PUBLIC ADMINISTRATION AND DEVELOPMENT, Vol. 4, 275–282 (1984)

Evaluation of training. A review of the literature

A. R. HOYLE

Canberra College of Advanced Education

SUMMARY

The literature on evaluation needs to be classified into education and training. The latter reveals many difficulties as regards evaluation. Scientific and quantitative methods are not popular. Evaluation appears to be undertaken reluctantly and with the simplest methods. Behavioural objects are rarely even set by trainers. Progress in the techniques of evaluation has been slow, though a good deal of research has been done. The literature is small but growing.

Few topics are more often raised in any discussion of the training of administrators than evaluation, although, at the same time, the practice of evaluation is limited to those few trainers who are concerned, as professionals, to have an appreciation of the value of their work which owes little to prejudice, hunch or gossip. That this is not an extreme statement became apparent to the writer when he was recently engaged in preparing an annotated bibliography on the evaluation of training.

A first look at the catalogued material on evaluation suggests that this is a field in which a great deal of work has been carried out by researchers, educationalists and trainers and the frequent references to evaluation in training journals appear to indicate that the concept and the practice are matters for lively debate among training professionals. A further look, however, reveals that much of the material is lightweight and ephemeral and that the range of important contributions to both the philosophy and methodology of evaluation is very limited. This is surprising as the first important works date from the 1950s.

In reviewing the literature on the evaluation of training it was necessary, in the first place, to differentiate between educational evaluation and training evaluation. The former, which has a large and important literature, deals almost entirely with the evaluation of lengthy educational programmes in schools and, to a much smaller extent, tertiary institutions. It deals with situations where scientific method can be employed and where objective measurements of performance can be made. Training evaluation, on the other hand, is usually concerned with the assessment of brief programmes for adults, in non-school situations, on subjects which are not susceptible to objective measurement.

These elements—brief courses and subjects which cannot be measured quantitively—present very real difficulties for even the most professional of training staff. When they are associated with the usual lack of interest of senior management in training it is understandable that evaluation of training has not thrived.

Of the training evaluation literature in English the basic works continue to be those of D. L. Kirkpatrick, P. Hesseling and A. C. Hamblin. Older useful texts exist, particularly that of Kohn (1969), but for training professionals and academic researchers alike, these three authors appear to provide the best starting point.

0271–2075/84/030275–08$01.00
© 1984 by John Wiley & Sons, Ltd.

The fundamental text on training evaluation is that of Hesseling (1966) who aimed his work at social scientists, policy makers and training officers. Unlike most of the periodical literature which attempts to give a cookery book recipe, Hesseling deals with the theoretical issues and overall strategy necessary for evaluation attempts while facing the difficulties encountered by the trainer when he is committed to carrying out a proper evaluation.

The work of Kirkpatrick (1975) is a useful collection of articles which originally appeared in the journal of the American Society for Training and Development and which are directed to the practitioner of supervisory and management training. Kirkpatrick is a very useful adjunct to Hesseling as he uses the theory to provide a practical model of how to evaluate a programme. The work of Hamblin (1974), although designed for the training specialist, has an extra dimension. Hamblin has produced a book which is useful, not only for the specialist, but also for those who are responsible for the management of training. All too often these managers have only the most elementary understanding of the training process—a situation which may be of short term advantage to the trainer but can be ultimately disastrous. Hamblin deals with the cost-benefit approach to training, which is increasingly popular in a period of strained resources, and is realistic enough to recognize that what is objectively right may not be politically expedient. He also points out correctly that evaluation is not a goal in itself—to determine if desired changes did occur—but is rather to determine what should happen next. The book is one of the most important contributions to evaluation at both the philosophical and practical levels. The question of using scientific method in evaluation is one that has exercised a number of writers on evaluation. Most practitioners, writing in periodicals, have no doubt that, in the real world of training—especially management training—the difficulties are insuperable. Hamblin believes that, as evaluation must be open-ended, it is hardly ever possible to set up a scientifically controlled evaluation experiment.

There is, however, some literature which deals usefully with the problems and techniques of experimental design in evaluation. Particularly useful for trainers who see evaluation as one of the major aspects of their job is the extended article by Kane (1976). For those training professionals with some knowledge of quantitative techniques, Kane has produced a very useful short work on the classic experimental approach, the generally more useful quasi-experimental design, the interrupted time series design and similar techniques. He also deals with non-experimental designs but the value of the article lies in the competent approach to experimental design.

For those interested in an experimental approach but lacking the experience or the necessary confidence to produce their own experimental design from scratch, the work of Morris and Fitz-Gibbon (1978) is particularly useful. They deal successfully with conceptualizing evaluation and how to play the role of formative evaluation and give a step-by-step guide for conducting a small experiment. For training professionals with a real interest in evaluation, the remaining books in the set are also of value as they deal with areas often glossed over quickly—how to deal with goals and objectives and how to present an evaluation report.

At a more sophisticated level is the extensive work of Tracey (1971) which has exhaustive chapters on evaluation dealing with the construction of evaluation instruments and criterion measures, administering and analysing these measures and following up graduates. The provision of ample diagrams and sample questionnaires should assist the reader.

2

The general unpopularity of scientific method and the quantitative approach to evaluation is revealed in the general dearth of articles on them in recent periodical literature. Where this literature does deal with these approaches it is either directed towards the researcher rather than the trainer (Williams, 1976) or is too brief to be of much assistance (Wallace and Weitzel, 1972) or, more usually, damns it with faint praise.

The overwhelming impression from the literature is that most trainers are reluctant to practice evaluation and when they do attempt it they want the simplest possible method. The fact that training is usually thought of as an investment for the future (Guyot, 1977) often removes from the trainer any managerial pressure to prove the cost-effectiveness of training. Again, training personnel are hesitant to undertake evaluation efforts that might appear to invalidate their work (Bunker and Cohen, 1978). The result is that such evaluation as is usually undertaken is unscientific and mainly to get feedback from course participants on their own performance as trainers (Brandenburg, 1972).

All serious writers on evaluation recognize that trainers need immediate feedback, but they are also aware that the real purpose of higher level training is to produce long lasting changes in values, initially in individuals and then in the organization (Cowell, 1972). What is also recognized is the very real difficulty in developing criteria by which to measure attitudinal change (Donald, 1970). Fundamental to developing these criteria are the tasks of making a needs assessment, establishing clear objectives and analysing the organization to determine factors which might make useless the best conceived training, e.g. organizational conflicts (Goldstein, 1979). However, even where these tasks are carried out, the key difficulty of determining proper criteria remains. In skills training this may be comparatively easy (Bunker and Cohen, 1977) but in management training the problem is one which has yet to be properly overcome. There are, as yet, no final solutions but the use of behavioural objectives will assist in overcoming the criterion problem in programme evaluation (Gordon, 1972).

Many trainers are deterred from attempting to set behavioural objectives because they feel they lack an education in psychology. This should not, in practice, be a hindrance because psychologists have had a continuing interest in evaluation and have made a considerable contribution to the literature. Works by authors such as Hartley (1973) and Schwarz and Krug (1972) go a long way towards bridging the gap between psychologists and trainers and permit the latter to develop more sophisticated measures.

Whether training professionals will use the work of psychologists to improve evaluation techniques remains to be seen but the auspices do not seem favourable. It appears probable that most trainers will search for a relatively simple model to use. There is no shortage of these with the most popular being the CIRO (context, input, reaction, outcome) of Warr, Bird and Rackham (1970) and the four stage (reaction, learning, behaviour and results) model of Kirkpatrick (1976).

The selection of a good model is only one of the early steps in evaluating training. Almost all models require the development of questionnaires and the involvement of management in the process. The former is not an easy task and should be unique to each organization—or even each course—but it is one of the areas in evaluation literature which is heavily supplied. A very considerable proportion of the books and periodical material contains advice on how to contruct questionnaires and/or gives examples. It is impossible to say which are the best but, for the public sector,

particularly useful examples are to be found in the publication by the now defunct Civil Service Department of the United Kingdom (1977) covering most fields of training, both administrative and technical, in the major ministries in the British government. For trainers concerned with evaluating the results of the well-known Kepner–Tregoe training programme in management problem analysis and decision making, very useful questionnaires are available in the work of Harman and Glickman (1965). For private sector training, Kirkpatrick (1975) is useful.

There are some occasions when it is desirable to give tests, rather than administer a questionnaire. Much test construction is best left to professional psychologists but evaluators without recourse to such persons can profitably use the work of Denova (1979), who emphasizes practical application rather than theory and offers the training officer the essential procedures for selecting test items, constructing different types of tests, administering the tests and evaluating the results. The particularly strong point of Denova's work is that it is outlined from the viewpoint of the trainer. For those with some background in psychometrics, much of the same area is covered by Dubois, Mayo and Douglas (1971).

The use of matrices has become popular in recent years and for those interested in using them in the evaluation of training a number of short articles are available. Brethower and Rummlen (1979) use a systems model to produce an evaluation matrix which will be of use to most trainers. An even shorter article by Brown (1976) presents an evaluation form in the shape of a matrix with techniques on the left and stages of adoption on the right. This is sent to participants some time after the conclusion of the programme. Kayloff (1971) points out that the matrix is useful not only for existing courses but also for the establishment of training levels for new courses.

Post-course evaluation has a number of difficulties. These include the difficulty of having a sufficiently large number of course participants respond, getting worthwhile reports of changed attitudes of behaviour from supervisors and taking account of the large number of possible contaminants. The time of post-course evaluation can be critical as Hand, Richards and Slocum (1973) found in their study. Ninety days after the programme little change was noticed in the attitudes and behaviour of participants. Eighteen months later significant positive changes existed in the experimental group whereas negative changes were observed in the control group. A highly practical approach to the subject is to be found in two short pieces by Coffman (1979, 1980). Coffman has produced a practical procedure by which the supervisors and managers, who are perceived as the real clients of the training system, are formally brought together with the training staff in two sessions to develop opinions on the strengths and weaknesses of a programme. The article is realistic in recognizing that this technique can be threatening to the trainer(s) involved.

Value analysis is a well-known tool in the world of management consulting but it is rarely applied to aspects of the personnel function such as training. This lacuna has been partly removed by Talbot and Ellis (1969) who attempt to apply the technique in evaluating training. Value analysis believes that what is being done can be improved by recognizing the problem factors and challenging assumptions and that average people working together with a common aim will achieve better than average results. The authors point out that neutral questioning will stimulate new thinking and 15 questions are listed that anyone interested in value analysis should ask about training.

It is well known that, in some societies, the messenger who brings bad news to the top figures in government is likely to be summarily executed. Much of the same situation applies when a trainer, having evaluated some training programmes and found them wanting, has to bring the bad news to management. To survive requires not only courage but also preparation. Morris and Fitz-Gibbon (1978) give details on how to present an evaluation report but it is mainly aimed at educational administrators. A succinct statement of how to handle this unpleasant task is contained in the article 'They shoot messengers, don't they? Breaking the news to management' (*Training*, 1980). The eight-step procedure set out—know your audience; communicate as you go; plan the best media—time style; pilot test the presentation; deliver the bad news first; leave nothing to chance; expect to be rejected; evaluate—is one which could, with advantage, be used by every serious evaluator of training.

Organizational development is one of the least precise terms in the training officer's lexicon but, in its many forms, it appears that it is here to stay for a long time yet. Because organizational development is more than simply training, the usual evaluation techniques may not be sufficient. What can be done is tackled by Armenakis and Smith (1978) who propose the use of an abbreviated time series design and demonstrate that it is more effective than the usual pretest-post-test design. For organizational development evaluators who do not necessarily accept ATSD as the best method Armenakis, Field and Mosley (1975) set out the results of a survey in the mid-1970s showing how American practitioners evaluated programmes and detailing the most common evaluation practices, the types of criteria used and guidelines for practice.

The high water mark for sensitivity training has passed as increasing doubts are shown as to its value but, for trainers who still believe in and practice the technique, the reports of two evaluation efforts will be of benefit. Bare and Mitchell (1972) report positive results and their article is useful for its methodology. However, a more generally useful study is that by Smith (1975) which deals with issues such as global measures of self concept, perception of others, perception of trainee behaviour and organizational behaviour after sensitivity training.

Tertiary educational institutions are notorious in the educational world for their reluctance to embark upon any form of evaluation of their teaching activities. However, economic factors and student stridency are gradually altering this situation. The first large-scale effort in the field was made by Andrews (1966) nearly 20 years ago. More up to date approaches to evaluation are common today but Andrews' work, in great detail and with a 16 page questionnaire, is still useful for any tertiary educator or researcher who seeks to evaluate a very large scale prrogramme. The more recent attempt by Burgoyne (1973) shows, much more clearly than Andrews, the intractable nature of much of the problem when faculty members are unable to enunciate clearly what the students are to learn and the great difficulty in developing a performance-related index. Because they were less ambitious, Ference and Ritti (1970) were able to obtain better results with a sample of recent MBA graduates, and their work appears to have a wider application. However, much the best work in this field is that of Dressel (1976) which presents a complete analysis of all phases of evaluation in academic institutions and shows how evaluation can pin-point future needs and appraise the university or college of unrecognized potentials as well as provide insight regarding curriculum and faculty.

Many of the books and journals reviewed contained case studies of evaluation

attempts and, to the trainer, these are often more valuable than lengthy monographs. Some which should be helpful are those for performance appraisal interviewing training (Allison, 1977), a training programme in the probation service (Ashton and Gibbon, 1974), a series of marketing courses (Cowell, 1975), management by objectives training (Fay and Beach, 1974), management skills in a computer firm (Honey and Morris, 1973) and administrative training in a developing country (Taylor, 1978).

Progress in the techniques of the evaluation of training has been slow, but this does not mean that there have not been considerable efforts made in research. Burgoyne (1973), Clement (1982) and Huczynski and Lewis (1980) are but a few of the many persons who are continuing to try to grapple with the all too obvious difficulties of evaluation.

Compared with other areas of concern to training professionals, the literature on evaluation is small but it is widely spread and there is room for bibliographies to be prepared every few years. For the early period the work of Kohn (1969) is helpful, whereas Kirkpatrick (1976) provides a useful list of mainly American material from before 1972. Much the same ground is covered by Inayatullah (1975). For recent attempts to provide bibliographical material trainers might look at the work of Obern (1982) as well as that prepared by the present writer (Hoyle, 1983).

Evaluation can be a costly exercise, not only in financial terms but also in human terms where it has the potential to threaten the self-image of trainers and destroy the often blind faith of management in the development possibilities of the process. None the less it is vital if training is to succeed in transforming or developing organizations and individuals to meet the increasingly complex problems facing management in the next few decades. That, as Easterby-Smith (1981) has pointed out, there are no longer any clear guidelines as to what constitutes a 'good' evaluation, only serves to make the task of training evaluation more challenging and more worthy of professional interest.

REFERENCES

Allison, C. W. (1977). 'Training in performance appraisal interviewing—an evaluation study', *Journal of Management Studies*, **XIV**, 179–191.

Andrews, K. R. (1966). *The Effectiveness of University Management Development Programs*, Harvard, Cambridge, Mass.

Armenakis, A. A., Field, H. S. and Mosley, D. C. (1975). 'Evaluation guidelines for the organizational development practitioner', *Personnel Journal*, **54**, (2), 99–103.

Armenakis, A. A. and Smith, L. (1978). 'A practical alternative to comparison group designs in organizational development evaluations', *Academy of Management Journal*, **21**, (3), 499–507.

Ashton, D. and Gibbon, B. (1974). 'Evaluating a training programme in the probation service', *European Training*, **3**, (1), 62–75.

Bare, C. E. and Mitchell, R. R. (1972). 'Experimental evaluation of sensitivity training', *Journal of Applied Behavioural Science*, **8**, (3).

Brandenburg, D. C. (1982). 'Training evaluation—what's the current status?', *Training and Development Journal*, **35**, (7), 14–20.

Brethower, K. S. and Rummler, G. A. (1979). 'Evaluating training', *Training and Development Journal*, **33**, (5), 14–23.

Brown, G. R. (1976). 'A model for post course evaluation', *Canadian Training Methods*, **8**, (5), 22–24.

6 1

Bunker, K. A. and Cohen, S. L. (1977). 'The rigors of training evaluation—a discussion and field demonstration', *Personnel Psychology*, **30**, (4), 525–542.

Bunker, K. A. and Cohen, S. L. (1978). 'Evaluating organizational training efforts', *Training and Development Journal*, **32**, (8), 4–11.

Burgoyne, J. G. (1973). 'An action research experiment in the evaluation of a management development course', *Journal of Management Studies*, **X**, 8–14.

Burgoyne, J. G. (1973). 'A new approach to evaluating management development programmes—some exploratory research', *Personnel Review*, **2**, (4), 40–45.

Civil Service Department (1977). *Evaluation of Training in the Civil Service—A Survey*, London.

Clement, R. W. (1982). 'Testing the hierarchy theory of training evaluation—an expanded role of trainee reactions', *Public Personnel Management*, **11**, (2), 176–184.

Coffman, L. (1979). 'An easy way to effectively evaluate program results', *Training and Development Journal*, **33**, (8), 28–33.

Coffman, L. (1980). 'Successful training program evaluation', *Training and Development Journal*, **34**, (10), 84–87.

Cowell, D. W. (1972). 'Evaluating the effectiveness of management courses', *European Training*, **1**, (1), 55–64.

Cowell, D. (1975). *Making Management Education Effective*, Management Decision Monograph, Bradford, 13 (2).

Denova, C. C. (1979). *Test Construction for Training Evaluation*, van Nostrand Reinhold, New York.

Donald, A. (1970). 'Evaluating middle management and supervisory training', *Management Education and Development*, **1**, (2).

Dressel, P. L. (1976). *Handbook of Academic Evaluation*, Jossey-Bass, London.

Dubois, P. H., Mayo, G. and Douglas, G. (eds) (1971). *Research Strategies for Evaluating Training*, American Educational Research Association, Washington, D.C.

Easterby-Smith, M. (1981). 'The evaluation of management education and development—an overview', *Personnel Review*, **10**, (2), 28–36.

Fay, P. P. and Beach, D. N. (1974). 'Management by objectives evaluated', *Personnel Journal*, **53**, (10), 767–769.

Ference, T. P. and Ritti, R. R. (1970). 'Evaluating the functions of management education', *Personnel Journal*, **49**, (8), 642–650.

Goldstein, I. L. (1979). 'The pursuit of internal and external validity in the evaluation of training programs', *Public Personnel Management*, **8**, (6), 416–429.

Gordon, M. E. (1972). 'Three ways to effectively evaluate personnel programs', *Personnel Journal*, **51**, (7), 498–504.

Guyot, J. F. (1977). 'How do we know which training is good medicine for managers?', *Public Administration Review*, **37**, (6), 698–705.

Hamblin, A. C. (1974). *Evaluation and Control of Training*, McGraw Hill, London.

Hand, H. H., Richards, M. D. and Slocum, K. W. (1973). 'Organizational climate and the effectiveness of a human relations training program', *Academy of Management Journal*, **16**, (2), 185–195.

Harman, E. L. and Glickman, A. A. (1977). *Evaluation of Training in the Civil Service—A Survey*, London.

Hartley, J. (1973). 'Evaluating the effectiveness of training', in Davies, I. K. (ed.), *The Organization of Training*, McGraw Hill, Maidenhead.

Hesseling, P. (1966). *Strategy of Evaluation Research*, Van Goocum, Assen.

Honey, P. and Morris, M. (1973). 'Practical experiences in evaluating management training', *Personnel Management*, January 1973.

Hoyle, A. R. (1983). *An Annotated Bibliography on the Evaluation of Training*, Commonwealth Secretariat, London.

Huczynski, A. A. and Lewis, J. W. (1980). 'An empirical study into the learning transfer process in management training', *Journal of Management Studies*, **XVII**, 227–240.

Inayatullah, M. (ed.) (1975). *Management Training for Development—The Asian Experience*, Asian Centre for Development Administration, Kuala Lumpur.

Kane, J. A. (1976). 'The evaluation of organisational training programmes', *Journal of European Training*, **5**, (6).

Kayloff, A. G. (1971). 'A method of evaluating the effectiveness of technical training', *Training and Development Journal*, **16**, (2), 185-195.

Kirkpatrick, D. L. (1975). *Evaluating Training Programs*, American Society for Training and Development, Madison.

Kirkpatrick, D. L. (1976). 'Evaluation of training', in Craig, R. L. (ed.), *Training and Development Handbook*, McGraw Hill, New York.

Kohn, V. (comp.) (1969). *A Selected Bibliography on the Evaluation of Management Training and Development Programs*, American Foundation for Management Research.

Morris, L. and Fitz-Gibbon, C. (1978). *Evaluator's Handbook*, Sage, Beverley Hills.

Obern, A. G. (1982). *Bibliography on Staff Training and Development in Developing Countries*, United Nations.

Schwarz, P. A. and Krug, R. E. (1972). *Ability Testing in Developing Countries—A Handbook on Principles and Techniques*, Praeger, London.

Smith, P. B. (1975). 'Controlled studies of the outcome of sensitivity training', *Psychological Bulletin*, **82**, (4), 597-622.

Talbot, J. R. and Ellis, C. D. (1969). *Analysis and Costing of Company Training*, Gower, London.

Taylor, L. (1978). 'Evaluation of administrative training programmes in the Southern Sudan 1975-77', *Journal of Administration Overseas*, **XVII**, (3).

Tracey, W. R. (1971). *Designing Training and Development Systems*, American Management Association, New York.

Training (1980). 'They shoot messengers, don't they? Breaking the news to management', *Training*, 99-102.

Wallace, M. J. and Weitzel, W. (1972). 'The feasibility of experimental evaluation of management training programs', *Academy of Management Proceedings*, August 1982, 320-323.

Warr, P., Bird, M. and Rackham, N. (1970). *Evaluation of Management Training*, Gower, London.

Williams, G. L. (1976). 'The validity of methods of evaluation learning', *Journal of European Training*, **5**, (1), 12-20.