Evaluation: Towards a definition and statement of purpose

M Jane Marsden

This paper explores various interpretations of the term "evaluation" in the context of training and development. It looks specifically at definitions and purposes of evaluation. It is proposed that, in Australia's current economic state and in the environment of the Training Guarantee, training professionals need to address the need for comprehensive evaluation practices.

The purpose of this paper is to develop a definition and statement of purpose for evaluation in the context of training and development. Evaluation is given a low priority in the instructional process, a contention that is supported by the small number of articles in the literature that deal with it. However, in the current economic environment and in the light of the Training Guarantee, training personnel are going to be faced with hard economic decisions about the viability and value of the programs they offer. They are going to need evidence of the quality of their programs in order to make such decisions and to influence the decisions of organisational management. This evidence can only come about through evaluation. Thus, evaluation must be given a high priority and must be fully incorporated into the instructional development process. In order for this to happen the meaning of evaluation must be clarified and its purpose(s) must be clearly identified.

In this paper it is my intention to report on a limited review of the literature and to explore some of the notions of evaluation that are presented.

Evaluation of the Literature

In 1989, Marguerite Foxon reported on some of the findings of a literature search that she undertook dealing with evaluation. She reviewed the Australian, American and British journals for a 15 year period to 1986 and

expressed surprise at the relatively small number of articles (80) on the subject. Six of these articles were in the Australian literature, five of which were in *Training and Development in Australia*. Of the remaining 74 articles, 34% were published in the *Training and Development Journal*. The total contribution to the literature on evaluation in these two journals averaged two articles per year for the 15 year period.

I have confined my own search to these two sources for the period 1987 to 1990 (May). In the three and a half year period I identified a total of 13 articles dealing with evaluation representing an average of three articles per year. Although an improvement, this still reflects the regard that the profession has for evaluation, it remains a low priority. Only two of the 13 articles located are in the Australian journal.

In a survey of *Training and Development Journal* readers 30% of respondents identified "evaluation of training as the most difficult part of the job" (Galagan, 1983, and Del Faizo, 1984 cited in Foxon, 1989, p 90). This finding is not surprising since evaluation is poorly defined having different meanings for different people in many different contexts. There is a strong dependence in the profession on the determination of trainee reactions to programs as a major means of evaluation. Foxon (1989, p 91) makes the point that many trainers see the "development and delivery of training as their primary concern, and evaluation something of an afterthought." She suggests that the reliance on post-course reactions results from an inability to deal with quantitative measurement techniques and a lack of finances, time and expertise in comprehensive evaluation. Further, she suggests that training practitioners are confused by the term and do not understand what its "essential features" are nor what "purpose it should serve".

In my own review, only one author attempts to define evaluation. Wigley (1988, p 21) defines it as "a data reduction process that involves the collection of large amounts of data which are analysed and synthesised into an overall judgement of worth or merit". The implication here is that the judgement of worth can be supported by the data. In her review, Foxon (1989, pp 91-92) found similar definitions referring to judgements of "value or worth". What is not clear in any of the definitions offered is what is entailed in the criteria of worth.

It has been suggested that a major problem in arriving at a definition of evaluation is confusion with related terms such as measurement, assessment and validation (Foxon, 1989 p 92). This suggestion is a reasonable one if the literature is any indication of the way the training population perceives evaluation. Only four of the articles (Wigley, 1988; Brinkerhoff, 1988; Birnbrauer, 1987; and Bushnell, 1990) in the current study refer to a comprehensive approach to evaluation, involving the collection of data from the beginning of program design through to

program completion and post program evaluation techniques utilising a variety of data collection methods. Of the remaining articles five (O'Donnell, 1988; Harbour, 1988; Erkut and Fields, 1987; Weatherby and Gorosh, 1989; and Newstrom, 1987) use evaluation when referring to post course trainee reaction assessment and two (Alliger and Horowitz, 1989; and Hewitt, 1989) use it to refer to performance and/or knowledge testing. Two articles (Bowsher, 1990; and Lombardo, 1989) highlight the cost and benefits of training within the organisation but fail to refer to cost/benefit analysis in the context of course evaluation!

Evaluation in Practice

Wigley (1988) describes a "production approach" to training in which evaluation activities are seen as being isolated from the training itself. In this approach evaluation is focused on statistics that describe the number of training days per year, the number of courses per year and the number of trainees attending each course among other things. Whilst these statistics are useful in providing data about how popular the programs offered by the training department are, they have little affect in showing whether the training department is fulfilling any useful purpose for the organisation - unless "bums on seats" is seen as a useful purpose.

Foxon (1989, p 90) reported the results of a number of surveys involving the HRD professionals in the United States, Britain and Australia in which it was found that the major form of evaluation, in many cases the only form, was end-of-course trainee reactions and that the data from these evaluations were seldom used. The current search revealed that a number of authors (Harbour, 1988; Erkut and Fields, 1987; Weatherby and Gorosh, 1989) continue to rely heavily on the trainee reaction form of evaluation. The trainee reaction questionnaire, often referred to as the "smile sheet," is relatively easy to construct and administer when compared to other forms of evaluation, and, if kept simple enough, it is easy to analyse and report on the findings. The data obtained can be useful in determining which trainers, training methods, aids and resources are popular and therefore likely to affect trainee motivation and participation. However, its usefulness is limited in that the data obtained is subjective and gives little or no information about whether the training program contributed to or achieved the goals and objectives of the organisation, the training department, the particular program or the individual trainees. Only one article (Newstrom, 1987) has referred to a statistical analysis to determine the reliability of evaluation - in this case applied to the trainee reaction form of evaluation.

The impression that one is left with is that, for the most part, evaluation is seen as an activity that occurs at the completion of a training program, and that the practice of evaluation is confined to a limited number of activities.

Those articles that describe a comprehensive approach to evaluation generally speak of the principle rather than the practice. In the current study there are no reports of a training program in which various evaluative data has been collected from the outset and continued through to follow-up assessment with collation and analysis of the data and compilation of a report with recommendations as the final outcome. This perhaps sounds like an idealistic expectation, and indeed it is, but I believe that it is not only possible but essential if we are to be responsible and accountable to our clients for effective training programs. This, I believe, is the ultimate goal of evaluation - to demonstrate the accountability of the training department/consultant to the client for effective training programs.

Purpose of Evaluation

If this is the ultimate goal, then what is/are the purpose(s) of evaluation? For it is only in determining the purpose(s) that we can develop a working definition of the term. Foxon (1989, p 93) reports that more that 20% of the articles reviewed "neither describe nor imply a purpose for the evaluation." Of those that did, 15% saw the purpose as "justifying the training department's existence and providing evidence of cost benefit to the organisation." (Foxon, 1989, p 93). Interestingly, only 2% of the articles in Foxon's study considered assessment of trainee reactions to be the purpose of evaluation. Given this evidence one can not help but wonder at the proliferation of this form of evaluation.

In the current review three of the articles provide a clear statement of purpose for training. Hewitt (1989, p 23) sees the purpose of evaluation as providing data demonstrating the program's effectiveness on targeted behaviour. Wigley (1988, p 21) has a broader view of the purpose to improve the program and facilitate informed decision making. Far and away the most comprehensive view of the purpose of evaluation is given by Bushnell (1990, p 41) who identifies four purposes: "to determine whether training programs are achieving the right purposes... to detect the types of changes they [the trainers] should make to improve course design, content, and delivery... [to tell the trainer] whether students actually acquire the needed knowledge and skills."; the ultimate purpose being to "balance the cost and results of training." All of Bushnell's purposes involve end stage evaluation methods, that is the data that will fulfil them can only be obtained at the completion of the program.

Another four of the articles in the current study (Weatherby and Gorosh, 1989; Erkut and Fields, 1987; Alliger and Horowitz, 1989; and Brinkerhoff, 1988) imply a purpose to the evaluation of training. These implications range from very vague phrases like "gauging the effectiveness of training materials and programs" (Erkut and Fields, 1987, p 74) to very specific and

comprehensive descriptions such as that given by Brinkerhoff (1988, p 67) when he provides "six stages of HRD program development and operation":

1. to determine that an identified problem represents a training need and to determine what the real goals are.
2. to determine the most appropriate training strategy
3. to determine if the chosen strategy is successfully implemented
4. to determine if learning occurred and to what extent
5. to determine usage outcomes (at individual level)
6. to determine impacts and worth (at organisational level)

Implicit in these six stages are seven corresponding purposes. These seven purposes require the utilisation of evaluation methods to obtain data continually from the beginning of the instructional process - needs analysis - to the completion of the process - performance assessment. In comparing the six-stage model to the widely referenced Kirkpatrick model (cited in Birkenhoff, 1988, p 66; Poulet & Moult, 1987, p 64; and Newstrom, 1987, p 57), it is clear that the latter falls short of an ideal model because it is entirely outcome-oriented whereas the six stage model is integrated to include instructional activities in the planning, design and implementation stages of the instructional process.

On first looking at the purposes implicit in Brinkerhoff's model, one can easily conclude that the two purposes identified at (1) above relate directly to needs assessment and, indeed, could even be interpreted to be the same as needs assessment. Closer examination of the model reveals this to be true to a certain extent, but it goes further to incorporate appraisal of the needs assessment tools and methods to ensure that they are truly exposing the real nature of the problem and the correct solution options. Similarly, with the purpose given at (2) above which could rightly be said to be the property of instructional planning or instructional design, yet in choosing the appropriate strategy one or more training strategies need to be considered, trialed and/ or analysed to ensure that the correct one is identified. Thus, an evaluative procedure must be employed. The remaining purposes equate readily to frequently stated or implied purposes, namely to determine trainee/trainer reactions to the training methods and materials utilised in a program, to assess trainee acquisition of knowledge and attitudes, to assess trainee performance, and to determine if organisational goals are met.

It is my view that evaluation should fulfil all these purposes with the overriding aim being to influence decisions about the need for future programs, the need to modify future programs, the need to modify instructional tools at all stages of the instructional process and the need to provide cost/benefit data regarding the programs offered by the training

department/ consultant. It seems clear that, if decisions are to be influenced, then the culmination of any set of evaluation activities is the compilation of a report containing recommendations

Evaluation in the context of training

Thus, any definition of evaluation in the context of training and development should include a number of elements: what it is, what it involves and what it leads to. Evaluation is an analytical process. Evaluation involves the collection of subjective and objective data from a number of sources using a variety of techniques about a training program and the reduction of such data. Evaluation leads to the synthesis of the data into a report containing a summary of results and recommendations, with validated rationales, about the program being evaluated. This last element - the synthesis of a report - is one which has not been addressed in the literature reviewed. Yet, if the overriding aim is to influence decisions, then the omission of a report is a grave omission indeed. A well written report presents arguments clearly and concisely so that decision makers have the evidence before them for consideration of the trainers recommendations. There is less reliance on verbal discussion and what is often construed as the trainers' attempt to justify their continued existence within the organisation.

Further, a written report provides a permanent record of the evaluation outcomes and the reasons for modifications to a program. If the reporting is not formalised then the reasons for modification of a program can tend to become more and more subjectively interpreted over time. The result being that the program can eventually revert to its original form rather than being progressively improved and in line with organisational goals.

Conclusion

In summary, evaluation can be defined as an analytical process involving the collection and reduction of data of all (or some) phases of the instructional process and culminating in the synthesis of a report containing recommendations about the instructional program being evaluated. The overall aim of evaluation is to influence decisions about the need for the program in the future; the need for modifications to the program; and the need to provide cost/benefit data about the program. Therefore, evaluation can be said to have at least seven purposes:

1. to validate needs assessment tools and methods.
2. to confirm or revise solution options.
3. to confirm or revise training strategies.
4. to determine trainee/trainer reactions.

5. to assess trainee acquisition of knowledge and attitudes.
6. to assess trainee performance.
7. to determine if organisational goals are met.

In light of the Training Guarantee, which requires employers to contribute a minimum of one percent of the payroll to training (in 1990-91), training practitioners need to become more cost conscious than they have ever been in the past. In my experience, trainers are often heard to complain that in hard economic times training is one of the first things to suffer budget cuts and staff stand downs. The Training Guarantee will ensure that organisations give careful consideration to the contribution that training makes to the organisation before taking such action in the future. However, in order for that to happen, training practitioners are going to be called upon to provide hard evidence of the value of the training programs they offer.

References

Alliger, G. M. & Horowitz, H. M. (1989). IBM takes the guessing out of testing. *Training and Development Journal*, April, 69-73.

Birnbrauer, H. (1987). Evaluation techniques that work. *Training and Development Journal*, July, 53-55.

Bowsher, J. (1990). Making the call on the COE. *Training and Development Journal*, May, 65-66.

Brinkerhoff, R. O. (1988). An integrated evaluation model for HRD. *Training and Development Journal*, February, 66-68.

Bumpass, S. & 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

Bushnell, D. S. (1990). Input, process, output: A model for evaluating training. *Training and Development Journal*, March, 41-43.

Erkut, S. & Fields, J. P. (1987). Focus groups to the rescue. *Training and Development Journal*, October, 74-76.

Foxon, M. (1989). Evaluation of training and development programs: A review of the literature. *Australian Journal of Educational Technology*. 5(1), 89-104.

Hewitt, B. (1989). Evaluation a personal perspective. *Training and Development in Australia*, 16(3), 23-24.

Lombardo, C. A. (1989). Do the benefits of training justify the costs? *Training and Development Journal*, December, 60-64.

Newstrom, J. W. (1987). Confronting anomalies in evaluation. *Training and Development Journal*, July, 56-60.

O'Donnell, J. M. (1988). Focus groups: A habit-forming evaluation technique. *Training and Development Journal*, July, 71-73.

Poulet, R. & Moult, G. (1987). Putting values into evaluation. *Training and Development Journal*, July, 62-66.

Weatherby, N. L. & Gorosh, M. E. (1989). Rapid response with
    spreadsheets. *Training and Development Journal*, September, 75-79.
Wigley, J. (1988). Evaluating training: Critical issues. *Training and
    Development in Australia*, 15(3), 21-24.

Author: Jane Marsden is a Human Resource Development (HRD)
professional with a range of expertise in policy development, personnel
management, education, training and development, and professional
writing. She has fourteen years experience in training needs analysis, job
analysis and skills audits, concept analysis, instructional systems design,
formal and informal presentation strategies, and assessment and evaluation
techniques. For the last three years, Jane has been involved in training
needs analysis, skills audits, instructional systems design and training
evaluation in the corporate sector.

Please cite as: Marsden, J. (1991). Evaluation: Towards a definition and
statement of purpose. *Australian Journal of Educational Technology*, 7(1), 31-
38. http://www.ascilite.org.au/ajet/ajet7/marsden.html