

Measuring participant performance: An alternative

Susan Bumpass and David Wade
Professional Education Division
Arthur Andersen and Co

This article presents the reasons for having faculty appraise participant performance during training. A methodology used to develop such an appraisal is explained and a sample of the behavioural description used is provided. The preliminary results suggest that faculty appraisal of participants' performance during training is a useful evaluation alternative.

As the amount and need for training increases, management is asking, "What am I getting for the money I spend?" The answer from the training community has been an increase in the evaluation of training but often without answering management's question.

This article explores why the prior statement is true and presents the preliminary findings of an alternative method of evaluation and training: faculty appraise participant performance during and at the end of a training program.

Levels of evaluation

Thirty years ago, Kirkpatrick (Kirkpatrick, 1959a, 1959b, 1960a, 1960b) introduced a model of training evaluation. His model has withstood the test of time and is useful in developing evaluation procedures, methods and instruments. Kirkpatrick's model has four levels:

<i>Reaction</i>	The reaction level evaluation measures the subjective views of the participants to the training, such as rating the overall quality of the training or were the objectives met.
<i>Knowledge</i>	The knowledge level of evaluation measure the extent to which the participant actually learned the material presented. Testing is most often used to evaluate the transfer of knowledge at this level.
<i>Performance</i>	Performance level evaluations are designed to measure if the participant can demonstrate the transfer of training on the job or in a simulation. Performance appraisals are most often used to measure this level.
<i>Organisational</i>	At the organisational level of evaluation instruments and methods are designed to determine what economic or psychological effects have occurred; such as is the organisation more productive, have attitudes changed?

Reaction and knowledge levels of evaluations are most commonly used. According to a recently published study by the American Society of Training and Development (Carnevale and Schultz, 1990), of all the organisations represented in the study:

... 75 to 100 percent of them evaluated training programs at the participants reaction level. Virtually all of them also evaluated participants' knowledge gain in some of their training programs. Twenty-five percent of their training programs were evaluated at this, the learning level.

Although data collected at the reaction level can provide valuable information, unfortunately it does not provide evidence that the transfer of knowledge has occurred. Therefore, reaction level evaluation cannot address management's question.

Testing is the method most commonly used to evaluate participant knowledge. When developed properly, tests provide an objective and reliable estimate of participant knowledge. In addition, a great deal of research has been done on testing; over 1000 articles in refereed journals since 1980. Testing, as an evaluation technique, is well understood. However, tests have the following pitfalls:

- Tests are time consuming and expensive to develop: A minimum of one hour development time per item with a pool of over 200 items is not uncommon.
- Often the item review and item analysis steps are eliminated or curtailed, decreasing the level of confidence that can be placed in the data.

- If the content is highly volatile, such as tax law, test items are expensive to maintain.
- Despite many safeguards, the potential for "cheating" remains quite high.
- Paper and pencil tests do not reliably measure interpersonal or performance skills such as presentations.

Although tests provide a measure of knowledge gains, the data collected does not indicate whether new knowledge is successfully used on the job to enhance job performance.

Performance and organisational levels of evaluations are not frequently conducted due to the difficulty and cost involved in collecting reliable and valid data. Again in the same study by the American Society of Training and Development, of all the organisations represented in the study, only about 10% evaluate training at the behaviour/ performance level and only about 25% at the organisational level.

Despite the difficulty, performance and/or organisational level evaluation data must be collected if the training community is to adequately respond to management's question.

The remainder of this paper presents a data gathering technique that is currently used to appraise participant performance during simulated job situations.

An alternative method

In the organisation under study, instructor-led training is primarily composed of job relevant simulations. The simulations require participants to apply knowledge gained during the completion of pre-requisite training to case studies that reflect the work and situations they are likely to encounter.

The skills to be addressed by the cases are determined by needs assessment and analysis. Once the skills are identified actual events are selected as the basis for the case studies. The simulations are built around the case studies and facilitated during the class by supervisory/ management level personnel who were originally involved in the case situation. The facilitators are selected based on their "exemplary performance" in the skill area and the case being discussed.

Evaluation data is collected and used at the reaction and knowledge levels; for example, testing is used to ensure mastery at the end of pre-requisite self-study training and follow-up interviews conducted with participants and their supervisors at three and six month intervals after training.

However, an evaluation technique that would provide immediate feedback and be relatively inexpensive and cost effective to implement and maintain was needed to measure knowledge gains and performance during the instructor-led training. Since the faculty are line managers who are experts in the content presented and have ongoing responsibility for ensuring and appraising the quality of work performed in this area, one alternative seem to be faculty appraisal of participant performance.

A literature search was performed to determine if and how such appraisals were performed. However, there were no articles on instructor appraisal of participant performance in a business training and development setting. But there were many articles and books on how to conduct performance appraisals. Therefore using this latter literature, developing forms upon which faculty could appraise participant performance during training appeared to be a viable alternative.

There are several reasons for having faculty appraise participant performance:

- | | |
|-----------------------------|--|
| <i>Motivation</i> | Participants seem to work and study harder when they know they are to be appraised. Because the criteria are distributed at the beginning of each school, participants can identify faculty expectations of them. This encourages participants to match their behaviour to the performance items being measured. |
| <i>Measurement</i> | Appraising participant performance provides a way to determine the extent of training transfer. The cumulative results can be used to determine the need for training revisions. |
| <i>Standards</i> | Faculty appraisals can be used to determine whether participants have met the criteria necessary for progressing to the next level of training or can be certified as having mastered a domain of expertise. |
| <i>Selection Confidence</i> | Only those skills that are critical to on the job performance are measured. Adding these appraisals to other performance measures helps promotion decision makers reach better selection judgments. Multiple ratings by various instructors helps to develop a consensus view of a participant's performance and as the pool of observations goes up, the possibility of making an inaccurate assessment/selection judgment goes down. |

Advantages and disadvantages

Like all measurement instruments and procedures, faculty appraisals of participant performance has advantages and disadvantages. Faculty appraisals of participant performance are less expensive than tests to develop, require only minimal maintenance (objectives must be updated as part of any content change to training), and if multiple faculty are used during each training session and each is trained in the use of the form, rating bias and idiosyncratic rating errors can be controlled. One way to control any impact of rating errors is to distribute a frequency distribution of the scores along with individual evaluation forms. This assists in more reasonable interpretations of the results.

The major disadvantages of faculty appraisals is that such appraisals are subjective. More analysis such as multi-rater/multi-method techniques or three way ANOVAs needs to be performed to determine the level of confidence training developers can have in the data collected and how best to use the data. Although additional research has been conducted to ensure the instrument is reliable, there is no consensus in the performance appraisal literature as to which statistical test should be used to make this determination.

Methods used to develop the instrument

The following describes the methodology used to develop the current training performance appraisal procedures and form:

1. Establish the appraisal criteria.

A group of line personnel (incumbents and supervisors) were asked to reach agreement on the broad areas that are critical for successful job performance at a given personnel level. Specific objectives were then determined to support each agreed to area. Examples of the areas established were:

- Technical knowledge
- Business knowledge
- Communication Skills

2. Obtain an expert to write the behavioural descriptions.

An assessment/ performance management expert in the central training function was asked to write a behavioural definition for each area. For example:

Communication skills (Oral). The participant used effective presentation skills during classroom discussion. Ideas were communicated logically and concisely using an authoritative image, clear enunciation, and good voice projection.

These statements were reviewed by a sample of the line personnel who would be using the instrument. This review was a check to ensure that the criteria descriptions were complete, unambiguous and accurately reflected those behaviours deemed to be critical to success. Three review cycles were needed to complete this portion of the process.

3. Establish a rating scale.

A 5 point Likert-type scale was selected to assess the degree to which the participant performed the skill described by each description.

Rating scale

5	Very much so
4	For the most part
3	Somewhat
2	Only slightly
1	Not at all
N/A	Not applicable

4. Develop implementation procedures.

Next, working with the line personnel, procedures, training, and guidelines were developed for administering the instrument. Some of the procedures developed are as follows:

- Instruct each faculty member to evaluate every participant at the completion of all topics within a school. (It is important to note that more advanced training requires a greater number of faculty to present the material.)
- Instruct faculty not to discuss impressions of participants until the results are summarised. This provides a larger number of independent ratings for use in completing the final evaluations.
- Aggregate all faculty ratings for each participant. Send the individual aggregated ratings, a frequency distribution of all participant ratings and interpretation directions to the local office (see Figure 1).
- Instruct the local office to review the ratings and discuss them with each participant. These tasks are usually assigned to the faculty member from that office. At this point supplemental work or other corrective action can be taken if necessary.

QUALITIES	RATINGS*					TOTAL
	5	4	3	2	1	
COMMUNICATION Communication Skills (Oral)	1	6	21	8	-	36

* The numbers in each column indicate the total number of participants in the section receiving that mark on their evaluation.

Figure 1: Example faculty appraisal of participant performance frequency distribution

Using the Figure 1 frequency distribution, if a participant was given a rating of 2 in Oral Communication Skills, the office would interpret that the individual performed below the minimum acceptable level of 3.0 and the level demonstrated by his/her peers during the training. However, if the summary had shown that a majority of participants were scored 2 and 1 in communication skills with only a few participants receiving a rating of 3, the interpretation would be different.

The course developers also use these frequency distributions as noted in the discussion of item 7 on the following page.

5. Test the instrument and procedures.

Feedback from faculty, participants and division heads who must use the instrument and interpret its results is being gathered. Users and participants are asked to identify any ambiguities or other problem areas. An analysis will be conducted to determine the degree to which the instrument accurately predicts successful performance on the job. The combination of measures will provide some additional insight not only into the gains made and the reason for those gains but other influences that may be affecting performance.

6. Refine and revise the instrument and procedures.

As feedback is collected, problem areas are logged (e.g., different people interpreting the criteria in different ways) and possible corrective action steps identified.

7. Analyse participant performance rated below 3.0.

The performance of participants who have aggregated ratings below the minimum acceptable level of 3.0 on any item is further analysed to determine the reasons for such deficiencies. As a result of these findings, other action may be taken. In some cases changes to the design of the training may be deemed an appropriate corrective action.

Such could be the case with the example described following Figure 1 on the prior page. In that example a frequency distribution indicated that almost an entire class rated below the minimum acceptable performance level in oral communication skills. One of the options to be considered, following an analysis of the cause of the deficiency, is modifications of the training to help participants enhance this skill area.

Summary

Faculty appraisal of participants performance in a training environment is "virgin territory." Organisations that spend money on training want to know:

1. That the training they use will positively affect on the job performance and
2. How well their participants perform.

As with all training investments, the costs and benefits of alternative methods of measuring participant performance must be determined and considered in order to make effective decisions. The authors believe that organisations will be more willing to continue to fund training if evaluation methods designed to capture results at Kirkpatrick's performance level are used. Accordingly, this evaluative technique deserves further exploration and development.

References

- Kirkpatrick, D. L. (1959a). Techniques for evaluating training programs. *Journal of ASTD*, 13(11), 3-9.
- Kirkpatrick, D. L. (1959b). Techniques for evaluating training programs: Part 2 - Learning. *Journal of ASTD*, 13(12), 21-26.
- Kirkpatrick, D. L. (1960a). Techniques for evaluating training programs: Part 3 - Behaviour. *Journal of ASTD*, 14(1), 13-18.
- Kirkpatrick, D. L. (1960b). Techniques for evaluating training programs: Part 3 - Results. *Journal of ASTD*, 14(2), 28-32.
- Carnevale, A. P. and Schultz, E. R. (1990) Return on Investment: Accounting for Training. *Training and Development Journal Supplement*, July, 44(7), 5-24.

Bibliography

- Alliger, G. M. and Janak, E. A. (1989). Kirkpatrick's levels of training criteria: Thirty years later. *Personnel Psychology*, 42, 331-342.

Wexley, K. N. and Baldwin, T. T. (1986). Post training strategy for facilitating positive transfer: An empirical Exploration. *Academy of Management Journal*, 29, 503-520.

Address for correspondence: Susan Bumpass, Professional Education Division, Arthur Andersen and Co, GPO Box 5151AA, Melbourne, Victoria 3001.

Please cite as: Bumpass, S. and Wade, D. (1990). Measuring participant performance: An alternative. *Australian Journal of Educational Technology*, 6(2), 99-107. <http://www.ascilite.org.au/ajet/ajet6/bumpass.html>