

Evaluating Training

*There is no “cookbook”
approach*

Fred Nickols

© Fred Nickols 2003
All rights reserved



www.nickols.us
fred@nickols.us

This is a close-to-the-original version of an article prepared for a 1992 *ASTD Tool Kit* edited by Karen Medsker and Don Roberts. The original paper I prepared was published in three separate pieces. This one is more or less intact.

Evaluate What and Why?

Evaluate? Evaluate what? Training? What do we mean by training? What's to be evaluated? A particular training course? The trainees? The trainers? The training department? A certain set of training materials? Training in general?

More to the point, why evaluate it? Do we wish to gauge its effectiveness, that is, to see if it works? If so, what is it supposed to do? Change behavior? Shape attitudes? Improve job performance? Reduce defects? Increase sales? Enhance quality?

What about efficiency? How much time does the training consume? Can it be shortened? Can we make do with on-the-job training or can we completely eliminate training by substituting job aids instead?

What does it cost? Whatever it costs, is it worth it? Who says? On what basis? What are we trying to find out? For whom?

The preceding questions illustrate the complexity of any effort to evaluate training and emphasize the importance of being clear about the purposes of and the audiences for any such evaluation.

It is the central thesis of this article that the evaluation of training poses a problem for many trainers, managers, executives, and other professionals with an interest in training. Further, it is my firm conviction that these problems are most productively addressed by examining their underlying structure. As Dewey (1910) wrote, "A difficulty clearly apprehended is likely to suggest its own solution (p. 94)". This article, then, will examine various elements in the structure of the problem of evaluating training.

The centerpiece for the collection of articles comprising the *ASTD Tool Kit* for which this paper was originally written is Donald Kirkpatrick's well-known framework for evaluating training, frequently referred to as "Level One," "Level Two," and so on. Much has changed since Kirkpatrick's framework first appeared and it might help to better understand and appreciate the truly seminal nature of his work if we attempt a very brief review of some of the major changes in the training and development world since then.

A Brief Historical Perspective: 1960-1990

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. These articles can be found in *Evaluating Training Programs*, a collection of articles compiled by Kirkpatrick from the pages of the *ASTD Journal* and published by ASTD in 1975.

In 1959, when Kirkpatrick launched his views, the American Society of Training Directors (ASTD) was about as close-knit a "good old boys" network as one could find. Since its inception in the 1940s, ASTD membership had consisted primarily of training directors, known also as training officers. Even as late as 1969 (the year in which I took up the training profession), ASTD

was still dominated by training directors. That the members of ASTD were in fact "old boys" is amply demonstrated by some figures from the 1969 ASTD national conference, which was held in Miami, Florida (Reith, 1970): Only nine percent of the attendees were 29 years of age or younger. Fully 59 percent were 40 years old or older. Only nine percent of the attendees were females. To elucidate the obvious, 91 percent were males. Any group consisting of more than 90 percent males past the age of 40 certainly seems vulnerable to charges of being a bunch of "good old boys."

Changes, however, were already evident. Of the 1,081 full-time attendees filling out the Miami conference feedback form, almost half or 49 percent were attending their first ASTD national conference. More than 77 percent had been in training assignments for more than three years and roughly 40 percent had been in training assignments for more than 10 years. But, at the same time, more than 50 percent of those attending had been in their present jobs for less than three years.

Elsewhere, the training business was stirring. The likes of Bob Mager, Susan Markle, Tom Gilbert, Geary Rummler, Joe Harless and Karen Brethower were shaking up the training establishment and would continue to do so for several more years. The development business was stirring too. Rensis Likert, Chris Argyris, Douglas McGregor, and George Odiorne were shaking up the management mindset and a new term had entered our vocabulary: "Organization Development (OD)."

The board of governors of the American Society of Training Directors, perhaps sensing some kind of shift in the tide of human and organizational affairs, changed the name of the society from the American Society of Training Directors to the American Society for Training and Development, and moved its headquarters from Madison, Wisconsin to the Washington, D.C. area (Alexandria, Virginia).

Other changes affecting the training and development worlds were taking place during this same time period. Behaviorism flowered for a while then wilted in the face of the shift to knowledge work. Peter Drucker, in book after book, beginning with *Landmarks for Tomorrow* (1959) and continuing through *The New Realities* (1989), kept reminding us that the center of gravity in the employed workforce was shifting from those who worked with their muscles to those who worked with their minds. By 1980, the shift to knowledge work was more or less complete and, three years later, I spelled out some of its consequences for training and trainers (Nickols, 1983).

As perceptions of the locus of working gradually and painfully shifted from the workers' muscles to their minds, the focus of managerial control over work and working shifted from the exercise of direct control over overt physical behavior to a search for ways and means of influencing covert mental processes. In short, the cognitive view gained sway (and it is likely to hold sway for the foreseeable future). Nevertheless, behaviorism, mostly through the efforts of Bob Mager, did give us this central question pertaining to the evaluation of training: "What is the trainee supposed to be able to do as a result of training?" – and the training business hasn't been the same since.

Programmed instruction blossomed for a while too, and was then displaced by its own progeny: self-instructional materials, job aids, and performance technology. Another society, the National Society for Programmed Instruction (NSPI), moved its headquarters from San Antonio, Texas to Washington, D.C., and changed its name to the National Society for Performance and Instruction. (It has most recently become the International Society for Performance Improvement.)

Systems concepts and the systems approach came rushing at us from two very different angles. We didn't stand a chance; we were overwhelmed by superior forces. Systems engineering, apparently obeying the biblical command to be fruitful and multiply, gave us the systems approach to this, that, and the other. Its primary legacy consists of (1) the instructional systems development (ISD) model originally developed in the military and (2) the computer systems development process found throughout business and industry.

General systems theory (GST) was fertile and prolific too, mostly on the organizational side of things. The concepts of "open" and "socio-technical" systems came into vogue and stayed. "Systems thinking" is with us still, so pervasive now that we hardly give it a second thought. Human relations was a burgeoning movement in this same period. Years earlier, Elton Mayo had given us the "Hawthorne effect" and, in the 1960s and 1970s, his legatees gave us sensitivity training, T-groups, and organization development (OD). One of Mayo's philosophical descendants, Len Nadler, coined the term "human resources" and people haven't been looked upon as people since.

Technology was at the heart of much of what was going on from 1960 through 1990. For 10 of those years (1965 to 1975) a brief war was waged between "educational technology" and "instructional technology." It was a civil war, of course, and like a lot of recent wars it ended in a draw; there weren't any clear-cut winners, but at least the hostilities came to an end.

Donald Kirkpatrick's four-level evaluation framework has survived all this turbulence. One might even say that it has prospered. At the very least, one must acknowledge its staying power – and rightly so, for, although his framework might not be the last or latest word on the evaluation of training, it certainly comes close to being the first word on the subject.

Let us now shift our focus from the past to the present and begin our examination of the evaluation of training problem. Our starting point is with the structural relationship between training and the workplace.

Training and The Workplace

Most training takes place in an organizational setting, typically in support of skill and knowledge requirements originating in the workplace. This relationship between training and the workplace is illustrated in Figure 1.

Using the diagram in Figure 1 as a structural framework, we can identify five basic points at which we might take measurements, conduct assessments, or reach judgments. These five points are indicated in the diagram by the numerals 1 through 5:

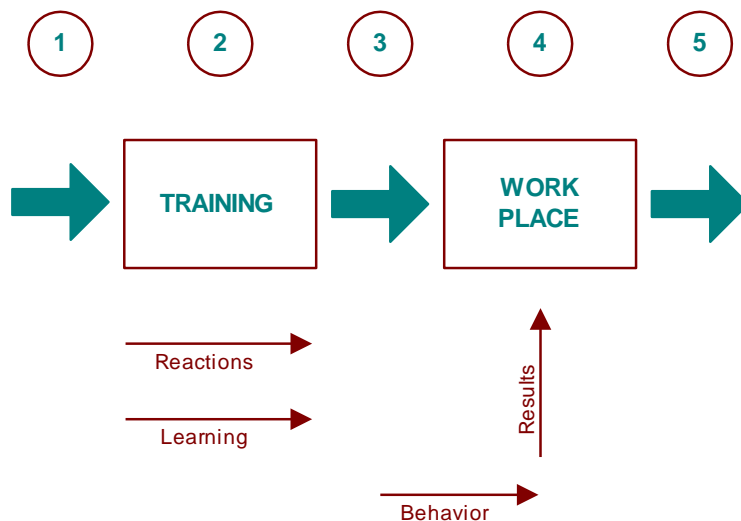


Figure 1 - The Structure of the Training Evaluation Problem

1. Before Training
2. During Training
3. After Training or Before Entry (Reentry)
4. In The Workplace
5. Upon Exiting The Workplace

The four elements of Kirkpatrick's framework, also shown in Figure 1, are defined below using Kirkpatrick's original definitions.

Reaction. "Reaction may best be defined as how well the trainees liked a particular training program." Reactions are typically measured at the end of training – at Point 3 in Figure 1. However, that is a summative or end-of-course assessment and reactions are also measured during the training, even if only informally in terms of the instructor's perceptions.

Learning. "What principles, facts, and techniques were understood and absorbed by the conferees?" What the trainees know or can do can be measured during and at the end of training but, in order to say that this knowledge or skill resulted from the training, the trainees' entering knowledge or skills levels must also be known or measured. Evaluating learning, then, requires measurements at Points 1, 2 and 3 – before, during and after training.

Behavior. Changes in on-the-job behavior. Kirkpatrick did not originally offer a definition per se for this element in his framework, hence I have not enclosed this one in quotation marks. Nevertheless, the definition just presented is taken verbatim from Kirkpatrick's writings – the fourth and final article. Clearly, any evaluation of changes in on-the-job behavior must occur in the workplace itself – at Point 4 in Figure 1. It should be kept in mind, however, that behavior changes are acquired in training and they then transfer (or don't transfer) to the work place. It is deemed useful, therefore, to assess behavior changes at the end of training and in the workplace. Indeed, the origins of human performance technology can be traced to early investigations of disparities between behavior changes realized in training and those realized on the job. The seminal work in this regard is Karen Brethower's paper, "Maintenance: The Neglected Half of Behavior Change" (see the references section).

Results. Kirkpatrick did not offer a formal definition for this element of his framework either. Instead, he relied on a range of examples to make clear his meaning. Those examples are herewith repeated. "Reduction of costs; reduction of turnover and absenteeism; reduction of grievances; increase in quality and quantity or production; or improved morale which, it is hoped, will lead to some of the previously stated results." These factors are also measurable in the workplace – at Point 4 in Figure 1.

It is worth noting that there is a shifting of conceptual gears between the third and fourth elements in Kirkpatrick's framework. The first three elements center on the trainees; their reactions, their learning, and changes in their behavior. The fourth element shifts to a concern with organizational payoffs or business results. We will return to this shift in focus later on.

Thinking about the Evaluation of Training

The diagram shown in Figure 1 not only depicts Kirkpatrick's evaluation framework, it also indicates the points at which it takes measurements, collects data, and so forth. We can create other possibilities for evaluating training by altering the points at which these same measures are taken.

Trainee reactions, for instance, could be assessed at Point 4, after the trainees have been on the job for a while, instead of so soon after the completion of training. In a slightly different vein, we could compare Points 2 and 4, which essentially amounts to comparing the training environment with the workplace environment. From such a comparison we might be able to gauge the "authen-

ticity" of the training, that is, how closely the training environment matches or resembles the workplace environment and, from this, draw some conclusions about the likelihood of a phenomenon known as the "transfer of training."

We can "get outside the box," so to speak, and pick points not even shown on the diagram.

Moving all the way to the left of Point 1, for instance, we can speculate that trainees arrive at Point 1 as a result of some kind of selection process. In the course of evaluating training, we (or someone else) might wish to measure the effect selection has on success in training.

Moving all the way to the right, beyond Point 5, we can inquire as to where people go when they leave the workplace, perhaps at the end of the day or perhaps at the end of a career. One answer is that they go home. Another is that they reenter the larger community in which the organization is embedded and from whence they came.

From this perspective, one might ask, "What good or harm comes to the community as a result of the organization's training and workplace practices?" Alternately, "Is the organization turning out skilled, self-supporting members of the community, or is it simply chewing up people and spitting out dull-eyed, unthinking, uncaring automatons who are of no further value to themselves or to society?" In short, by moving all the way to the right in Figure 1, we begin examining the societal impact of organizations – and of the training they provide – or don't provide, as the case may be.

Another way to make use of the structure depicted in Figure 1 is to change the time perspective being used. Kirkpatrick's "Reactions" element is a retrospective or after-the-fact view. The trainees are looking back at the training (to the left from Point 3). Why not substitute a perspective of looking forward? At Point 3, the notion of looking forward raises the possibility of asking the trainees to provide their predictions regarding the nature of the workplace they're about to enter. In other words, we might consider assessing the image of the company and the workplace that is communicated by the training experience.

As seen earlier, learning is typically assessed through before and after measures. This is a point-to-point measurement and comparison; it spans a "chunk" of the framework. By varying the points used, we can identify other "chunks" and come up with other evaluation issues. We could, for instance, create a span encompassing all of Figure 1 – Points 1 through 5 – and this might suggest larger learning issues that involve training and development in an integrated fashion. How do training and workplace developmental experiences dovetail, for instance, in mapping out career paths?

Create a span from Points 1 through 3, the same span used in gauging learning, but take the perspective of the manager of the people who are going through training. A couple of likely evaluation issues from this perspective can be expressed in two terse questions: "How long is it going to take? What is it going to cost?"

Let's pick yet a different audience for the evaluation of training: The professional training community. And let's use Point 2, the training process, as our focal point. It could well be the case that, at this point in the structure we are using, an evaluation for this audience would center on matters like adherence to standards for design and delivery, that is, the "professionalism" of the training.

Stay at Point 2 and adopt the trainees' perspective. Perhaps the chief evaluation issue in this case can be expressed in a single question: "How does all this (the training) relate to my job?"

Suppose we go to Point 1, adopt a looking forward (to the right perspective), and put on our executive's hat. What might we be interested in from that perspective? One quick answer is the results that can be expected in the workplace, at Point 4. Another is the resources required to achieve those results.

Training, like all organizational functions, must compete for resources. Moreover, resources must be allocated before any effort can be undertaken. From this it follows that resource allocation decisions must be made before the resources can be expended. Consequently, from the resource allocation perspective, the case to be made regarding the results of training must be made before the training is conducted, not after.

The preceding examples of evaluation possibilities were arrived at by varying elements of the structure of what might be termed "the evaluation of training problem." One of the elements varied was the point or span of points in the process at which measurements might be taken. Another element varied was the audience for the results of the evaluation. Yet a third element varied was the time perspective employed. Varying these elements, singly or in combination, permits us to identify some of the many purposes for evaluating training. In turn, the purposes for evaluating training are inextricably bound up with the purposes of the training being evaluated.

The Many Purposes of Training

Almost 20 years ago I wrote a brief article addressing what I saw as the need to adopt a "strategic view" of training (Nickols, 1981). My aim then, as now, was to point out that "training is a management tool, not the private domain of those who specialize in its development or delivery, nor of those who make its development and delivery contingent upon some other methodology." By "some other methodology," I mean performance technology, which seems to me to view training as little more than an occasionally useful remedy for skill or knowledge deficiencies.

As a management tool, training serves many masters and many purposes. In the article just mentioned, I presented and explained examples of three such purposes (the first three in the list below). Additional purposes for or uses of training are given in the list below. It is not my intent here to elaborate upon these many purposes. Instead, I wish merely to prompt you to think about how the evaluation of training might vary with the purpose or use of the training itself.

- Focusing energy on issues
- Making work and issues visible
- Supporting other interventions
- Legitimizing issues
- Promoting change
- Reducing risk
- Creating a community based on some shared experience
- Building teams
- Indoctrinating new staff
- Communicating and disseminating knowledge and information
- Certifying and licensing
- Rewarding past performance
- Flagging "fast trackers"
- Developing skills

Given the diverse array of purposes listed above, it seems reasonable to conclude that the results sought from the training would also be diverse. And so they are. It is time now to return to the issue postponed earlier; namely, the fourth element in Kirkpatrick's framework, the results of training.

The Results of Training

When we speak of measuring the results of training – and we mean results beyond those of simply equipping people with the skills and knowledge necessary to carry out their assigned tasks and duties – we are redefining training as an intervention, as a solution to some problem other than equipping people to do their jobs.

In cases where skill and knowledge deficiencies are leading to mistakes, errors, defects, waste, and so on, one might argue (and many do) that training which eliminates these deficiencies and in turn reduces mistakes, errors, defects, and waste, is a solution to a performance problem. This argument is extended to assert that the reductions in mistakes, errors, defects, and waste, as well as the financial value of any such reductions constitute the "results" of training.

The logic of this argument has a certain superficial appeal but it is far from impeccable and even farther from compelling. In short, it does not withstand serious scrutiny. It is frequently pointless to ask, "What business results were achieved as a result of training?" because the goal of training is generally one of preventing mistakes, errors, defects, and waste, not correcting them. Thus, by a strange twist of circumstances, the only way to prove that such training is successful is to shut down the training. As is the case with some other things, it is sometimes the case with training that the true measure of its value lies in its absence, not its presence, but shutting down training is hardly a practical way of testing that proposition.

At this point, it seems worthwhile to see if the evaluation of training problem can be cast in a more practical light. To accomplish this aim, we will use a completely fictitious, hypothetical, situation, one in which an equally fictitious executive, Lee Resnick, will play a central role. In short, let's pretend.

Let's Pretend

Pretend you are Lee Resnick, senior vice president for systems and operations at the Cowardly Lion Insurance Company. You are cutting over to a new, multi-million dollar insurance policy administration system in just a few months and your neck is on the line to the CEO for a "smooth, problem-free introduction" of the new system. You know that's a joke and so does the CEO – there's no such thing as a "problem-free introduction" of a new system – but the underlying message is also clear: If things get too screwed up, it'll be you that gets the ax, not the CEO.

The new system radically alters the way the clerical staff members do their jobs; indeed, the jobs themselves have been radically restructured. Obviously, the people need to be retrained. They need to know how the new system works and how to carry out the many new and different procedures they'll encounter. They'll also have to be sold on the new system, so as to reduce the friction at installation time. Moreover, you don't need some training consultant to tell you all this. You also know that, given enough time, the clerical staff wouldn't need much in the way of formal training at all. Sooner or later, they would figure out how to make the system do what it was supposed to do. In short, they would learn how to do the job even if they weren't trained how to do it. But you don't have time. And you can't afford to live with the financial and political costs of the error rates you'd encounter in a world where people are learning solely from their mistakes. You don't need to be told this, either. So, you know you're going to spend some money on training. The primary issue facing you is how much? How much money and for how much training?

Depending on the riskiness of the situation, your personal circumstances, your career ambitions, and a host of other factors, you might be inclined to go for the minimum amount of training and the minimum expenditure of cash or, conversely, the cost and length of the training might be no

object. Which of these is the case is more or less immaterial because your choice, in either case, will be governed by what is essentially the same criterion: Of the options available to you, which seems most likely to serve your purpose?

When you follow up, which you're very likely to do, you're likely to make do with a few phone calls, a few questions, and a few answers. Formal, structured, and expensive after-the-fact evaluations are of little use and could even pose an inadvertent threat. What would you do, for instance, if you commissioned the kind of evaluation the training people are pressing for and it revealed that the money you spent on training was wasted? Now how's that going to look come performance appraisal time? (Fortunately, you can always hang the blame on the trainers.)

As Lee Resnick, you can probably relate very quickly to item six in the list of training purposes presented earlier: Reducing risk. Your primary motive in providing the training is simply to ensure that the lack of training doesn't create a problem during cutover. Training, in this case, is insurance; prevention as much or more than intervention.

Let's Pretend Some More

Suppose now that you are a new general manager and that your department heads have a long history of isolation and compartmentalism, a history of not talking to one another. Further, suppose you decide to use some training sessions as a means of bringing them together and getting them started talking with one another. How would you evaluate this training?

Suppose instead that, historically, a deaf ear has been turned to laments and complaints about the company's performance appraisal system. A new CEO, charged with changing the corporate culture, is willing to modify it. How could training be used in support of this objective? Which of the purposes in the list above might this kind of training serve? How would you evaluate this training?

Suppose, finally, that the officers of the company are dissatisfied with the quality of their own training and education and decide to institute an advanced management program. First, they attend. Next, some but not all of the senior managers in the pool from which the officers are selected also attend. What's going on here? Which purposes are being served? How would you evaluate this training?

The root word of interest in this article is a verb: "Evaluate." To evaluate some thing is to determine its value, to find its strength or its worth. To evaluate training is to determine its value. Value is relative. What is of great value to one person is of little or no value to another. In evaluating training, then, it is important to know one's audience – the person or persons for whom the determination of value is to be made. As noted earlier, there are several possible audiences for evaluation results. These include the trainees, their managers, the trainers and their managers, the executives of the organization wherein the training is taking place, members of the training profession and even, as we saw at one point, members of the larger community in which the organization is embedded.

Because the definition and perception of value varies from person to person, so do the purposes of evaluation. Moreover, the various audiences for evaluation frequently act as their own evaluators. If you look carefully about you or if you reflect upon your own experiences as a "trainee," you will quickly discover that training is being evaluated every day, but by trainees, managers, and executives — and in accordance with their criteria and purposes.

Conclusion

The concluding point to be made here is very, very simple and very, very important: *There is no "cookbook" approach to the evaluation of training.* To properly evaluate training requires one to think through the purposes of the training, the purposes of the evaluation, the audiences for the results of the evaluation, the points or spans of points at which measurements will be taken, the time perspective to be employed, and the overall framework to be utilized.

References

1. Brethower, K. S. (1967). "Maintenance Systems: The Neglected Half of Behavior Change," in *Managing the Instructional Programming Effort* (Rummler, Yaney, & Schrader, Eds.). University of Michigan.
2. Dewey, John. (1910). *How We Think*. D.C. Heath.
3. Drucker, Peter F. (1959). *Landmarks for Tomorrow*. Harper & Row.
4. Drucker, Peter F. (1989). *The New Realities*. Harper & Row.
5. Kirkpatrick, Donald L. (1975). "Techniques for Evaluating Programs." Parts 1, 2, 3 and 4. *Evaluating Training Programs*. ASTD.
6. Nickols, Frederick W. (1982, April). "Training: a strategic view." *NSPI Journal*.
7. Nickols, Frederick W. (1983, October). "Half A Needs Assessment: What is in the world of work and working." *NSPI Journal*.
8. Reith, Jack (1975). 1969 "Miami Conference Evaluation Methods and Results." *Evaluating Training Programs*. ASTD.

Contact the Author

Fred Nickols can be reached by e-mail at fred@nickols.us. Other articles of his can be found on his web site at www.skullworks.com.