

Training Evaluation Methods—Who, What, Why and HOW

**William “Bob” Coffey, CSP, CPEA
President
WRC Safety and Risk Consultants, LLC**

It is a shared experience, the class is over and each attendee is given a piece of paper with the word “Evaluation” written across the top. The boxes are checked, comments are made and names are scribed (optionally). The attendee leaves, having evaluated the class. The scenario may be somewhat different. Possibly the attendee is given a written multiple choice test before (or instead of) the post class evaluation. In either case the attendee leaves the classroom setting wondering if this is the extent of training evaluations. It is, of course, not, but this brings up a number of other questions:

- What is training evaluation?
- Why evaluate training?
- Who is being evaluated and who is performing the evaluation?
- How is/how should training be evaluated?

The above are among the most basic of questions. From this inquiry, another basic question arises, do the training providers (instructors, facilities, companies, etc.) have answers for these questions? These answers underscore the very reasons for performing evaluations. Without a clear understanding of the answers to these questions, it must be asked if the training evaluations have any meaning. By analyzing the situation and answering the What, Why and Who questions, guidance will be provided on answering the How question.

What is training evaluation? Webster’s defines evaluate as “To ascertain or fix the value of.”ⁱ So, training evaluation would, therefore, be to determine the value of training. If the value of training is to be determined, it must first be determined why a person is being trained.

The author has postulated that training is conducted for one of three reasons:

- To convey knowledge
- To change a behavior/s
- To meet a regulatory requirement.

Some notes on these definitions: while the above reasoning can be applied to training in general, since the author is an EHS professional, EHS training examples will be used as examples for

highlight purposes. In an EHS sense, conveying knowledge can be as simple as relating to the attendee the location of MSDSs. The phrase “to change a behavior” can be applied in a specific or general manner. An example of use of the phrase in a specific manner can be conducting forklift training on the use of a seatbelt during forklift use. This is a specific behavior a training session is trying to change. In general, the change of behaviors can be synonymous to teaching a skill. A skill can be viewed as a series of behaviors which lead to an outcome (hopefully the desired outcome). As an example, consider safe crane use. To use a crane safely, an operator must perform inspections, analyze situations, understand and apply loading and lifting principles of the crane and depending on the circumstance, interpret and apply hand signals from co-workers. These items can be broken down to behaviors the operator must apply to use the crane safely. Quite often, training will be a combination of two, or all three of the above reasons.

The discussion on training was entered into for a simple reason; if the reason why the training was conducted is not understood, how can the training be gauged? If a training evaluation is to be conducted, it must be understood why the training was conducted in the first place. Or to phrase in another way, how can the training be gauged (evaluated) if it is not clear what the attendee was expected to learn from the training in the first place?

While some organizations will make the determination of what the attendee should learn from the class informally; in a formal training development scenario, this determination would be expressed in the course goals and objectives. Determining Goals and Objectives is the step after a training needs assessmentⁱⁱ. In goals and objectives development, it is outlined what is to be accomplishedⁱⁱⁱ. From this perspective, evaluations are easy. It is a method to determine if the course’s goals and objectives were met. So to reduce the question of what is training evaluation to a somewhat simple answer, once it is understood what is desired out of a training session (expressed as goals and objective) the gauge or “value” (as per the Dictionary definition) can be set from these. The comparison of what was desired to what is accomplished is the evaluation.

Why evaluate training? The answer in this case is simple, resource management. Does an organization’s training meet what is needed of it? If it was determined that a set of skills are needed, if the training is not meeting these needs the original issue is still present and being resolved. This simple answer, however, brings another element to the training evaluations issue, efficiency. If training is evaluated for resource management issues, how well is it accomplishing its task? Is the training being performed in the most efficient fashion? With this question, the evaluation can be expanded to evaluate the instructor and/or the course material. By efficiency, it must be understood that while business resources are an issue (i.e. money spent and time used), they are not the major factor. The major factor is how well is the knowledge and/or skills transferred. If the meeting of goals and objectives is viewed as only a pass/fail scenario, this element is a non-issue. If, however, the goals and objectives are viewed as individual pieces, an analysis can be performed on each element. Passing can now be viewed as retaining a certain percentage of the information. By conducting the analysis, it can be determined if certain elements of the training are not being conveyed to the general population and improvements to the system can then be made. There are aids in this endeavor. In 1994 Donald Kirkpatrick outlined four levels of evaluation (referred to as Kirkpatrick’s Evaluation Levels)^{iv}. Kirkpatrick outlines how (according to this model) training should start with the first level and (as opportunity arises) should progress to the fourth level. The lower levels feed up to the higher level. It revolves around metrics and can require a high level of analysis (and, therefore, time). Kirkpatrick’s levels are summarized in Table 1.

Level	Title	Explanation	Measures*
1	Reactions	This is the most basic level. It is literally how the attendees felt about the training. This is a purely subjective level. You are gauging an attendee's perception on attending the training and if they "feel" it will be applicable. This is exemplified by the class evaluation sheets filled out at the end of a training session.	Individuals
2	Learning	This is a comprehensive level. This level is evaluating the retention of the information/skills presented. This level would correspond to a post-class knowledge test.	Individuals/Groups**
3	Transfer	This is an application level. At this level the evaluation revolves around whether the attendee can apply the knowledge/behavior in a suitable fashion.	Groups**/Units***
4	Results	This is the final level. It is analyzing whether the original issue which generated the training is addressed	Units***

* This section was not part of Kirkpatrick's original work and is an addendum by the author to aid in discussion

** Groups is used to mean any situation where a number of individuals that work together. Depending on the situation it can refer to cells, departments or whole facilities

*** Units is used to mean any situation where a number of Groups work together. It will commonly refer to a facility, division or a whole business

Table 1.

The "Measures" column was added to the above table to aid in answering the questions on who is being evaluated and who is performing the evaluation. Kirkpatrick starts with the reaction level. This level is looking at an individual's reaction to the training and their intent to apply it. This is solely an individual level, gauging a personal reaction to the training. The Learning level is looking at retention of knowledge/skills presented in the class. While this is evaluating an individual, it is in reference to the individual applying this knowledge/skill in relation to a "group"- the individual's cell, department or, facility. The Transfer level is a critical level, it is looking at whether the knowledge or skill presented in the training is being utilized and applied. This is where the individual issues which were identified in a Needs Assessment are being corrected. Individual responses are of smaller importance at this level as compared to the Group. Can you say that a training session was a success and a need is being met if one person of a thirty five person department (all of which have had the training) is showing the behaviors utilized? This level is actually gauging the Groups and possibly the Units. The Results level is the last of Kirkpatrick's levels. At this level the behavior and knowledge being utilized by the training attendee is compared to the knowledge and behaviors identified by the Needs Assessment. This occurs at the highest level (Units) as it is measuring total success of the training endeavor. If a company instituted a new crane training program to reduce crane accidents and near misses, is the reduction occurring? To use an example to illustrate the logic of the process:

- Trainee attends crane training and felt the training was worthwhile (level 1)

- Trainee absorbed the information and understands why and how to inspect the crane and safe operation (level 2)
- Attendees inspect crane properly every day and observe safe practices (level 3)
- Company has less crane incidents (level 4)

To continue the discussion on who is evaluated, in all cases it will be the attendees who will be evaluated. The training is conducted to engage the attendees to utilize the knowledge and behaviors taught and they will be the ultimate gauge. With the use of after class evaluation forms and other formal techniques, the instructor and course materials can also be evaluated. The evaluation, however, does not have to be conducted during the class or necessarily by an instructor. Since the goal of the training scheme is to have knowledge and behaviors utilized in application, evaluators can be any person in the application (most likely work) setting who is knowledgeable on the subject and behaviors.

So to recap up to this point, training evaluations are performed to determine if the attendees have gained the knowledge and skills put forth in training sessions. It is a process where the attendees are measured and judged as to the extent (if any) in which they apply desired behavior and use information presented. Evaluations can also be conducted of the instructor and course materials, to gauge how well the material is presented in a manner it will be retained by the attendees. The attendees will be evaluated (as well as possibly the instructor and training materials) but the evaluation may be conducted outside of the training session. There are various “levels” for which the training can be evaluated, but the ultimate evaluation is whether the training corrects the issues it was designed to correct. This recap was provided to lead into the discussion on HOW to conduct evaluations.

Before reading the previous paragraphs, the readers must have had some (if not all) knowledge on training evaluations. They will most likely applying some, if not all, of the concepts presented. With the knowledge of previous discussion, it is possible to move forward, building a scheme to measure the performance of the training at various levels. To do this, a question must be asked: what was desired out of the training?

The ultimate goal of all health and safety programs is to prevent injuries or decrease the likelihood of an injury occurring (which is, in fact, the same thing). So in the end, safety training should show a reduction in injuries. While this may be true in theory for application, there is more that is needed. The question becomes, how was training determined to reduce injuries? This can occur in a number of ways (training needs assessment, trend analysis, procedure review,...). Once the need is determined, the training is developed and delivered. So at this point, a reduction in injuries should be noted, correct? That is a fallacy. How is it known that the training is effective? The lack of a reduction in injuries may be due to a deficit in the training. This is where Kirkpatrick provides guidance. Kirkpatrick highlights that evaluation occurs at various levels.

The process will begin by applying Kirkpatrick to common evaluating schemes. The post class evaluation form is actually evaluating the presenter and the course material. It is providing feedback on the attendees’ feelings and intentions, but not whether the skills and knowledge will be applied effectively. A post -class knowledge test is again testing the instructor, material, and the knowledge transfer portion of the test (for example, can the attendee remember the location of the MSDS binder or the evacuation assembly point). It is at the Transfer level where behavior and knowledge is actually applied. This is where the knowledge and behaviors must be used in an

actual situation or a simulated situation. The final level asks has this skill transfer had the desired results? An example will be used to illustrate.

OSHA's Powered Industrial Truck (PIT) standard requires a skills demonstration^v. The attendee must undergo classroom and practical training and then undergoes a certification (which is the skills test). A company providing this training wants to determine the effectiveness of the training. This can be done by applying Kirkpatrick as follows:

1. A post class evaluation form is completed
2. A knowledge exam is taken by the attendee at the end of the exam
3. A skills test is given of the attendee
4. A comparison of Powered Industrial Truck incidents are compared pre-training vs. post-training

In-depth comments into each of these points will aid in understanding their application and highlight necessary elements for success.

The post-class evaluation can provide guidance on course materials and structures. Since it only signals intent, it is not a true gauge of whether knowledge and skills are transferred, but if there are deficits in material or instructor technique, future classes will suffer and they, therefore, need to be noted and corrected. Individual evaluation forms and techniques varied and have various advantages/disadvantages. While this paper will not delve into this subject, the evaluator should be aware of these advantages/disadvantages. Also, due to individual tastes, different learning styles and different intelligences, the evaluators should be analyzing the post class evaluation for trends as opposed to individual comments.

The knowledge based test enables the organization to again gauge the material and instructor. In all cases there will be key pieces of information that must be conveyed to the attendee. This test allows the company to gauge if this information has been conveyed effectively. Again, due to individual differences, the trend is of the highest value for future improvement.

The evaluation of the transfer level is of key importance. This is where the actual knowledge or behavior is to be utilized. At this level of evaluation, the attendee is expected, at some level, to show the knowledge or behavior. In the case of the Powered Industrial Truck standard, it could be by performing a certification and "driving test". There are some key factors to consider at this point. What are you comparing the attendee to? This expectation must be clearly laid out. In formal training development the evaluation is determined by a comparison to the class goals and objectives^{vi}. Training evaluations do not always have to follow this formal process, but it must be clear what the attendee will be compared to. The PIT standard can be used to illustrate this. What tasks will the attendee be expected to perform as part of their normal job function? They will have to drive around the facility, properly lift loads, and place the loads on racking. If this is the extent of their job tasks, this may be an adequate evaluation. What if their job include loading/off loading of trucks? Then a component must be built into the evaluation.

Finally, the facility (or company) should see less forklift accident, incident, near misses or property damage some time after the completion of the training. The issues with this level is has enough time passed since the training for the knowledge and behaviors to be effective and was there a process in place to measure these item before the training began.

When these levels are used in conjunction, they provide a wealth of data. If the only criterion of success was the reduction in accident/incidents, what happens when the expected reduction in accident/incidents is not actualized? Is the training ineffective? In what way? How long until it is decided that the training is lacking? By utilizing the Kirkpatrick scheme, data is obtained to highlight where the problem is occurring. Consider:

Evaluations and knowledge tests can indicate deficiency in the material or instruction styles. The basic concepts are sound, but the presentations themselves may not be conveying to the attendees the necessary information. At this point a reworking of the presentation may correct the problem.

Issues at the transfer level can provide useful information. If the attendees are not applying the skills (and passing the evaluation) the problem may be the course is not conveying the correct behaviors or knowledge. Even if the attendees are “passing” the evaluation, consecutive failure on one particular area (a trend) will highlight the need to improve that area in the class materials.

If in the end, the previous three levels are passed consecutively, and the facility does not realize a reduction in forklift accident/incidents, the initial assessment must be questioned. The knowledge and behaviors are being observed, but the benefit is not observed so the basic premise must be readdressed. Perhaps the issue is one of enforcement or policy.

So by using Kirkpatrick’s principles the deficiency in the program can be pinpointed and corrected in the shortest amount of time, as opposed to a method without the various steps of evaluation in which trial and error would be used to try to correct the problem. This trial and error method will waste resources which could have been used in a more constructive manner.

In conclusion, the need for training is obvious. If the resources are placed on conducting training, it must be determined if that training is meeting its goals. How a “Unit” does this is dependent on several logistical and cultural factors of the unit. In some instances informal evaluation may be a valid and efficient technique. However, for those instances where formal evaluation is needed, by arranging the evaluation by the schemes put forth in this paper the unit can obtain the most information from the evaluation and use their resources in an effective manner.

Bibliography

Cantonwine, S. C. *Safety Training That Delivers*. Des Plaines, IL: American Society of Safety Engineer, 1999

Dell Publishing. *The American Heritage Dictionary, 3rd Edition*. New York: Dell Publishing, 199

Federal Emergency Management Agency, “Management and Supervision of Training- Student Manual”, FEMA’s Management and Supervision of Training Course.

Kirkpatrick, D.L. (1994). *Evaluating Training Programs: The Four Levels*. San Francisco, CA: Berrett-Koehler

ⁱ (Dell Publishing 293)

ⁱⁱ(Cantonwine 36)

ⁱⁱⁱ(Cantonwine 36)

^{iv} The information on Kirkpatrick was summarized from Kirkpatrick, D.L. (1994). *Evaluating Training Programs: The Four Levels*. San Francisco, CA: Berrett-Koehler

^v 29 CFR 1910.178 (I) (4)

^{vi} FEMA Student Manual