

---

# Evaluation: The Link between Learning and Performance

By Roger Chevalier, CPT

---

Management has traditionally asked the questions:

“How many employees attended the course?”

“How did they like the course?”

“What did they learn?”

However, with more of a bottom line orientation, managers are beginning to ask:

“Are they using what they learned?”

“Are there other things we should look at to change performance?”

“What business results were improved?”

“What is our return on investment for the training?”

The answers to these new questions can be found by systematically evaluating the training we provide. But while the idea of evaluating training is nothing new, instructional designers and human performance technologists have been slow to embrace this powerful tool for measuring the impact of training. In this article readers will be presented with the distinction between Formative and Summative Evaluation, will learn about Kirkpatrick’s Levels of Evaluation with additional levels of evaluation by developed Phillips and Kaufman, be presented case studies of three evaluations, and have summarized various lessons learned about evaluation.

Instructional Systems Design (ISD) has given trainers the structure needed to systematically produce and refine training. While analysis, design, development, and implementation are routinely used to create and deliver courses, evaluation of the results has been typically limited to measurement of how participants reacted to the training and whether they improved their knowledge of the subject.

## Formative and Summative Evaluation

The distinction between formative and summative evaluation has been widely used since its first description in the late 1960s (Scriven, 1969). These two types of evaluation are done for different purposes, at different times, using different mean

Formative Evaluation is a process of evaluation done during the design, development, and implementation stages to improve the quality of the training or other performance improvement intervention. The goal of this evaluation is to provide information with which to ensure the quality and consistency of the intervention. Formative evaluation can be done with a subject

matter expert review, a user review, or a pilot test. In my experience, an intervention should be designed and developed to about an 80% level of completion, then piloted with both subject matter experts and users who will provide the added feedback needed to further refine the intervention.

Summative Evaluation is a process of evaluation during and after the training or other performance improvement intervention. Its purpose is to determine the impact on individual and organizational performance. Summative evaluation can be done by direct observation, surveys of those affected by the training or intervention, measurement of performance indicators for the activities done by those affected (such as quality, productivity, or customer satisfaction), and/or measurement of a business outcome (such as number of sales, customer retention, profitability, or market share).

Stated more simply in an often-used example, when the cook tastes the soup, that's formative evaluation; when the customer eats the soup, that's summative evaluation.

#### **Four Levels of Evaluation**

Another way to classify evaluations was developed by Donald Kirkpatrick who described four levels of training evaluation (Kirkpatrick, 1998). Level 1 evaluations focus on the reaction of individuals to the training or other performance improvement intervention. These are commonly described as "smile sheets" as they have been traditionally used to report the reaction of the student to the training. Level 2 evaluations assess what has been learned and has been traditionally been done with end of course tests. Level 3 evaluations measure the transference of what has been learned back to the workplace. Level 4 evaluations measure the impact of the intervention on a business results.

Levels 1 and 2 can be used as part of the formative evaluation while Levels 3 and 4 are typically associated with summative evaluation. Two approaches to "Level 5" evaluation have also been proposed. Specifically, Roger Kaufman has advocated evaluating the contributions that an organization makes to society (Kaufman, Keller, and Watkins, 1996) while Jack Phillips has espoused the need for calculating the return on investment for the intervention (Phillips, 2000).

The American Society for Training and Development's (ASTD) 2003 State of the Industry Report provided the results of a benchmarking survey of 276 U.S. organizations. These organizations reported whether they evaluated their training programs at each of four levels (ASTD, 2003, p. 19). Seventy five percent reported that they evaluated their training programs to determine the participants' reaction (Level 1) and 41% reported that they evaluated what knowledge and skills were learned (Level 2).

Only 21% of those organizations that were surveyed reported that they evaluated whether the behavior of their students changed because of the training (Level 3) and only 11% evaluate whether organizational performance as measured by a business results improved (Level 4). If anything, these numbers are high for the training industry as a whole since these organizations were selected from those that voluntarily participate in the study each year.

The act of evaluation not only measures the impact of the training provided but also can actually encourage the training participants to use what was learned because it is being evaluated. Stated

another way, “What gets measured gets done.” By evaluating our training programs we can encourage the use of what was taught as we measure how well the graduates were able to apply what was learned.

### **A Level 3 Evaluation Example**

In 1983, the Coast Guard Training Center in Petaluma, California trained about 4,000 students in 25 courses, none of which had ever been systematically evaluated. While courses were periodically rewritten, no feedback was ever obtained from the field to assess the effectiveness of the training. A more systematic approach to refining the training was necessary and the best tool available was evaluation.

After rewriting the curriculum outlines to identify the desired behavioral outcomes for each course, surveys were sent to recent graduates and their supervisors to assess: a) how well the graduates were able to do the desired behaviors, b) how often they did those behaviors on-the-job, and c) how important it was for them to be able to do the behaviors. Open-ended questions were also asked to determine what else graduates should have learned during the training that would have prepared them for their roles in the field.

As the result of this systematic assessment, courses were continuously refined making the training more efficiently delivered and more relevant. Student years of training were reduced as outdated objectives were eliminated and job aids were introduced. The net savings to the Coast Guard was found to be \$3,000,000 a year recurring from a base cost of \$9,000,000 for student and instructor salaries alone (Chevalier, 1990).

In this case, the value of Level 3 evaluation was in making training more efficient and effective by closing the feedback loop between the provider (the training center) and the customer (graduates and supervisors). The overall course refinement system became one of systematic assessment and continuous improvement.

The Coast Guard Training Center Petaluma has recently returned to conducting Level 3 evaluations of their courses by using an on-line survey of graduates and their supervisors. The results so far have been impressive, with over a 65% return rate for the surveys. This is a substantial improvement over the manual system (printed surveys mailed to graduates and their supervisors) that began in the mid 1980's and was later replaced with optically scannable forms.

There are some in our field who would say that self-reports are not reliable enough for Level 3 evaluation. While I would agree that direct observation would be a better indication of Level 3 performance, this view should be tempered by the reality of Coast Guard graduates being dispersed around the world after completing their training. The feedback received from the graduates and their supervisors is the next best thing and was gotten for a fraction of the cost of other measurement methods. It is an example of the old 80/20 rule in action: for 20% of the cost, the Coast Guard got 80% of the information they needed to refine their training programs.

### **A Level 4 Evaluation Example**

In 1995, Century 21 Real Estate Corporation redesigned the one-week training program it provided to over 15,000 to 20,000 new real estate agents each year by over 100 trainers in

various locations around the country. A pre and post-test to determine what was learned (Level 2) was added to the end of course participant evaluations (Level 1) used in the previous version of the course.

To measure the effectiveness of the training, the progress of the agents was tracked after graduation through a sales performance system. The number of sales, listings, and commissions for each graduate was cross-referenced to the office where they worked and to the instructor who delivered the training.

While the initial goal of the evaluation was to measure the value of the training, the more important return was in being able to provide feedback and improve the performance of the sales managers and the instructors. Before this level four evaluation, instructors were rated primarily of the feedback received from the participant evaluations. By tracking their students' performance post-graduation, a better measure of instructor performance was made available.

When the graduates were tracked by office, different patterns of success were observed. A typical class had 18 to 24 new agents from various offices in the same geographic area. In reviewing the agent performance during the six months that followed the training, unusual patterns of success were found. Several agents from the same office might greatly outperform agents from another office who took the same course with the same instructor at the same time. It became obvious after tracking the performance over several classes that agents from some offices had a higher probability of success than those from other offices.

Armed with this information, the graduates were surveyed to identify what was different in the work environment that contributed to their performance. Not surprisingly, the high-performing graduates came from offices where they received direction when they needed it, had access to ongoing training, and received psychological support as they did their jobs. The low performing graduates tended to be from offices where all or some of these factors were not present.

The titles of the 100 trainers were changed to "Sales Performance Consultant." These individuals, while still responsible for delivery of the training, now worked with the offices in the field to try to shape the environment to which their students returned. About 100 business consultants were also given access to the performance data for the offices they served to provide feedback to the brokers on the work environment of their offices and the impact it was having on their newly trained salespeople.

Another lesson learned was centered on the over reliance of Level 1 evaluations to assess the instructors' performance. In one case, a trainer who was rated in the bottom third of all trainers by his students in Level 1 satisfaction evaluations, was found to be one of the most effective in how his students performed during the first three months after they graduated. The trainer was running evening sessions that gave the students practice with what they learned. After learning about how to make "cold calls" to identify potential clients, that evening students would spend two hours making calls and another hour debriefing what was done. Some of the students made appointments to meet with prospects that they contacted during the evening session. While the students were unhappy about working evenings (as was reflected in their Level 1 evaluations), they were much better prepared for the real world.

### ***Level 3 and 4 Evaluation as Part of the Intervention: A Case Study***

Sometimes it is easier to sell evaluation to our internal and external clients as a part of the intervention rather than as a separate event. Jack Welch, the former CEO of General Electric said, “You get what you measure” (Welch, 2001). The act of measurement encourages trained personnel to use what has been learned in the classroom.

While many organizations view evaluation as “nice to have” at best, managers need to learn that evaluation can be an important part of their overall intervention strategy. The act of measuring performance encourages the use of what has been learned in the training as well as reinforces the desired change associated with other performance improvement initiatives. This may be another application of the Hawthorne Effect in that the act of evaluation is a form of observation that leads to increased performance (Roethlisberger, 1941).

In 1999, a software company was about to launch a new product line with a new sales team. While training was necessary for the sales team, evaluation was also needed to provide timely, behaviorally specific feedback for the salespeople to use what was learned and continue their development.

The following performance improvement intervention started the way most do, with a request for training (Chevalier, 2003). The company provided software products and services for large financial organizations and government agencies. They had recently decided to develop products and services for medium size financial institutions and formed a sales team to bring these products to market.

While the sales manager was looking for training to enhance the selling skills of his sales team, he was open to a broader solution of building systems that would systematically track and continuously improve his team’s performance. Not only would the solution have to contain selling and sales management systems to support the training given the sales team, it would have to bridge the gap between the sales team and the customer service people in another division.

Sixteen hours of consultative sales and customer service training were delivered in a weekend workshop. The consultant provided the structure by presenting the various sales and customer service models and used the participants to cross train each other as they moved from the generic models to practical courses of action.

Two sales management systems were developed from the coaching and sales performance aids and the employee input given during the training programs. The systems focused on the means (how salespeople were perceived by their clients) as well as the ends (how clients moved through sales channels to become customers).

A client and customer survey, derived from the sales performance guide, was developed to gather information from clients and customers regarding the selling process and the value of the products and services being offered. As soon as a sales lead was “declared dead”, a one-page survey was sent to the failed prospect. A more comprehensive two-page survey was sent to customers after the purchase was made, the software installed, and the first month of service provided.

When the surveys were returned, a copy was immediately given to the salesperson. The sales manager’s administrative assistant would analyze the feedback received for each salesperson and

prepare a monthly summary for the sales manager who would then provide feedback on selling tendencies and ways for each salesperson to become more effective.

A parallel survey was also developed to gather information every three months from the salespeople on the leadership; they were receiving from the sales manager. The net effect was to develop an interrelated survey system for the systematic assessment and continuous improvement of the sales and coaching processes. Continuing customer service surveys were later developed to bring that division under the same system

While training was the first part of the intervention, the use of evaluation provided systematic assessment and continuous improvement for the new sales team. The act of tracking their individual and collective performance after the training encouraged use of what was learned in the classroom while providing them with an ongoing source of feedback with which to improve their performance.

### **Lessons Learned**

- Evaluation should use existing measurement systems where possible. Since they already exist, there is baseline information available, management believes the metrics measure a desired business outcome, and it is cheaper than setting up a new evaluation system.
- Level 1 (reaction) and Level 2 (knowledge and skills) evaluations can lead to a false sense of security; there may be no relationship between how participants feel about the training and improved individual and organizational performance.
- Level 3 evaluations can be used to refine the training provided but Level 4 will determine whether it has value. Level 5 (ROI) evaluations will determine if the gain was worth the pain.
- It may be not desirable, practical, or necessary to do all five levels of evaluation. Select the level that will produce the information needed to evaluate the intervention.
- Use multiple measures so performance is not driven by just one desired outcome. Trade-offs are often seen between measures for quality, quantity, time and cost.
- Summative evaluation should focus on intermediate goals (such as improving quality, customer service and cost reduction) as well as desired business outcomes (such as sales, profitability and market share).
- Evaluation can be imbedded in the intervention itself encouraging improved performance in the areas measured.
- The act of measuring encourages the use of what was learned. Evaluation can improve transference of what was learned in training back to the workplace.
- Evaluation documents the impact that training has made and makes the training function more defensible during budget cuts and more credible when asking for more resources.

### **Conclusion**

Evaluation is a powerful stage of both the instructional design and the human performance technology models. Unfortunately, it is not often done at a level that will measure the impact of training or other performance improvement interventions.

If done properly, systematic evaluation can provide the information needed for continuous improvement. Once the evaluation process has begun, more opportunities for added evaluation

may be discovered to further understand the patterns of performance identified in the initial evaluation.

Stated another way, “If your going to teach a bear to dance, you better be prepared to dance as long as the bear wants to dance.” When the value of evaluation is fully realized, the opportunities to improve performance will take on a life of their own.

## References

- Annual Report of the Industry (2003). American Society for Training & Development, Alexandria, VA.
- Chevalier, R. D. (1990, May/June). Systematic Change, *Performance and Instruction*, 29 (5), 21 – 23.
- Chevalier, R. D. (2003, May/June). Updating The Behavior Engineering Model, *Performance Improvement*, 42 (05), 8 – 14.
- Kaufman, R., Keller, J., and Watkins, R. (1996). What Works and What Doesn't: Evaluation Beyond Kirkpatrick, *Performance and Instruction*, 35(2), 8 – 12.
- Kirkpatrick, D. L. (1998). *Evaluating Training Programs*, Second Edition, San Francisco, CA: Berrett-Koehler.
- Phillips, J. J. (2000). *Return on Investment in Training and Performance Improvement Programs*, Woburn, MA: Butterworth-Heinemann
- Roethlisberger, F. J. (1941). *Management and Morale*, Boston, MA: Harvard University Press.
- Strayer, M. J. and Chevalier, R. D. (1995). *From Training to Performance, Performance Technology 1995*, Washington, DC: International Society for Performance Improvement.
- Scriven, M. (1967). *The Methodology of Evaluation, Curriculum Evaluation*, American Educational Research Association Monograph Series on Evaluation, No. 1, Skokie, IL: Rand McNally.
- Welch, J. F. (2001). Presentation at the TechLearn 2001 Conference.

Roger Chevalier, Ph.D., is an independent performance consultant specializing in integrating training into more comprehensive performance improvement solutions. With over 25 years experience in performance improvement, Roger is a former vice president of Century 21 Real Estate Corporation's Performance Division and a former training director for the U.S. Coast Guard's west coast training center.

Roger has earned a Ph.D. in Applied Behavioral Science as well as two master of science degrees in personnel management and organizational behavior. His previous education includes bachelors and masters degrees in English Literature. Roger is the Human Performance Technology Forum editor for ASTD's *Performance in Practice*.

Roger may be reached at his home office at 924 Hudis Street, Rohnert Park, CA 94928, 707.584.7160 or rdc@sonic.net.