"Half the money I spend on advertising is wasted, and the trouble is, I don't know which half."

David Ogilvy attributed this quote to John Wanamaker, the legendary department store mogul, in Confessions of an Advertising Man, but the concept could just as easily be applied to training. How much of the $58.6 billion budgeted by U.S. companies for training last year actually paid off? How can such return on investment (ROI) be measured? Training executives today are more concerned about these questions than ever; in fact, the 1997 National HRD Executive Survey conducted by the American Society for Training and Development (ASTD), showed that “determining the impact of training on financial performance” was the top challenge for more than half (55.9%) of the HRD professionals surveyed. However, bottom line accountability is elusive, and some say it’s down right impossible.

Kirkpatrick’s Four Levels

Evaluation research has developed as a result of substantive support by the federal government, beginning in World War II training and evaluation activities. It provides answers to the questions of “Do we implement or repeat a program or not?” and “If so, what modifications should be made?” Today, measurement in the form of instructor and course evaluations is a fixture of most training programs.

However, when what goes on in the classroom is not the outcome of interest, these are the wrong measurements -- or at least the unimportant ones -- to take. An outcome that is of importance answers the question of “How have you used what you learned?” This type of evaluation is difficult to conduct as it requires being done at three months, six months, or even at 12 months from the time of the training. Adding to the difficulty is the aspect that the evaluators need to be co-workers, managers and outside customers of the participant who took part in the training.

In order to classify areas of evaluation, Donald Kirkpatrick created what is still one of the most widely used approaches, even though it was first developed in 1959. At the time, he was a professor of marketing at the University of Wisconsin. His four levels of evaluation are:

Level 1: Reaction - a measure of satisfaction
Level 2: Learning - a measure of learning
Level 3: Behavior - a measure of behavior change
Level 4: Results - a measure of results
Here are questions that should be asked at each level:

| Level 1: Reaction | Were the participants pleased?  
|                  | What do they plan to do with what they learned? |
| Level 2: Learning | What skills, knowledge, or attitudes have changed?  
|                  | By how much? |
| Level 3: Behavior | Did the participants change their behavior based on what was learned in the program? |
| Level 4: Results  | Did the change in behavior positively affect the organization? |

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.

Current Trends

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.

When looking at ROI and cost benefit analysis, it is important to remember that:

- Improving efficiency means achieving the same results with lower costs.
- Improving effectiveness means achieving better results with the same costs.
- It is possible to get better results with lower costs, and this is called improved productivity.

In order to calculate ROI, evaluation experts such as Jack Phillips are recommending the addition of a fifth level to Kirkpatrick's model for some programs. This requires collecting level 4 data, converting the results to monetary values, and then comparing those results with the cost of the training program.

Here is Phillips' basic formula for calculating ROI:

- 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. Phillips recommends dividing training results into hard data and soft data. He says hard data are the traditional measures of organizational performance because they're
objective, easy to measure, and easy to convert to monetary values. They include output (units produced, items assembled, tasks completed); quality (scrap, waste, rework); time (equipment downtime, employee overtime, time to complete projects); and cost (overhead, accident costs, sales expenses). Conversely, soft data includes such things as work habits (tardiness, absenteeism); work climate (grievances, job satisfaction); attitudes (loyalty, perceptions); and initiative (implementation of new ideas, number of employee suggestions).

- Total the costs of training.
- Compare the monetary benefits with the costs. The non-monetary benefits can be presented as additional - though intangible - evidence of the program’s success.

It takes a lot of time and effort to conduct evaluations at this level, and not every program needs this much attention. But when it’s important to know the real value of a program, ROI measurement can go a long way to justify company efforts. For example, Magnavox Electronics Systems Company in Torrance, CA, maintains an 18-week literacy program covering verbal and math skills for employees. Here are the results of a five-level evaluation the company conducted:

<table>
<thead>
<tr>
<th>Level 1:</th>
<th>reaction was measured by post-course surveys.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 2:</td>
<td>learning was measured with the Test of Adult Basic Education.</td>
</tr>
<tr>
<td>Level 3:</td>
<td>changes in behavior were measured by daily efficiency ratings</td>
</tr>
<tr>
<td>Level 4:</td>
<td>business results were measured through improvements in productivity and reductions in scrap and rework.</td>
</tr>
<tr>
<td>Level 5:</td>
<td>ROI was calculated by converting productivity and quality improvements to monetary values. The resulting ROI was 741%.</td>
</tr>
</tbody>
</table>

According to Phillips, 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:

- determine whether or not a program is accomplishing its objectives
- identify the strengths and weaknesses in a Human Resources Program.
- determine the cost/benefit ratio of an HRD program
- decide who should participate in future programs
- identify which participants benefited most or least from the program
- reinforce major points made to the participants
- gather data to assist in marketing future programs
- determine if the program was appropriate (Phillips, 1983)
A New Model?

Not everyone agrees that the Kirkpatrick model should be used for evaluations. Elwood Holton, writing in HRD Quarterly (1996, 7:1, p. 5-21), goes so far as to say it isn't even a model, but rather merely a taxonomy. 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.
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. [For a more detailed outline of Holton's paper, see http://131.96.116.76/msit/vmw/evaluation/new.htm]

Conclusion

In implementing the evaluation process, Phillips cautions us to remember Murphy's Laws:

1. If things can go wrong, they will.
2. Nothing is as easy as it looks.
3. Everything takes longer than you expect.
4. Projects take longer than they do.