

**Concept Paper**

**On**

**Improving Training Evaluation in Organizations**

Prepared by

Dr. Paul Squires  
President



*"A Leader in Measuring Success"*

20 Community Place  
Morristown, NJ 07960  
973-631-1607 fax 973-631-8020  
[www.appliedskills.com](http://www.appliedskills.com)

## Improving Training Evaluation in Organizations

---

According to the American Society for Training and Development, U.S. corporations spend approximately \$70 billion annually to train their employees. For most corporations between 2% and 4% of their annual revenues are spent on training. For a corporation whose annual revenue is \$100 million, this represents an expenditure of \$2 million or more annually. Considering the size of this expense it is somewhat surprising that the effectiveness and return on investment of this money is not more carefully assessed. The most common framework for evaluating training is Kirkpatrick's four levels of training evaluation. These include evaluating trainee satisfaction, trainee learning, trainee job improvement and business impact. According to a survey by ASTD, 80% of organizations conduct trainee satisfaction evaluations, 40% evaluate trainee learning, and less than 10% perform evaluations of trainee job improvement or business impact.

There are two main reasons why organizations do not perform adequate evaluations of training. These include:

1. Management's training evaluation strategy is not well formed.
2. Training evaluation is technically complex.

The following describes some thoughts about better approaches to training evaluation. These thoughts will address the two reasons why poor evaluations of training are commonplace and suggest solutions. In addition, the strengths and weaknesses of the Kirkpatrick framework will be discussed and alternatives suggested.

**Management's training evaluation strategy is not well formed.** Most organizations do not have a training evaluation strategy. Management doesn't require it. Instead, training is determined and evaluated by managers' "gut feelings" about what should be trained and how well it worked. Needs analyses are performed infrequently and competency models are not up-to-date or do not exist. Training managers are often unfamiliar with key business measures. There is no strategic intent to training because there is no alignment it with the objectives of the business.

One measure of the effectiveness of training is to assess how well it is aligned with critical business needs. If the training offered by an organization is not based upon an up-to-date competency model and tracked against a key business metric then the "quality" of the training is suspect.

This misalignment with the business strategy can be remedied by establishing a "Board of Education" comprised of training managers and business leaders. They select and fund training initiatives based upon appropriate business objectives, skill analyses, and needs analyses associated with the objectives.

The Board of Education decides what courses require an evaluation and what levels of evaluation are most appropriate. They may confer with training evaluation experts to make these decisions. Not all courses require training evaluation or more than one level

# Improving Training Evaluation in Organizations

---

of training evaluation. The training evaluation strategy should include a determination of the courses that will be evaluated and what levels will be used for each.

Management must have good measures of progress toward their business objectives and trainers need to understand them and conduct training that positively impacts these measures. In short, the Board of Education develops the training evaluation strategy and tracks the results.

**Training evaluation is technically complex.** Training evaluation is a good example of social science research. All the expertise and tools required to do good social science research are required for good training evaluation. The development of measures to evaluate training, the choice of a research design, and the analysis of the data are all part of the research and are technically complex. They require an understanding of psychometrics, experimental design and statistical analysis. Psychometrics is the special application of mathematics to the development and validation of measurement tools, including business measurements. Experimental design requires knowledge of randomization, control groups, longitudinal and cross-sectional studies, and other topics important to successful research method. Statistical analysis is a more general application of mathematics to evaluate the study results quantitatively. This expertise is not usually available in organizations. Those experienced in social science research such as industrial psychologists are available to perform the technical work needed to create a training evaluation program.

With these thoughts in mind, let's look at Kirkpatrick's four levels of training evaluation and discuss their strengths, weaknesses and solutions.

## 1. Trainee reaction to the training experience.

The purpose of this first Kirkpatrick level of training evaluation is to determine if the trainee liked the training. Most training organizations collect what are unflatteringly called "smiles sheets." Sometimes this information provides little in the way of useful feedback. But, done properly, this level of evaluation has value because the evaluation is a sort of customer satisfaction measure, if you consider the trainee to be the customer. This evaluation is administered as an end of class survey completed on a voluntary basis by the students. Questions commonly used are, for example, was the training content job-related, was the instructional material well organized and easy to follow, was the instructor prepared, were all trainees encouraged to participate, was the physical environment clean and comfortable, etc.

While customer satisfaction data are valuable, training organizations must determine the purpose and content of end-of-course evaluations on the basis of their training evaluation strategy. For example, if the course is product training and is for sale to external customers, then customer satisfaction data are important. If the course is a mandatory safety course for employees, then customer satisfaction is probably less important. If the course is newly released, questions about the quality and organization of the materials are important. If the

# Improving Training Evaluation in Organizations

---

course was delivered to 1,000 trainees and three rounds of revisions were completed, questions about the quality and organization of materials are probably less important.

Data that are probably always valuable to collect are those that provide information about the instructor(s). These data should be stored in a database and tracked to compare performance for different groups of students, different courses and over time. Some organizations have used them as part of a performance appraisal process and they are always useful to assess instructor performance.

The administration of training evaluation measures has become much easier and less expensive due to the Internet. Because the Internet is easily accessible from anywhere and at any time, data collection becomes simplified. The trainee responding to the survey inputs the data and this eliminates a large expense. Web-based data collection and analysis software is available and organizations with the expertise in-house can manage the evaluation process themselves or it can be outsourced to companies with the software and training evaluation expertise.

As a general rule, some topics that should be included in end of course evaluations are:

- job relatedness of training topics,
- instructor effectiveness,
- quality of materials,
- difficulty level, and
- class management.

Once the purpose(s) of the evaluation is decided and topics selected, then writing the statements that comprise the evaluation is the next step. Writing clear, unambiguous, non-redundant statements that accurately achieve the purposes of the measurement is more difficult than most imagine. Those inexperienced in creating measurements often write ambiguous and redundant statements.

## 2. Trainee gains in learning

The purpose of the second Kirkpatrick level of training evaluation is to determine if the trainee learned anything from the training. For example, if programmers are learning object-oriented programming, an assessment of skill in object-oriented programming before and after the training can indicate the impact of the training. Obviously, an assessment such as a mastery test is required to determine objectively if learning occurred. The development of a fair and valid assessment requires psychometrics expertise. The creation of a poor quality assessment creates the risk of misdirecting conclusions. In addition, the evaluator must be assured that the pre-test and post-test are of equal difficulty and coverage of topics. And, simply administering a pre-test and a post-test often misleads the evaluator. Without a “control-group” to compare gains in learning, the evaluator cannot safely say that the gains were due to the training. Also, if the students

## Improving Training Evaluation in Organizations

---

were not randomly assigned to the training group and the control group, then there may be motivational or other differences between the groups that account for differences in gains. There are other threats to the accuracy of decisions made when comparing pre-tests and post-tests.

### 3. Trainee increase in job performance

The purpose of the third Kirkpatrick level of training evaluation is to determine if the trainee performs more effectively after returning to the job. This is clearly a critical part of the evaluation of training. Employers send employees to training in order to increase their job performance. But, this form of evaluation is rarely done. The challenging part of level three evaluations is the creation of a fair and accurate measure of job performance. The evaluator must use a job performance measure that reflects only the parts of the job that were included in training. For example, if a stock broker participated in mutual fund accounts training, the measure of job performance must focus on the use of that knowledge. Using monthly sales results would be a poor choice of job performance measures because there are many skills required of a stock broker that influence sales results, not just product knowledge.

A common solution to this problem is to develop special performance appraisal tools based upon the learning objectives of the course. The learning objectives are used to create job task statements that are related to the training and will be reflected in the job. A performance appraisal tool is created using these statements and ratings on the job task statements are gathered from the supervisor and the trainee. The ratings are combined to determine the extent that job performance improved.

When gathering data in this way, the evaluator must consider research design issues – for example, whether to use a control group or conduct a longitudinal study or to gather pre and post training job performance measures.

The third level of analysis assumes that an evaluation at level two indicates that students gained skill and knowledge during the training. If no gains can be demonstrated then level three evaluation is not likely to show positive results.

### 4. Improved business performance.

The purpose of the fourth Kirkpatrick level of training evaluation is to determine if the business improved due to the training. In other words, what was the return on investment? It is very rare for a course to be evaluated in this way. Various approaches have been suggested to estimate ROI. Capital budgeting analysis, utility analysis, cohort analysis and combinations of these approaches have been considered. Organizations that have good business measures in place are in a stronger position to determine the business impact of training. These are all effective approaches but are based upon data and statistics gathered at levels two and three.

## **Improving Training Evaluation in Organizations**

---

The fourth level of analysis also assumes that evaluations at levels two and three were successful. Students gained skill and knowledge during the training and it had an impact on their job performance. If no gains can be demonstrated at levels two or three, then a level four evaluation is not likely to indicate any positive return on investment.