

TRAINING EVALUATION PROCESS AND TOOLS

Dana C. Reinke, Sociologist
Launa Mallett, Ph.D., Sociologist

Pittsburgh Research Laboratory, NIOSH

One of the outcomes of the promulgation of 30 CFR, Part 46 has been increased awareness of the need to evaluate mine health and safety training. The new law explicitly cites knowledge of evaluation techniques as part of the definition of a competent trainer. It further encourages evaluation by including it as one of the required components of a training plan that can be implemented without MSHA review. With the attention given evaluation in Part 46, it is likely that mine trainers are also reviewing the evaluation requirements of Part 48. Under this law, task training must be evaluated. The question for many health and safety managers, resource personnel, and executives is how do they begin to evaluate their training program? To answer this question, they need to have an understanding of the theory and methodologies used in evaluation research. This paper will provide a background for understanding what evaluation is and the techniques that can be used to examine the effectiveness of a training program.

The term evaluation can be used to define multiple activities that result in a critical review. For purposes of this paper, evaluation is defined “as the process of determining the value and effectiveness of a learning program” (Clark, 1997). Value and effectiveness are subjective terms and they must be defined with standards set for each in an evaluation plan. The plan should answer the questions below which will be discussed in the following sections.

- 1 - What is the purpose of the evaluation?
- 2 - What will be measured?
- 3 - How comprehensive should the evaluation be?
- 4 - Who determines and oversees the evaluation?
- 5 - What is the source of the data and how will the data be collected and compiled?
- 6 - How should the data be analyzed and presented? (Clark, 1997)

An evaluation strategy addressing these questions should be created at the same time that the training is being developed. Incorporating an evaluation component into a training plan will help to ensure that the evaluation focuses on topics, practices, or skills that are desired results of the training and that data can be gathered at appropriate times before, during, and/or after training implementation. Even if the training program is being developed by a contractor or outside vendor, the contracting organization should be involved in this initial planning phase.

Purpose of the Evaluation

Just as a training program should have specific goals, the evaluation plan should also be based on explicit goals. Identifying evaluation goals should be completed by the organization before the training is designed and implemented. Specific goals that might be included in an evaluation plan can be determined by the following categories.

- 1 - Feedback- quality control to determine how learning outcomes link to initial training objectives.
- 2 - Control- assessing the training program in relationship to other organizational activities and to consider cost effectiveness.
- 3 - Intervention- to make changes to improve future training program implementation. (Blamey and Newby in Clark, 1997)

In mine safety and health training, one kind of feedback often needed is whether or not training is meeting the requirements of federal and state training regulations and policies. This goal should be stated and the specific rules to be met should be identified in the evaluation plan.

Another important determinant of the purpose of an evaluation is the audience who will be interested in the results. This could include trainers, trainees, managers, government officials, future employers of contract training firms, and others. Determining who will have access to the results and their needs should be part of the planning process. For annual refresher training, under Part 46, could include MSHA personnel and state mining officials who will want to see how the training plan and/or specific training activities meet legal requirement. Trainers will want results that can assist them as they plan future training sessions. Mine managers may want cost/benefit related information for future planning needs. On the other hand, do not include goals that are not of real interest to some specific party. If findings will not be used, it is not effective to put resources toward obtaining those results.

Evaluation Target

Evaluations can address multiple questions about a particular training program. In the late 1950's, Donald Kirkpatrick developed one framework that can help explain what can be learned by evaluation (Nichols, 2000). Over forty years later, Kirkpatrick's four levels are still used in discussions of training program assessments. Level 1 is the easiest and least resource intensive

to conduct. Level 4 is the most difficult and expensive (Kirkpatrick, 2001, pp. 122-132). "As we move from level 1 to level 4, the evaluation process becomes more difficult and time-consuming, although it provides information of increasingly significant results." (Clark, 1997) Each level is briefly described in the table below. A discussion of how the levels could be defined for safety and health training follows in Table 1.

Level 1 – This is the most common kind of training evaluation. It frequently consists of post-training questionnaires asking for the trainees' opinions about course content and/or the instructor. It can include questions related to the class setting and other factors that could impact the effectiveness of training. This level of evaluation is particularly useful to trainers who are interested in improving future training sessions. It is also possible to determine outside variables that could be impacting training such as uncomfortable classrooms or unfamiliarity with training tools or methods. While questionnaires are often used to gather data regarding trainee reactions to training, it is not the only method to obtain this data. Other methods will be discussed in a later section. Level 1 will be included in most evaluation plans.

Level 2 – This level focuses on any change in knowledge or skill that can be attributed to training. This requires measuring the knowledge or skill before and after training. Additionally, any other factors that could have caused the change must be explored. The most rigorous study model requires testing two equal groups, then administering training to only one group, and finally retesting both groups. Assuming no outside influence has made the groups dissimilar during the training phase, any differences recorded between changes in performance of the untrained and trained groups would be attributed to training. In safety and health training it is sometimes not ethical or wise to withhold training from one group and therefore a true experiment cannot be

Table 1. Evaluation Levels

Level	Measurement Focus	Questions Addressed
1 - Reaction	trainees perceptions	What did trainees think of this training?
2 - Learning	knowledge/skills gained	Was there an increase in knowledge or skill level?
3 - Behavior	Work-site implementation	Is new knowledge/skill being used back on the job?
4 - Results	impact on organization	What effect did the training have on the organization?

conducted. In these cases, measures of pre- and post-training scores should be reported along with anything outside of the training that could have impacted those scores.

Level 3 – Complexity increases for this level because some of the data should be gathered some time after the training has been completed. To assess whether or not what was learned during the training is making any difference in the workplace, trainees must have the opportunity to use their new knowledge and skills. Ideally, activities in the workplace should also be measured before training so change can be assessed. But “...something beats nothing, and I encourage trainers to at least do some evaluation of behavior, even if it isn’t elaborate or scientific. Simply ask a few people: ‘Are you doing anything different on the job because you attended the training program?’ If the answer is yes, ask, ‘Can you briefly describe what you are doing and how it is working out? If you are not doing anything different, can you tell me why? Did you learn anything that you can use on the job?’” (Kirkpatrick, 2001. p. 128) If Level 2 data shows that something was learned, but Level 3 data reveals it is not being used in the workplace, then the problem may be related to organizational variables such as authority hierarchies or access to equipment or tools.

Level 4 – To conduct an evaluation at this level a clear link between the training goals and the needs of the organization must exist. This level shows what impact the training program has at an organizational level and is frequently tied to costs/benefits types of analyses. Accident and injury data is often used in computations for

safety and health training at this level. Other information such as changes in use of sick leave, absenteeism, costs of personal protective equipment could also provide information about organizational change resulting from training. For this large-scale analysis, variables outside of training may interfere and make measurement of training effects difficult. Because of this proof that training has made an organizational change is not required for findings to be of value. Kirkpatrick says to, “...be satisfied with evidence, because proof is usually impossible to find.” (Kirkpatrick, 2001. p. 129) A well-planned evaluation can provide compelling evidence about the organizational impact of a training program.

Scope of the Evaluation

The comprehensiveness of the evaluation should follow its initial goals. An evaluation of an entire training program will be resources intensive. It should be conducted when an assessment of the program is needed for policy decisions about the worth of the program and/or changes needed to improve on ongoing program’s effectiveness. Often only one portion of a program is the appropriate target for an evaluation. One course or a set of courses covering related material is assessed at completion to report back to interested people about the course(s). Similarly certain instructors can be assessed by gathering data only about their work. The desire for information should always be balanced with the time, personnel, and funding available for the evaluation tasks. A clearly defined evaluation plan will keep data-gathering focused on what is really needed to

answer the question of importance for the intended audience of this evaluation.

Internal versus External Evaluation

The answer to the question of who should design and conduct an evaluation depends on the purpose of the assessment and the resources available. If a person who is part of the program being evaluated also has a role in the evaluation, he or she can add insight obtained by their other program tasks and involvement. There will be little time needed to learn about the structure of the program and its goals. It is often also cost effective to use someone already being paid by the organization for program-related tasks. On the other hand, it may be difficult for a program staff member to assess it objectively. This is particularly problematic when an evaluation could have a negative impact on this individual's career. Even when an internal evaluator wants to conduct an objective evaluation, participants are sometimes reluctant to give negative feedback to instructors or others with direct relationship to the program.

When an external evaluator is selected to assess some aspect of training, the chance for bias decreases. Participants are more likely to report their opinions to outsiders if they have anything negative to say. Even constructive criticism is provided more openly to a neutral party. But there are some cons to choosing external evaluation. Usually they require more resources than internal assessments. Even if the evaluator is brought in from another part of the organization and therefore is not a contracting cost, the evaluation will take longer than an internal one so it will cost more in salary. This is because the external evaluator will have to "get up to speed" on the structure and purpose of the program as part of assessment activities. An external evaluation may add objectivity, but will also add cost. These two competing factors will have to be weighted when determining evaluation goals.

Methods for Data Collection

Data gathering techniques are the most well known aspect of evaluation. Related tools are used to collect data for assessment and for validation purposes. Assessment is the measurement of the practical results of the training and validation determines if the objectives or goals of the training were met (Clark, 1997). The first column of Table 2 lists techniques that can be used to gather evaluation data. The second column tells when this technique would be used and the third column addresses what findings can be gathered with each method. (Mallett and Reinke, in press) The following text discusses the use of each technique in more detail.

Questionnaires. A relatively quick way to get a limited amount of information from a large number of people is a written questionnaire. A well-designed questionnaire can get valuable data, but a poorly designed questionnaire will simply be a waste of time. Some important considerations when designing or assessing a questionnaire are the ambiguity of the questions, the kinds of answers allowed, the formatting of the questions, and the appropriateness of the questions with relationship to evaluation goals.

Questions must be written clearly so that they will be interpreted in the same way by all respondents (Babbie, 1998). For example a question that could be asked to assess a safety training class is: "*Did you learn something new during this class?*" This question can be answered in terms of what was learned about the instructor, the format of the class, a specific topic covered or not covered, or the trainee's perception of management support for the training. The trainee may also elect to put a simple Yes or No as a response. A more specific way to ask this question would be: "*Please list any new first-aid techniques you learned during this class?*" This more targeted question is more likely to meet the evaluation goals.

Table 2. Data Collection Methods

Method	When Used	What can be Learned
Questionnaire	Before, during, or after training	<ul style="list-style-type: none"> • Perceptions of trainees or supervisors • Opinions of content or training experts • Knowledge or skills • Transfer of training to job • Organizational impact
Interviews	Before, during, or after training	<ul style="list-style-type: none"> • Perceptions of trainees or supervisors • Opinions of content or training experts • Knowledge or skills • Transfer of training to job • Organizational impact
Classroom observations	During training	<ul style="list-style-type: none"> • Perceptions of trainees
Performance tests	Before, during, or after training	<ul style="list-style-type: none"> • Trainee skills
Written tests	Before, during, or after training	<ul style="list-style-type: none"> • Trainee knowledge
Workplace observations	Before or after training	<ul style="list-style-type: none"> • Trainee knowledge or skills • Transfer of training to job
Games	During training	<ul style="list-style-type: none"> • Trainee knowledge or skills
Group discussion	Before, during, or after training	<ul style="list-style-type: none"> • Perceptions of trainees or supervisors • Opinions of content or training experts • Knowledge or skills • Transfer or training to job
Analysis of statistics	Before or after training	<ul style="list-style-type: none"> • Organizational impact

Questions can be designed to be closed-ended or open-ended. Closed-ended questions offer specific answers for the respondent to select. It is important that all possible answers are covered and that each answer is distinct from the others. Close-ended questions result in data that are easy to analyze but limit the amount of depth provided in the responses (Babbie, 1998). An example of a closed-ended

question is: *“Would you recommend this class to other employees with your level of experience? (Please circle your answer) Yes No”* An open-ended version of this question could be: *“What about this class would make you recommend or not recommend it to other employees with your level of experience?”* An open-ended question can be used to gain more detailed information, but it is more difficult to

analyze. Sometime open and closed-ended questions are combined. For example you could ask: *“Would you recommend this class to other employees with your level of experience? (Please circle your answer) Yes No, If Yes, why?”* The amount of detail wanted has to be balanced with the effort available to analyze responses.

Even if all the questions are well written, optimal data will not be gathered if they are not formatted appropriately on the page. A questionnaire should not intimidate the respondent with too many words, too small a font, unfamiliar technical terms, or too high a reading level. There should be plenty of space available for answers to be written and directions should be clear. It is better to get fewer well thought-out answers than to get many that are hurried through or are incomplete because of limited space on the page. After a questionnaire is drafted it should be tested with people similar to those who are the targeted respondents.

Interviews. Interviews can also be used to capture trainee impressions of the training session. Many trainers already ask trainees their opinions about training sessions without thinking of the conversations as interviews. Certain techniques can be followed that can make these conversations valuable evaluation tools. Compared to a survey, interviews are more flexible and ongoing, but interview techniques apply many of the skills used to construct effective surveys (Babbie, 1998). Interviews must also be designed with clearly written questions. But when the questions are asked orally, further explanation is possible if the trainee doesn't understand. This of course requires a full understanding of the questions by the interviewer. If multiple people will be gathering the information, they must be trained so that all of them understand the questions and are asking them in the same way. Interviews are best when detailed information is wanted from a limited number of people.

During an interview, open or closed-ended questions can be used. The best feature of the interview is that follow-up questions can be asked to illicit responses with more depth or that more closely target the evaluation needs. An example of a question with targeted follow-ups is: *“Have you ever taught someone to operate this kind of equipment? Yes No; If they answer yes ask: What written resources (examples: checklists, manuals, laws, safe operating procedures) were available to you to use during that training? What kind of written materials would help someone who is training another miner to operate this equipment?”* The interviewer must know enough about the topic to understand the answers give to open-ended questions and to ask for clarification of answers when they are not easily understood.

Classroom Observations. Trainers observe the activities occurring during their classes. They watch trainee behavior and responses to ideas, activities or questions. Documentation of these observations is another source of data for training evaluation. It is best to write down what the trainer or another observer sees while the session is occurring. Actions (both verbal and non-verbal) should be written down as well as comments that seem to form a pattern of activity, change the course of the session, or stand out as noteworthy. If notes cannot be taken during a session, then quick notes using short words and phrases to describe events should be written as soon as possible after session has ended. The trainer or observer should record what they ‘know’ happened and what they ‘think’ happened during the session (Babbie, 1998). After completing this quick sketch of the session, they should review the notes and fill in the details as soon as possible. The longer the delay in writing down observations from a session the less likely actions and events will be recalled accurately and completely. It is helpful to have an observation guide such as the one found in Appendix A to help the trainer or observer to remember to take note of important actions.

Written Tests

Trainee knowledge can be assessed with written examinations. Care should be taken when using this method, as many working adults will become anxious when they think they are being tested. The tests must be constructed in a manner that allows trainee knowledge to be written and interpreted by the trainer. A well-designed exam should allow those with more knowledge of a topic to perform better than those less knowledgeable. To determine if trainee knowledge change is a result of training, a pre-test and post-test will have to be administered. Outside influences can impact the results of this kind of study as was discussed earlier. If the purpose of the test is to assess the quality of the training rather than the performance of the individual trainees, they should be aware of that purpose. Their anxiety level may be somewhat lessened in those situations.

Workplace Observations. Like classroom observations, workplace observations are already taking place in many instances. To include them training assessments, they need to be formalized and documented. The trainer, a supervisor, or some other observer can conduct the observations. This method is useful for determining if skills learned during training can be transferred to the work environment. A checklist of what should be observed should be used. This checklist could be created as a job-aid and also used in training.

Games. Trainers often use games during their courses. These break up the day and allow trainee interaction. They can also be effective evaluation tools. Trainers can determine if trainees know select materials by the answers they provide during an adapted version of football, bingo, or a television game show. A crossword puzzle can be used as a review or to test knowledge of terminology. Games are a good way for trainers to informally determine if their course is working for this particular group of trainees. The games should always focus on important information and anytime an answer is incorrect or false, the correct answer must be presented and discussed. Trainees often enjoy

classroom games and do not think of them as tests, but they can be used to evaluate training effectiveness.

Group Discussion. Interviews can also be conducted in groups. The questions asked should be defined, but the interaction created by multiple individuals answering will make recording the responses more difficult. There may be more information obtained as people make comments that stimulate thoughts in others in the group. On the other hand, if the group members are not comfortable with each other, they may not express their real opinions. For example, care must be taken when discussing work with a group containing both employees and supervisors. Group composition, finding a comfortable location for the discussion, the questions to be asked, and how the answers will be captured must be considered designing an evaluation with this method included.

Analysis of Statistics. Sometimes analyses of statistics such as accident rates, near misses reported, maintenance costs can be used to evaluate training. When designing a training evaluation, the availability of records should be explored. Whether or not the numbers can be tied directly to the training activity is an important consideration. Outside variables that could impact the statistics must be considered and clearly stated in the final report if this type of data is used.

Analysis and Reporting of Results

If the implementation of the evaluation followed a clear design, then analyses will be straightforward. The data collection methods that were used are tied to given analysis procedures. The most important thing to remember is to target the analysis to questions asked when the evaluation goals were defined. While this may seem obvious, it can be tempting to explore other things that can be found in the data. While it may be appropriate to scan the data for unplanned outcomes, this will take resources away from the initial goal and

should be done only after the primary evaluation goals are met.

Reporting of evaluation results can take many forms. Written reports, oral briefings, or multi-media presentations can be used to convey findings. Often results should be reported in more than one way to meet the needs of different interested parties. More or less detail can be provided and the level of sophistication of the results can be varied. In all cases it is important to limit any public availability of information that can harm or embarrass any individuals. When assessing trainers, it may be necessary to provide negative feedback to that person and/or related managers. In those cases, care should be taken to offer the data in as positive or constructive manner as possible. Any time public harm or embarrassment results from an evaluation report, everyone involved and people who simply hear of what has occurred will be less likely to participate openly in future evaluations. A determination of who will have access to what kinds of results should be determined during evaluation planning and reporting should follow those guidelines.

In Summary

Developing and implementing a training evaluation is sometimes thought of as an abstract and unordered process. In fact a well designed evaluation is a structured process that produces results useful to an organization.

When designing a training evaluation it is critical to address specific issues. These evaluation characteristics include determining the purpose of the evaluation, what aspects of a training program need to be evaluated and the individuals who want to review or need access to the evaluation results. After those items are defined, choosing who will create and execute the evaluation, the resources available for implementing the evaluation, and the collection

methods used to gather data can be determined. An organization taking the initial steps towards producing a training evaluation needs to integrate their specific organizational qualities and needs into the evaluation design plan. A worksheet is provided in Appendix B to serve as an outline for the development of a training evaluation.

REFERENCES

Babbie, Earl (1998) *The Practice of Social Research*. Wadsworth Publishing Company Belmont, CA.

Clark, Donald (1997). "Instructional System Development - Evaluation Phase - Chapter VI", <http://www.nwlink.com/~donclark/hrd/sat6.html>

Dopyera, John and Louise Pitone (1987). "Decision Points in Planning the Evaluation of Training", in More Evaluating Training Programs: A Collection of Articles from Training and Development Journal, Compiled by Donald L. Kirkpatrick, American Society for Training and Development, Alexandria, VA, pp. 74-77.

Kirkpatrick, Donald (2001). "The Four-Level Evaluation Process", Chapter 12 in What Smart Trainers Know: The Secrets of Success from the World's Foremost Experts, Lorraine L. Ukens, editor, Jossey-Bass/Pfeiffer, San Francisco, pp. 122-132.

Nichols, Fred (2000). "Evaluating Training: There is no 'cookbook' approach", <http://www.nwlink.com/~donclark/hrd/sat6.html>

U.S. Code of Federal Regulations. Title 30 – Mineral Resources; Chapter I – Mine Safety and Health Administration, Dept. of Labor, Parts 46 and 48, July 1, 2002.

Appendix A

Guide for Observation of Simulation/ Response of Trainees

Before training starts, observe the trainees and make note of this starting point. The trainees' voices, sitting positions, body language and comments can serve as a control or point of reference to compare those same actions as the class progresses.

Observe Changes/Lack of Changes in the Following:

Body Language

- Facial expressions that signify stress, anxiety, comfort, confidence, etc.
- Changes in posture, position in chair
- Movement of hands
- Movement of feet/legs
- Changes in voice/language/delivery

Social Interaction

- Observe group dynamics – did they work together or separately?
- If they worked in groups, did those groups change over the course?
- Were there particular points when trainees asked each other for help?

Interaction with Training Materials or Multimedia

- Were trainees comfortable with the format of the material?
- Did they have problems using the technology?
- Did they assist each other with the materials?
- Did they seem to understand the content of the materials?

Appendix B Training Evaluation Worksheet

Use this worksheet as a guide to help organize the development of a training evaluation.

Plan the Evaluation

1. What are the goals of the evaluation?
2. What questions about the training will the evaluation answer?
3. Who will design and conduct the evaluation?
4. What resources are available for evaluating the training program? (Time, Money, Equipment)

Gather the Information

1. What method(s) will be used to gather information?

Data Collection Method	Performed (circle choices)
	Before, During or After Training
	Before, During or After Training
	Before, During or After Training
	Before, During or After Training

2. How should the gathered data be presented?

PROCEEDINGS THIRTY-THIRD ANNUAL INSTITUTE ON MINING HEALTH, SAFETY AND RESEARCH

**ROANOKE, VIRGINIA
AUGUST 27 - 30, 2002**

Editors

George R Bockosh
Senior Scientist
National Institute for Occupational Safety and Health
(NIOSH)
Pittsburgh Research Laboratory

Jeffrey Kohler
Research Director
National Institute for Occupational Safety and Health
(NIOSH)
Pittsburgh Research Center

John Langton
Chief, Division of Safety, Coal
Mine Safety and Health Administration
U.S. Department of Labor

Tom Novak
Professor and Head
Department of Mining and Minerals Engineering
Virginia Tech

Michael K. McCarter
Professor and Chair
Department of Mining Engineering
University of Utah

Angelo Biviano
Writing and Communications Program Coordinator
Department of Mining and Minerals Engineering
Virginia Tech

Sponsors

Department of Mining and Minerals Engineering
Virginia Tech

Department of Mining Engineering
University of Utah

Mine Safety and Health Administration
U.S. Department of Labor

National Institute for Occupational Safety and Health
Centers for Disease Control and Prevention

National Mining Association

National Stone, Sand & Gravel Association

Virginia Aggregates Association

Bituminous Coal Operator's Association

Society of Mining, Metallurgy, and Exploration

Published By

Department of Mining and Minerals Engineering
Virginia Polytechnic Institute and State University
Blacksburg, Virginia 24060
(540) 231-6671

TN 295
.I59
2002
C.2