key motivator, you need to know about it and measure it.

Online questionnaires. Web page forms can be used in much the same way as paper happy sheets.³ In fact, they have quite an advantage, in that the results can be automatically stored in a file on the web server, for analysis by a spreadsheet or database program. With a little programming, you can even have the analysis performed automatically.

³ Shepherd, Clive, “Evaluating online learning”, www.fastrakconsulting.com
7.13.3 One-to-one interviews

Interviewing participants individually is an excellent way to capture changes in job-related behaviours and to garner specific details. More versatile than questionnaires, interviews can probe issues, concerns, and actions related to the training.

A face-to-face interview involves an individual responding orally to oral questions asked by an interviewer. Interviews can be either structured or unstructured. Structured interviews consist of a list of predetermined questions. Unstructured interviews begin with standard questions but base subsequent questions on the interviewee’s responses to the previous questions. Interviews provide a means to collect in-depth information from participants who are reluctant to fill out a questionnaire. Interviews can be time-consuming and expensive and must be given by a skilled interviewer. Some participants may be less willing to reveal information in an interview. However, interviews (vs. surveys) are most likely to get people to “tell stories” and give specific illustrations.¹

Focus groups. An extension of interviewing, focus groups involve collecting post training information from 8 to 12 participants in a structured setting. Focus group members are asked specific questions about what they have changed as a result of training. The exchange of information often triggers creative thinking among the participants, which provide high quality data. They can be used to collect in-depth qualitative information.

Interviews with superiors. The trainees’ superiors can also be interviewed to find about changes in the trainees’ performance. Superiors should be asked objective questions about improvements they have noticed in the trainees’ knowledge, skills and attitudes. The questions should be tight scripted and restrict peoples’ answers using multiple-choice or number based formats.

7.13.4 Tests

Tests can be administered as a standardized method for measuring knowledge (paper and pencil test) and skills (performance test).

Pre- and Post-testing

To be able to measure training effectiveness, pre- and post tests are given to determine change after training. This is probably the single most useful evaluation technique if your aim is to measure the amount of change in learners’ knowledge, skills or attitudes. It provides a true measure of achievement.¹

With clear learning objectives, it is fairly simple to set the same or similar questions to test people’s knowledge before and after training. Equally, with skills or attitudes which have well-defined opinions or responses, learners could be asked to complete rating scales as a form of pre- and post-training test.

There is a potential bonus with pre-test. They help to focus the learners’ minds on the key messages of training. This will help the learners to pick up new ideas more quickly. Similarly, post-tests will help slower learners to measure the progress they have made.

Learning index

After obtaining the pre-and posttest scores, the amount of learning can be worked out by calculating the learning index for each trainee. The formula for learning index is:²

\[
\text{Learning} = \frac{\text{Post-training score} - \text{Pre-training score}}{100 - \text{Pre-training score}} \times 100
\]

More than the pretest score or the pre-training score itself, the unknown or unlearnt score (i.e., 100-pretraining score) is important since it is an indicator of the degree to which the trainee stands to gain through the training program. Hence the index of learning is based on this unknown portion which obviously varies from trainee to trainee.

7.13.5 Discussion of progress during the program

Reviewing progress and giving feedback has a motivational aspect in that it clarifies, for the learner, the gap between present and desired performance. It is necessary for the trainee to be aware of these gaps and be willing to do something about them if learning is to take place. It also has the function of establishing for the tutors, what learning has taken place and what is still needed. This information may also be necessary for the redesign of the later stages of the program.

One way of doing this is to have regular learning reviews, at which participants write down what they have found to be particularly interesting or useful during the session. They should also be asked to make a note of how they intend to use these and what they think will result from this use. The subsequent discussion can be as a group, where everyone selects something to read out, or one-to-one with tutors. The discussion should give a good feel for what is going well and what learning they are likely to take away.

The frequency of these reviews is a matter of judgment. In a program of some weeks’ duration, a review at the end of each week would seem to be appropriate. For a four-day workshop, each evening is a good frequency. Perhaps the most valuable purpose of these reviews is to focus on the usefulness of the learning during the program. The process can also give a basis for later evaluation of any changes. Those aspects that participants have found interesting and potentially useful can be grouped into clusters towards the end of the program to form an action plan. This takes them further around the learning cycle to ‘active experimentation’. The action plan can be used as a basis for establishing what they intend to do in six months after the program, and it can be later used to assess what they have actually achieved.

7.13.6 Action plans

The most powerful way to measure training’s effect is through the use of action plans. A good deal of adult behaviour is motivated by setting goals. People have plans for what they want to do, how far they want to be promoted, how they are going to impress senior people to achieve this, where they would like to be working in five years’ time, and so on. There is also little doubt that setting these goals effects performance, both in terms of
direction and effort. The greatest value of action planning is that it taps into these sources of motivation and can thus provide a bridge for transferring learning to the work situation.

Job Improvement Plan. Training aims at improving the performance of the executives. Measurement of the impact of training on the executive’s job performance is a difficult task. The Job Improvement Plan (JIP) is an action plan prepared by each trainee for improving his job performance on the basis of what has been learnt during the training period. It helps ascertain the impact of training on the trainee, his organization and the lessons that can be drawn by the training institution for improving the programs in terms of adequacy, quality and relevance of course content. Thus, the JIP helps in establishing a direct link between training and its transfer on the job. (Refer Annexure 7.6)

The JIP is one of the most important aspects of the training activity as it enables the trainees to concretize thoughts still fresh in their minds in terms of training inputs transferable to their jobs. The JIP also sets into motion a course of action which enables the executive to do his job better. It activates for the executive a process of thinking about the current or future problems in his job and how changes in procedures or methods will contribute to improved performance and profitability for the organization.

The JIP also forms the basis for the follow-up. It indicates the specific problems, if any, faced by the trainee which prevented him from being able to translate his knowledge into action. By following the action steps enunciated by the trainee, his organization is also able to link training with the job. Areas where no action steps have been indicated would be pointers to the probability that the training has been insufficient or inadequate.

Once the trainee returns to his workplace with his Job Improvement Plan – it could be discussed with his boss and if need be modified, changed, or certain aspects of the JIP could be deleted in the light of his organizational requirements. This step has the advantage of increasing the organization’s receptivity to new ideas that the trainee had acquired during training and also clarifies to the trainee, the scope and areas of improvement. The action plan can be signed as a ‘learning contract’ by each trainee at the end of the

---

training. It should, of course, be thoroughly discussed with the trainee’s superior on return, and this would probably be the first action listed on it.

If the trainers are really interested in studying the application of the training to the job then the progress has to be monitored after a sufficient time gap. A questionnaire can be designed to suit the purpose which basically checks the implementation of the trainees’ JIP. The follow-up some six months later, would ask such questions as:

- How much of your action plan have you been able to implement?
- Which actions are still likely but now need a longer time frame?
- Which actions have been shelved and why?
- What positive organizational benefits have come from your actions?

This method of data collection is combined with interviews and case studies. This kind of follow-up analysis helps compare the actual gains from the training with the expectations of the trainees and their sponsoring organization.

7.13.7 Performance Records

Without delving into the realms of statistics in a big way, you can ascertain how suitable a given piece of training is for learners from different sub-groups of the population. All you need to do is keep records of the levels of performance pre- and post-training, and then look at these data in the light of the qualification, age, and experience of the learners. You will soon see which groups are benefiting most and which least from your training.

Performance records can be used to evaluate a training program’s effect on the company’s bottom line. Data such as costs incurred, amounts produced, revenue generated, or time required to complete tasks, would be measured both before and after training to be able to quantify the effects of training. Any measurable savings to the company could be compared with the actual cost of delivering training.

This approach requires examining the organization’s overall performance data to obtain before-and-after comparisons of each data item. Because it

---

can provide the most convincing evidence, it’s often the preferred approach of senior managers.¹

7.13.8 Follow-up assignments

Perhaps the easiest approach to post training data collection is to ask participants to complete a task or a project, to serve as evidence of their successful application of the training content. Typically assignment results are reported to participants’ superiors. This approach is essentially helpful for level 3 evaluations.²

7.13.9 Special follow-ups

It can be effective to reconvene participants one to three months after the initial training segment was conducted so they can report on their success. Follow-up sessions also provide opportunities for additional training such as refining new skills. This approach is appropriate for level 3 and level 4 evaluations.³

7.13.10 Follow-up visits

The trainer can visit the trainees at their workplace, some weeks after the training. Merely visiting your trainees at work isn’t going to achieve much. But if you visit them:

- With a wide range of data gleaned from interviews, questionnaires and assessments carried out under training conditions, and
- With a view to confirming and updating that data by further evaluative work with themselves and their managers,

Then a follow-up visit can achieve a lot.⁴

---

³ Ibid
7.13.11 Performance contracts

These contain a training agreement between participants, their superiors, and sometimes the facilitators. The parties meet prior to training in order to develop measurable goals related to the training content. Later, they can determine whether the goals were met. (Refer Annexure 7.7)

7.13.12 Portfolio of learner work

Another assessment tool is the portfolio of learner work. Portfolios often include samples of class work, checklists where learners rate their progress in basic and workplace skills, and journals where they record their reactions to class and workplace activities. Like interviews these measures can provide vital information on learner attitudes and concerns. They are also a venue for self-assessment, and allow participants who are unable or unwilling to express themselves orally, or who have difficulty with formal tests, to demonstrate progress towards their goals.

7.13.13 Using behaviour scales

Change in behaviour and on the job performance can also be evaluated by using behaviour scales. These scales list desirable actions required from a manager, for which he is evaluated both before and after training.

Suppose, for example, that we want to change the management style in our organization to a more consultative one. We would then need to focus on the behaviours consistent with a more consultative style. For instance, we might try to increase the frequency of doing things such as target setting and discussing priorities, asking for views before making decisions and delegating authority.

One way of doing this is to define, by listing examples of relevant behaviour, what is meant by the desired approach and then to assess to what extent this is being achieved. The people best able to carry out this assessment of middle managers are the junior managers who report to them. Some intervention would then be made, based upon feeding back the subordinates’ views to the middle managers and then offering opportunities

---

1 Miriam, Burt. & Saccomano, Mark. “Evaluating workplace ESL instructional programs”, ERIC digest
to attend workshops on the skills that they wished to develop. A possible set of such behaviours is:

- Communicates company objectives to you
- Jointly sets clear targets with you
- Jointly reviews progress on targets at timely intervals
- Lets you know exactly what is expected of you
- Delegates sufficient authority and responsibility to you
- Accepts full responsibility for your actions even when things go wrong
- Encourages you to suggest new ideas on how things could be done
- Asks for and uses your input in decision making
- Expresses appreciation personally when you do well
- Shows genuine interest in you and your work
- Demonstrates concern about your development
- Coaches and guides effectively
- Is available and spends time with you when needed
- Encourages good relationships among his staff
- Knows when individuals have problems, is helpful and supportive.

The subordinates would be asked to apply each of the 15 statements to the behaviour of their own departmental head and mark each statement ‘agree = 1’, or ‘tend to agree = 2’ and so on, to ‘disagree = 5’. This is best done in a way that allows subordinates to give their views anonymously. The opinions should be collated and used for individual feedback to the managers whose behaviour is being reviewed. The departmental heads thus learn of their subordinates’ views on their management styles, discuss these (with a consultant or with each other) and agree upon an action plan or learning contract. Opportunities to attend short workshops on joint target setting, holding team meetings and developing subordinates would be offered. Six months or a year later, the process would be repeated and differences noted. (Refer Annexure 7.8)

7.14 Quantifying qualitative measures

To increase credibility and help ensure reliability of qualitative measures, evaluators can collect multiple types of evidence such as interviews and observations from various stakeholders around a single outcome. Data collected from the various measures can then be arranged into matrices. This
chart like format organizes material thematically and enables analysis of data across respondents by themes.

Questionnaires and interview data can be quantified by creating a scale that categorizes responses and assigns them a numeric value. Improvement in such subjective areas as worker attitudes can then be demonstrated in a numeric or graphic form.¹

7.15 Summary of Evaluation Instruments

To collect accurate information with evaluation instruments, you need a basic knowledge of statistics and research methods. You need to know how to use various instruments and be able to select the most appropriate instrument for each evaluation. Multiple instruments should be used in any evaluation. Each instrument has inherent strengths and weaknesses. Multiple instruments can compensate for the weakness in another instrument or complement the strengths in another. Multiple instruments also provide more credibility and may produce different results that could be missed with a single evaluation instrument.²

Conducting evaluations, particularly Level 3 and Level 4 evaluations, can seem overwhelming, and if it is done as an afterthought it can be very difficult to do. However, it is much easier to design evaluation into the course as the course is being developed. Evaluation must be plotted while the training course is still a fresh idea. The following evaluation matrix worksheet could used to assist in the design of the evaluation.³

¹Miriam, Burt. & Saecomano, Mark., “Evaluating workplace ESL Instructional programs”, ERIC digest
### Evaluation Matrix Worksheet

<table>
<thead>
<tr>
<th>Levels</th>
<th>What might be measured?</th>
<th>What are the data sources?</th>
<th>How should data be collected?</th>
<th>What are potential problems?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Reaction)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Learning)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Behavior on job)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Results)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table: 7.4 Evaluation matrix worksheet**

Evaluation instruments need to be selected based on the design of the training program, as discussed earlier. Before selecting an evaluation instrument, the following should be considered:

- Will the instrument answer your question? The instruments you select must be appropriate for the questions you ask.
- Does the instrument suit the evaluation design?
- Is the instrument valid? The instrument selected must accurately measure course objectives.
- Is the instrument reliable? The instrument must provide consistent information.
- Is the instrument practical? Consider reading and vocabulary levels of printed materials and test; consider time and monetary resources needed to produce the evaluation.
In a survey conducted by the ASTD on evaluation practices, the following was found to be the usage of various evaluation instruments in industry:

<table>
<thead>
<tr>
<th>Tool/Technique</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Questionnaires</td>
<td>94%</td>
</tr>
<tr>
<td>Management/peer feedback</td>
<td>74%</td>
</tr>
<tr>
<td>Interviews</td>
<td>69%</td>
</tr>
<tr>
<td>Tests</td>
<td>58%</td>
</tr>
<tr>
<td>Observation</td>
<td>57%</td>
</tr>
<tr>
<td>Attitude surveys</td>
<td>49%</td>
</tr>
<tr>
<td>Focus groups</td>
<td>48%</td>
</tr>
<tr>
<td>Testimonials</td>
<td>42%</td>
</tr>
<tr>
<td>Performance records</td>
<td>40%</td>
</tr>
</tbody>
</table>

Table: 7.5 Usage of evaluation instruments

---

7.16 New Directions in Training Evaluation

When people ask Laurie Bassi, vice president for research at Alexandria, Virginia based ASTD, what to consider in calculating an ROI, she answers, “The first thing to consider is: Is it worth doing?”\(^1\) Before embarking on your ROI quest, you should first contemplate the ROI of calculating your ROI. Is it worth the time and money you will spend? The answer says Bassi depends on the type of training you are doing.

The ideal type of training suited for calculating ROI is one-time training on a specific skill – teaching customer-service staff the latest computer program vs. putting the executive team through a series of leadership seminars. The former is a discrete module in which the outcome can be readily tested in a before-and-after scenario. This is the easiest and most clear-cut way to calculate an ROI. The latter, although potentially will have a greater effect on the company, is more difficult to quantify. It’s hard to do a before-and-after test on leadership skills, and other factors will come to play in the trainees’ performance.

The four-level approach may work better for managerial and soft skills training. Consider a seminar in sales skills, for instance. To test learning, you could have participants explain – before and after the training – how they’d approach different sales opportunities. To test application to the job, you could track sales in the months before the training and after. To show business results, you use the sales figures – acknowledging that other factors may have affected them. However in such scenarios the focus is best placed on value and results, rather than a number-based ROI. In managerial skills training, you’ll have to use subjective judgments to determine the effectiveness of the training. Some other ways would have to be used to assess the value of training in these kinds of cases, rather than ROIs.

Indeed, some companies are turning away from calculating returns on investment, and are measuring return on expectations, instead. In this kind of approach, those who are involved decide exactly what they expect to achieve from the training. This set of expectations becomes the baseline for determining success. Months after the training, the stakeholders review their agreed-upon expectations. They then decide if the results are in line.

This approach allows for more anecdotal analysis. If a group of managers completed a communications-training course, stakeholders can then discuss the ways they feel communication has improved. This provides a more realistic picture than if HR were forced to place a convoluted dollar amount on improved communications.

Toshiba’s Electronic Imaging Division, California, is one company that made a conscious decision to focus on the value of training rather than the ROI of training. Because the company made a shift to multifunctional products, dealers had to make a shift too. In October 1996, training on selling the new products began. “We had to educate the dealers to sell our products,” says the director of training and dealer development. “Training is an investment. Obviously, how we estimate return is on increased sales. But we don’t have a one-to-one correlation. We feel it’s an investment in our dealers to help them move towards the 21st century.” The company has been able to double its market share, although it doesn’t credit training alone with that success. Rather, the company treats training as one key piece in an overall business strategy.¹

If your training lends itself to a simple calculation of ROI, doing so gives it more credibility. But if your training is too abstract to support ROI, your time and money would be better spent elsewhere. ROI is a means by which to measure the value of training, there are other ways too – returns on expectations, ROI thresholds and anecdotal information – to evaluate the training’s success. Choose the one that will work best for the training you are doing. Ultimately, that is the best ROI.

Thinking outside the box about training evaluation isn’t easy. But some organizations, such as Andersen Consulting Education, are doing that.² Several years ago, the staff created an evaluation method that rejects dollar-based ROI and the idea that business benefits from training can be isolated from other factors and proved. Andersen’s method – Concept Mapping and Pattern matching – uses stakeholder expectations to design and evaluate training courses. It’s based on the idea that managers know what skills and behaviours their employees need to do their jobs, and that training them in those areas will produce bottom-line results.

² Abernathy, Donna J., “Thinking Outside the Evaluation Box”, www.astd.org
William M.K. Trochim, a professor at Cornell University says that education evaluation must encompass methods that can deal with psychological and sociological complexity of the education enterprise. Such methods need to be able both to measure value in the development and delivery of a complex and communicate clearly the results to businesspeople not necessarily interested in the fine points of complex statistic and research design.

General Electric CEO and chairman Jack Welch, believes in investing more dollars in training than on detailed evaluation. GE uses surveys to realign training, as necessary, but the major emphasis is on training, not its measurement. Returns come in the form of tangible and intangible business results, increased consumer satisfaction, and career development for GE workers. That, to Welch, is great evidence that training has value.

Most experts advocate that training evaluation should be kept simple. When evaluating training, businesses should apply the same common-sense values to internal customers as to external customer. Businesses should evaluate training in terms of whether the trainees were educated, amused, and enriched. Did they receive the learning they needed to do their jobs better? Was it an enjoyable experience, such that they were motivated to learn and inspired to go apply the learning? Were they, the company, and the company’s customers enriched by the learning, in that it translated into improved skills and performance on an individual and company level?

Joe Furando, manager of organization development at Merk-Medco, also casts a vote for simplification. He says, “Too often, trainers feel that they need to commission lengthy and often costly research projects to determine the value and effect of their training. I have found that for the majority of clients, simpler is better. A full-blown study that strives to determine statistical correlations to performance metrics or the number of standard deviation gains in tests is fine for the training researcher. For most clients, its sufficient to explain how competencies from a model are related to performance metrics and that by going through the training, gains in participant performance can be directly tied to the training.”

---

1Abernathy, Donna, J., “Thinking Outside the Evaluation Box”, www.astd.org
Evaluating training is a matter of determining its value, and that's a matter of people's perceptions. Fred Nickols, executive director of strategic planning and management services at the Educational Testing Service, identifies the people who matter: trainees, trainers, supervisors of the trainees, and senior managers and executives. He says that there is no real ROI for training. “In strict accounting terms, it's an expense, not an investment. Moreover, it's often an act of faith. And faith is confirmed by the consequences of our deeds, not by financial rewards. The bottom line? If you want to measure training, determine the impact it is having on people, and through them, on processes. In short operations. Then with particular emphasis on supervisors, managers and executives, determine the perceived value of those impacts. That will tell you what training is worth.”

When deciding which programs to evaluate for ROI, develop criteria for what kinds of training are the most important for you to track. These may include the more expensive or controversial programs, programs that run frequently or involve a lot of people, programs that are linked to crucial strategic objectives, or any other programs for which you especially want to measure impact.

The further down the five-level model, you evaluate any given training program, the more information you will have on whether it is working well and if not, why. This will help you make informed decisions on which programs to continue, and give you a better idea of the potential of future programs. However, evaluations do take planning, time, and money, so remember that it's not necessary to evaluate every training program through all four (or five) levels. A good general rule of thumb might be to evaluate all training for reactions, but only 50% to 70% for learning, 30% for behavior, 10% to 15% for results, and 5% for ROI.

---

1 Abernathy, Donna J., “Thinking Outside the Evaluation Box”, www.astd.org
2 Nancy Chase, Quality Online September 1997 Special Training Issue,
Annexure 7.1

The comment sheet shown below was used to measure reaction at an ASTD summer institute that was planned and coordinated by the staff of the Management Institute of the University of Wisconsin.¹ Those who planned this ASTD program were interested in reactions to subject, technique (lecture versus discussion), and the performance of the conference leader. Therefore, the form was designed accordingly. So that the reactions could be readily tabulated and quantified, the trainees were asked to place a check in the appropriate spaces.

ASTD INSTITUTE

Leader

Subject

Date

1. Was the subject pertinent to your needs and interests?
   No          To some extent    Very much so

2. How was the ratio of lecture to discussion?
   Too much lecture    O.K.    Too much discussion

3. Rate the leader on the following:

<table>
<thead>
<tr>
<th></th>
<th>Excellent</th>
<th>Very good</th>
<th>Good</th>
<th>Fair</th>
<th>Poor</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. How well did he state</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>objectives?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B. How well did he keep the</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>session alive and interesting?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. How well did he use the</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>black board, charts, and other aids?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D. How well did he summarize</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>during the session?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E. How well did he maintain a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>friendly and helpful manner?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F. How well did he illustrate and</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>clarify the point?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G. How was his summary at the</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>close of the session?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

What is your overall rating of the leader?

Excellent      Very good      Good      Fair      Poor

4. What would have made the session more effective?

Signature (optional)
Annexure 7.2

Occasionally it is felt that the group reaction is not a fair enough evaluation of the effectiveness of the program. Sometimes the trainer's personality makes such an impression on the group that he may receive a very high rating. In other sessions, the trainer may receive a low rating because he does not have a dynamic personality. Therefore it is suggested that along with the evaluation done by the trainees an evaluation should also be done by the training coordinator, independent of the group's evaluation. A comparison of the two would give the best indication of the effectiveness of the program.

COORDINATOR'S RATING OF LEADER

<table>
<thead>
<tr>
<th>Rating</th>
<th>Name of Leader</th>
<th>Subject</th>
<th>Date:</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>A. PREPARATION</th>
<th>Very much so</th>
<th>To some extent</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Did he prepare for the meeting?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Was his preparation geared to the group?</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>B. CONDUCTING</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Did he read his material?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Did he hold the interest of the group?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Was he enthusiastic-dynamic?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Did he use visual aids?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Did he present his material clearly?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Did he help the group apply the material?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Did he adequately cover the subject?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Did he summarize during conference and at end?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Did he involve the group?</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>C. CONSTRUCTIVE COMMENTS</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1. What would you suggest to improve future sessions?</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>D. POTENTIAL</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1. With proper coaching what would be the highest rating he could achieve?</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>E. ADDITIONAL COMMENTS</th>
<th></th>
</tr>
</thead>
</table>
Annexure 7.3

The forms and suggestions that have been described above will apply to internal training programs. Since many companies send their management people to outside training programs at universities, for conferences, etc, it is suggested that the reaction of each person attending such a training program should also be measured

REACTION TO SUPERVISORY INSTITUTE BY FOREMEN AND SUPERVISORS WHO HAVE PARTICIPATED

IN GENERAL

1. How worthwhile was the Institute for you?
   a) Very worthwhile
   b) Fairly worthwhile
   c) Not very worthwhile
   d) A waste of time

2. The Institute had:
   a) Too much theory and not enough of the practical
   b) Too much of practical and not enough theory
   c) About the right combination of theory and practical

HOW THE INSTITUTE WAS CONDUCTED

3. On the whole, the course was conducted
   a) Very well
   b) Fairly well
   c) Poorly
   d) Very poorly

4. Lecture and discussion
   a) Too much lecture
   b) Too much discussion
   c) About right amount of each

5. Discussion leaders
   a) Too many from the university
   b) Too many from business and industry
   c) O. K.

6. Visual aids
   a) Not enough movies, charts etc.
   b) Too much use of demonstrations, blackboards, movies, charts etc.
   c) O. K.

---

APPLICATION OF THE COURSE
7. Did the Institute apply to your particular operations?
   Yes   Partly   No

FOLLOW UP
8. Would you like to attend another Institute?
   Yes   No
9. These Institutes should run for 5 days 4 days 3 days
10. Please list 3 of your main problems:
    1.
    2.
    3.
11. Comments or suggestions
Annexure 7.4

Supervisory Inventory on Human Relations

PLEASE ANSWER ALL STATEMENTS EVEN IF YOU ARE NOT SURE

1. Anyone is able to do almost any job if he tries hard enough. A DA

2. Intelligence consists of what we’ve learned since we were born. A DA

3. If a supervisor knows all about the work to be done, he is therefore qualified to teach others how to do it. A DA

4. A well trained working force is a result of maintaining a large training department. A DA

5. In making a decision, a good supervisor is concerned with his employees feelings about the decision. A DA

6. The supervisor is closer to his employees than he is to management. A DA

7. The best way to train a new employee is to have him watch a good employee at the job. A DA

---

Annexure 7.5

AT&T has incorporated into their personal factors in Management Program a short test measuring trainee sensitivity and empathy. First, each individual is asked to rank, in order of importance, 10 items dealing with human relations. The participants are then assigned to groups which work 15 minutes at the task of arriving at a group ranking of the 10 statements. Following this 15-minute “heated discussion”, each individual is asked to complete a short inventory. The successive class sessions then attempted to teach the participants to be more sensitive to the feelings and ideas of other people. Later in the course, another “empathy” test was given to see whether there was an increase in sensitivity.\(^1\)

1. (a) Were you satisfied with the performance of the group?

   Yes ______  No ______

(b) How many will say that they were satisfied with the performance of the group?

   ______

2. (a) Do you feel that the discussion was dominated by two or three members?

   Yes ______  No ______

(b) How many will say that they thought the discussion was dominated by two or three members?

   ______

3. (a) Did you have any feelings about the items being ranked that, for some reason, you felt it wise not to express during the discussion?

   Yes ______  No ______

(b) How many will say that they had such feelings?

   ______

4. (a) Did you talk as often as you wished to in the discussion? Yes ___ No ___

(b) How many will say that they spoke as often as they wished? ___

Annexure 7.6

Action Plan for Training Transfer

Subject: (The specific area(s) you have picked for improvement)

Objectives: (What is to be accomplished, the purpose of the plan)

Goals: (The specific targets by which you will measure progress)

Problems: (The barriers that might hinder you in carrying out your plan)

Solutions: (How you plan to avoid or deal with the problems)

Activities and Time: (What actions, sequence, and time are needed)

<table>
<thead>
<tr>
<th>Step</th>
<th>Action</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Resources: (What people, time, equipment, etc. you need to carry out the plan)

People: 
Time: 
Equipment: 
Other: 

Costs: (What the overall cost of implementing your plan will be)

Benefits: (What benefits you expect and their estimated dollar value)

<table>
<thead>
<tr>
<th>Benefit</th>
<th>Estimated dollar value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Commitment: (When you and your manager will next review progress)

I will meet with my manager to review the progress of this action plan.

Time: Date: Location:

1 Source: Paul L. Garavaglia, The ADDIE Group Inc., 57 Dennison St., Oxford, MI 48371
### A Learning Contract

<table>
<thead>
<tr>
<th>Objective set</th>
<th>Strategy for achieving the objectives</th>
<th>Criteria and means of evaluating progress</th>
</tr>
</thead>
<tbody>
<tr>
<td>Objective 1</td>
<td>How you intend to do it</td>
<td>How you intend to measure achievement</td>
</tr>
<tr>
<td>Objective 2</td>
<td>How you intend to do it</td>
<td>How you intend to measure achievement</td>
</tr>
<tr>
<td></td>
<td>How you intend to do it</td>
<td>How you intend to measure achievement</td>
</tr>
<tr>
<td>Etc</td>
<td></td>
<td>Signed</td>
</tr>
<tr>
<td>Etc</td>
<td></td>
<td>Signed</td>
</tr>
<tr>
<td>Date:</td>
<td>Signed</td>
<td>Signed</td>
</tr>
</tbody>
</table>

### A Questionnaire for Following up Knowledge-Based Programmes (Transfer of training)

<table>
<thead>
<tr>
<th>Topic (A detailed list of the areas covered on the program)</th>
<th>How useful is knowledge of this for your job?</th>
<th>Have you used knowledge in this area since the course?</th>
<th>Have you had any difficulty in applying this?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Very</td>
<td>Quite</td>
<td>Not</td>
</tr>
<tr>
<td>Topic 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Topic 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Topic 3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Etc</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---


Annexure 7.8

A behavioural scale for positive management

<table>
<thead>
<tr>
<th>Positive management</th>
<th>Never 0-19%</th>
<th>Seldom 20-39%</th>
<th>Sometimes 40-59%</th>
<th>Generally 60-79%</th>
<th>Always 80-100%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thinks ahead &amp; develops plans rather than constantly clearing up problems</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grasps the essential nature of the problem, knows what information is necessary and where/how to get it</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thinks in terms of objectives rather than vague generalizations, &amp; makes them both clear &amp; realistic</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Takes decision rather than procrastinating or passing problems up to the next level</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Is consistent about, &amp; effective in, obtaining high productivity in the short and longer term</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Acts as a model for the group, being firm, getting commitment and encouraging participation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Co-ordinates the group’s activities and checks on progress to achieve objectives</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deals with subordinates as individuals and makes each accountable for a specific set of responsibilities</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Knows what to delegate and has the courage to risk errors by subordinates</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Minimises immediate pressure & problems, & maximizes long term productivity |
| Sets high standards and gets them |
| Knows what is wanted and how to get it without resentment |
| Seeks increased value for money and year on year improvements in efficiency |
| Rewards outstanding performance |
| Makes opportunities to develop people as individuals |