Evaluation of training and development programs:
A review of the literature

Marguerite Foxon
Coopers & Lybrand

This paper outlines some of the findings of a research project on evaluation, which involved a review of the Training and Development (journal) literature for the period 1970-1986. An annotated bibliography was produced by the author as part of the project.

As part of a larger research project on evaluation, I reviewed the relevant Australian, British and American journals for the period 1970-1986. My intention was to identify themes or trends in the evaluation of T&D programs, and ultimately to extract from the literature some practical guidelines, techniques or models useful to T&D/HRD professionals, particularly in the area of management development and Human Resource programs.

I was initially surprised by the relatively small number of articles on the subject of evaluation. A total of six articles in Australian journals was found (five by Australian practitioners), and the Australian National Library has no record of any publication dealing with HRD evaluation for the period 1980-86. In British and American journals, some eighty articles were located, the most prolific period being 1982-84.

The other impression one gains is of the uneven quality of this material. Much of it is rather superficial and general; some on the other hand is so academic in style it would be difficult for many practitioners to understand or apply.

The lack of extensive bibliographies and literature reviews was also a surprise finding. As a result, one of the products of this research project was the development of an annotated bibliography of more than eighty articles. This is included at the end of this article.

In reviewing the literature I undertook a content analysis of the articles. In this article I will relate my findings in relation to the definition of evaluation, the purpose of evaluation as expressed by the author, and the models or techniques proposed.

Current evaluation practice

There is ample evidence that evaluation continues to be one of the most vexing problems facing the training fraternity. Catanello and Kirkpatrick's 1968 survey of 110 industrial organisations evaluating training (Burgoyne and Cooper, 1975, 60) revealed that very few were assessing anything other than trainee reactions.

Looking at similar data and the emphasis in much of the literature, one wonders if there has been much change in 20 years (see, for example, Brown, 1980, 11). Galagan (1983,48) and Del Gaizo (1984, 30) both refer to a survey of *Training and Development Journal* readers in which 30% of the respondents identified evaluation of training as the most difficult part of their job. Easterby-Smith and Tanton (1985, 25) report on their British survey involving HRD practitioners in fifteen organisations. In virtually every case the only form of evaluation being done was end-of-course trainee reactions, and the data so obtained seldom used.

Such findings are similar to my own 1985 survey of a sample of Public Service and private company trainers in Sydney to determine both their attitude to evaluation and what was being carried out by them in practice. All expressed a firm belief in the principle of evaluation, and all administered end of-course forms of varying degrees of complexity to gauge trainee reactions to the instructors, content, and facilities. But 75% admitted that was as far as their evaluation went, mainly because they did not know what else to do. As Easterby-Smith and Tanton (1985) observe, much current practice is only a ritual, and in many cases the evaluation that counts is done before the course is ever given; post-course data merely confirm prior judgements that the training is satisfactory.

In the minds of many practitioners evaluation is viewed as a problem rather than a solution, and an end rather than a means.

Where evaluation of programs is being undertaken it is often a 'seat of the pants' approach and very limited in its scope. Overawed by quantitative measurement techniques, and lacking both the budget and the time as well as the required expertise for comprehensive evaluations, trainers often revert to checking in the only way they know - post-course reactions - to reassure themselves the training is satisfactory.

If the literature is a reflection of general practice, it can be assumed that many practitioners do not understand what the term evaluation

encompasses, what its essential features are, and what purpose it should serve. Consequently the use of training courses far outstrips what is known of their usefulness. When such programs are evaluated, the common sources of data (other than trainee reactions) are numbers of participants, decreased absenteeism at work, high rating of instructors, etc. Many trainers are therefore making judgements on the basis of activities ("employee days of training") and not on relevant results. Many practitioners regard the development and delivery of training courses as their primary concern, and evaluation something of an afterthought.

On the other hand, adopting the premise that no news is good news, many practitioners still avoid the evaluation issue. Preferring to "remain in the dark", and worried that evaluation will only confirm their worst fears (since they have no other alternative to offer management if the current program is shown to be educationally ineffective), they choose to settle for a non-threatening survey of trainee reactions.

Towards a definition

Providing a sound definition is more than a lexicographic exercise; it can clarify and refine concepts, generating a framework within which to develop a pragmatic approach to the subject. Evaluation is no exception, and the apparent confusion in the minds of many as to the purposes and functions of evaluation corresponds to the ignorance or misunderstanding of what is meant by this and related terms such as research, validation, and assessment. A variety of definitions can be found in the literature, many of them stipulative, and the inconsistencies in the use of the terminology has "muddied the waters" of training evaluation a great deal, affecting the success of evaluation efforts (Wittingslow, 1986, 8).

Bramley & Newby (1984a) summarise the diversity of terminology used over the past decade, and offer a most helpful comprehensive table showing the interrelationships between various concepts of evaluation.

Rackham (1974, 454) offers perhaps the most amusing and least academic definition of evaluation, referring to it as a form of training archaeology where one is obsessively digging up the past in a manner unrelated to the future!

In the literature reviewed, where a definition of evaluation is given, the majority of writers tend to view it as the gathering of information in order to make a value judgement about the program, such as necessary changes or the possible cessation of the program. Williams (1976, 12) defines evaluation as the assessment of value or worth. Harper & Bell (1982, 24) refer to the planned collection, collation and analysis of information to enable judgements about value and worth. However, as Williams (1976, 12) observes, value is a rather vague concept, and this has contributed to the different interpretations of the term evaluation.

Some definitions (Goldstein, 1978; Siedman, 1979; Snyder et al, 1980) focus on the determination of program effectiveness. Several definitions emphasise evaluation as a basis on which to determine program improvements (Rackham, 1973; Smith, 1980; Brady, 1983; Morris, 1984; Foxon, 1986; Tyson & Birnbrauer, 1985). The distinction between formative and summative evaluation is not mentioned by most of these writers, but is implicit in their definitions.

Many writers not only differ in their definition of evaluation - they also use evaluation terminology interchangeably and in some cases quite confusedly. Burgoyne & Cooper (1975) for example, use the term evaluation research as synonymous with evaluation. While evaluation and research may appear at first sight to be similar, there are clear differences. Research is aimed at the advancement of scientific knowledge - it is not a given that it should be immediately useful or practical. Control groups, experimental designs, and total objectivity characterise research projects. Unlike research, it is the context of the evaluation which defines the problem, and the evaluator's task is to test generalisations rather than hypotheses. The evaluator may not be able to avoid making value judgements at every stage whereas the researcher must avoid any subjectivity.

Evaluation is also confused by some with the terms measurement and assessment. Evaluation involves description and judgement; measurement and/or assessment provides the data on which to base the evaluation. This confusion of terms is most obvious when considering the use of "evaluation" and "validation". While most American writers do not see validation as separate from evaluation, there are still British writers who appear to draw the distinction (Hawes & Bailey, 1985; Rae, 1985). Rae regards assessment as the measuring of the practical results of the training in the work environment; this, with validation of the training and training method, comprises evaluation). It must therefore be borne in mind that the terms "validation" and "evaluation", often used in HRD literature, do not always mean one and the same thing.

The literature reveals a broad range of definitions and considerable confusion in the use of associated terms, and it would seem that HRD practitioners have yet to give serious consideration to what the term evaluation actually means.

Purpose of evaluation

As well as the lack of agreed-on definition of evaluation, there is an equally broad range of opinions as to the purpose of evaluation. More than 20% of the writers neither describe nor imply a purpose for the evaluation. Where purposes are outlined, they provide some telling insights. For example, 15% see the purpose of evaluation as justifying the

training department's existence and providing evidence of cost benefit to the organisation. The majority of these articles surfaced in the period 1980-83, and clearly reflect the preoccupation of many practitioners with keeping their jobs during an economic downturn and resultant HRD budget cuts!

While a mere 2% consider assessing trainee reactions to be the purpose of evaluation, and 50% see the purpose as judging the quality and worth of the program in order to effect improvements and/or identify the benefits of the training it should be remembered that studies already referred to provide evidence that many trainers are not evaluating beyond the level of trainee reactions. What trainers believe should be done, and what they do in practice seem to differ markedly.

Despite the regular reference in the literature to Kirkpatrick's (1983) four stage model, only a small percentage consider the purpose of evaluation specifically in these terms.

Several writers resist stating a purpose for evaluation, adopting the view that the purpose depends on various factors (Thompson, 1978; Brinkerhoff, 1981; Salinger and Deming, 1982). Evaluation, according to Salinger and Deming (1982,20) is the response to the question "What do you want to know about training?" Nor should its purpose "self-serving" but designed in terms of someone doing something with the information (Brinkerhoff, 1981, 67).

Bramley and Newby (1984a) identify five main purposes of evaluation: feedback (linking learning outcomes to objectives, and providing a form of quality control), control (using evaluation to make links from training to organisational activities, and to consider cost effectiveness), research (determining relationships between learning, training, transfer to the job), intervention (in which the results of the evaluation influence the context in which it is occurring), and power games (manipulating evaluative data for organisational politics).

Burgoyne and Cooper (1975) and Snyder et al. (1980) discuss evaluation in terms of feedback and the resultant issue of control. A decision must be made about how and to whom evaluation feedback will be given. Evaluators are usually conversant with the purpose of the evaluation once they commence it, but this may be because they have a generalised view that the purpose of evaluation is to produce a certain set of data, or because they have determined what purpose the client wishes the evaluation to have. It is possible however that an evaluator may have no specific purpose. The identification of unanticipated side effects of the program may be an important evaluative purpose. Lange (1974) suggests it is often difficult to determine the purpose - there may be several; furthermore, the evaluator may not discover the real purpose until the end of the exercise.

Models and techniques

As with definitions and purposes, there is great variety in the evaluation models and techniques proposed. In some cases it is very difficult to separate the techniques from the 'model' - the writers are actually presenting an evaluation approach using a specific technique rather than a model.

Nearly 50% of the literature discusses case study or anecdotal material in which models and techniques are referred to, but seldom provides detail useful to the reader wishing to implement these. More than 80% of these articles lacked evidence of background research and many failed to offer practical applications.

If the literature reviewed is a reliable guide, Kirkpatrick's four stage model of evaluation is the one most widely known and used by trainees. Perhaps this is because it is one of the few training-specific models, and is also easily understood. Nearly one third of the journal articles from all three countries made reference to his model, and of the eleven writers actually presenting a specific model of evaluation (as opposed to the development of an evaluation strategy), five have drawn inspiration from Kirkpatrick's work.

The objectives-driven model also surfaces in various forms in the literature, although Tyler's name with which it is associated is rarely mentioned. This model of evaluation focuses on the extent to which training objectives have been met, and the common method of evaluating transfer of learning is by control groups. The desirability of setting measurable objectives, following a cost-effective plan to meet them, and evaluating to determine the degree to which they are met is a recurring theme in the HRD literature (Elkins, 1977; Freeman, 1978; Keenan, 1983; Del Gaizo, 1984; Larson, 1985).

The literature is cluttered with suggested evaluation techniques ranging from simple questionnaires to complex statistical procedures. Often the one technique is presented under several different names, such as pre & post testing which is variously referred to as pre-then-post testing (Mezoff, 1981), the 3-Test Approach (Rae, 1983), and Time Series Analysis (Bakken and Bernstein,1982). Similarly, Protocol Analysis (Mmobuosi, 1985) and the journal method of Caliguri (1984) are basically one and the same technique.

Much of the literature reviewed could be regarded as presenting "general techniques" and as such much of it is superficial. For example, in addressing the problem of evaluating the degree to which participants after training use the skills learned back on the job, one reads such statements as "Be sure the instrument [you design] is reliable and delivers

consistent results", and "Measure only what is actually taught and measure all the skills taught". Sadly, such broad brush advice is all too common. Even some of the case study articles gave no insight into their methodology or techniques.

There are three categories of evaluation techniques covered in the literature. The first is the interview. This can be of the trainer, trainee or trainee's superior. It may be pre, during or post training; structured or unstructured. Questionnaires can be used to evaluate at several levels, either qualitatively or quantitatively; as self assessment or objective measures. Finally, there are quantitative and statistical measures including control groups, experimental and quasi-experimental designs. These are far less likely to be used.

There appears to be no mid-point between reasonably subjective measures and scientifically controlled measurement available to the HRD evaluator. Evaluation linked to performance indicators is not common and as Goldstein observes, "The field is in danger of being swamped by questionnaire type items. The failure to develop methodologies for systematic observation of behaviour is a serious fault" (1980, 240).

There is an emerging awareness of the need to perform longitudinal evaluation to evaluate more than the immediate reactions or learning of trainees, although some of the suggested techniques lack objectivity, and data are therefore open to whatever interpretations best suit.

Conclusion

The literature reviewed for the 17 year period to 1986 suggests that there is a widespread under-evaluation of training programs, and that what is being done is of uneven quality.

It is not difficult to sympathise with the practitioners who agree with the principle of evaluation but express concern about the practice of it. The literature contains a confusing array of concepts, terminologies, techniques and models. For instance, more than 80% of the literature reviewed makes no attempt to define or clarify the term evaluation, yet one in four writers propose evaluation models of some description. It was particularly surprising to find this failure to define evaluation in some otherwise quite well researched articles.

Associated with the issue of definition is that of determining the purpose. Many imply their definition when they outline the perceived purpose. If one is unclear as to purpose, the choice of appropriate strategy and methodology will be affected. Nearly one quarter of the articles neither present nor imply any specific purpose for evaluating training. A similar proportion display a superficial understanding of the more complex issues involved, and a paucity of realistic applications.

Woodington (1980) encapsulates these views by highlighting five distinct impressions which can be gained from an overview of training evaluation. Firstly, many practitioners do not perceive the training program as an instructional system, nor do they fully understand what constitutes the evaluation of training. The nature and type of organisation exerts a subtle influence (possibly control?) over the scope and methods of evaluation, and the conduct of evaluation is also dependent on whether internal or external evaluators are used. Finally, he draws attention to the lack of personnel trained in evaluation methodology. The obvious constraint determining the type of evaluation chosen is the availability of resources. This includes time, money, and personnel, as well as the evaluator's own expertise. Possibly the latter is the major constraint. Lange (1974, 23) expresses similar concerns, stating, "Too many bad evaluations are being presented ... evaluation is a good concept based on solid theoretical thinking. But its practice is not well developed".

The definition and purpose of evaluation enable the evaluator to determine what strategy to adopt. Practitioners need to see evaluation in a broader context than merely a set of techniques to be applied. In a systems approach, evaluation is an integral part of the HRD function which in turn is part of the whole organisational process. This integrated approach contrasts with the more popular view of evaluation as something that is "performed" at certain points and on certain groups; the integrated approach means it is difficult to separate evaluation from needs assessment, course design, course presentation, and transfer of training.

It is not within the scope of this article to expand on this further, but the belief that training programs should be continually evaluated from the earliest design phase in order to modify and improve the product goes unrecognised by many trainers. This would account for the popularity of Kirkpatrick's model, which tends to promote retrospective evaluation rather than formative or summative.

Evaluation techniques are not well written up in the literature, and the use of experimental control groups, statistical analysis and similar methods may be concepts which exist only in academic journals according to Bramley and Newby (1984b,18). The need for measurement of training effectiveness is often referred to, but there are few good examples of rigorous evaluation of training programs. One conclusion must be that practitioners do not know how to do much more than basic assessment. Much of what is labelled evaluation is basically an assessment of the actual training activity (Zenger and Hargis, 1982; Morris, 1984). The choice of techniques will depend on some combination of methodological and pragmatic questions, and there is a need to settle for 'sensible' evaluation - one cannot measure the impact of management training on the whole organisation but must make some compromises. Questionnaires, surveys and structured interviews should be carefully designed and field tested to ensure that worthwhile information is received.

The literature review confirms the belief of Morris (1984) that evaluation is regarded by most practitioners as desirable in principle, difficult in practice. It also highlights the lack of well written and documented articles for practitioners to learn from.

Annotated bibliography of evaluation literature

Altschuld, J., Thomas, R, McColskey, W. (1984). An Evaluation Model for Development of Technical Training Programs. *Evaluation News*; 5,4, 3-36.
    Adoption of Cronbach's Lifecycle model to course development and evaluation.
Anon (1986). Focus On Results. Report of 42nd Annual ASTD Conference. *Behavioural Sciences Newsletter*; Book XV, 13, 14 July.
    Reviews material presented on evaluation, referring to Kirkpatrick's model, identifying frequent evaluation pitfalls and barriers to skill transfer.
Bakken, D. & Bernstein, A. (1982). A Systematic Approach To Evaluation. *Training Development Journal*, 36,8, 4-51.
    Evaluation considered in terms of key diffusion makers and what information they want to know. Believes most training has multiple objectives so requires multiple measures. Key is to know what to measure in order to determine how.
Blakeslee, G. S. (1982). Evaluating A Communications Training Program. *Training & Development Journal*, 36,11, 84-89.
    Case study on Communications program using a post course questionnaire after 6 months to evaluate application back on the job.
Bramely, P. & Newby, A. C. (1984a). The Evaluation Of Training Part I: Clarifying The Concept. *Journal of European & Industrial Training*, 8,6, 10-16.
    Summarises the diversity of terminology used in training evaluation; differentiates numerous facets of the training process about which evaluation data may be useful, and provides a framework for linking different evaluation purposes with specific evaluation techniques; discusses the main purposes of evaluation and criteria for selecting an evaluation strategy.
Bramely, P. & Newby, A. C. (1984b). The Evaluation Of Training Part II: The Organisational Context. *Journal of European & Industrial Training*, 8,7,17- 21.
    Examines some organisational factors needing consideration in an evaluation study, including 'politics' and the extent to which evaluations can be truly objective; looks at specialised techniques developed outside the profession by non-trainers.
Brethower, G. & Rummler, G. (1979). Evaluating Training. *Training & Development Journal*, 33,5, 14-22
    Present a framework for considering evaluation alternatives in terms of a general systems view of training. Identifies four levels of evaluation studies. Looks at ability of various designs e.g. control group, reversal, multiple baseline, pre/post measures.
Brinkerhoff, R. (1981). Making Evaluation More Useful. *Training & Development Journal*, 35,12, 66-70.
    Evaluation is the systemic inquiry into training contexts, needs, plans, operations and effects and must be linked to three stages of HR programming: planning, delivering, recycling. Evaluation should collect information to decide what is needed, what is working, how to improve program, what has happened as a result.
Brion, M. & Newby, T. (1981). Research & Training-A Two Way Exchange. *Training Officer*, 17,9, 254-56.
    Case study involving evaluation of T&D function by Housing Training Project for Dept of Environment, UK. Evaluation project involved five interwoven roles (catalyst, educator, sponsor, source of power, researcher).

Brook, J. A., Shouksmith, G. A. and Brook, R. J. (1983a). Research Report: An Evaluation of Management Training. Pt I Training Needs. *Journal of European Industrial Training*, 7,4, 23-28.
Define and develop various evaluation concepts, and discuss the setting of key performance indicators against which to judge the effectiveness of training.

Brook, J. A., Shouksmith, G. A. and Brook, R. J. (1983b) Research Report: Training. Pt II Changes In Understanding. *Journal of European & Industrial Training*, 7,7, 11-15.
Evaluation is cyclic, corresponding to various levels of objective setting to determine whether time and money are well spent; provides basis for well informed decisions concerning future improvement. Use of control groups and variety of statistical measures discussed.

Brook, J. A., Shouksmith, G. A. and Brook, R. J. (1984). Research Report: Training. Part III Changes in Work Behaviour. *Journal of European & Industrial Training*, 8,3, 11-16.
View training as having three stages - evaluation necessary at each point: (period of prelearning, learning phase, on job application stage) Argue the need for 3, 6, 12 month follow-up and present a case study outlining methods used over a 12 month period.

Brookfield, S. (1982). Evaluation Models & Adult Education. *Studies in Adult Education*, 14 Sept, 95-100.
Provides an overview of educational models and their relevance for adult education.

Brown, M. G. (1980). Evaluating Training Via Multiple Baseline Designs. *Training Development Journal*, 34,10, 11-16.
Discusses internal validity and considers four major research designs based on Brethower and Rummler's work. Purpose of evaluation is to determine what bottom line results can be directly attributed to training.

Bryson, J. M. & Cullen, J. W. (1984). A Contingent Approach To Strategy & Tactics In Formative & Summative Evaluations. *Evaluation & Program Planning*, 7,2, 267-290.
Argue for a move away from 'one best way' approach to a contingency approach.

Burgoyne, J. G. & Cooper, C. L. (1975). Evaluation Methodology. *Journal of Occupational Psychology*, 48, 53-62.
Considers some current issues in the methodology used in research evaluation in the managerial and training fields by comparing US and European approaches. Discuss the 'patient' vs. 'agent' framework in evaluation and point out the framework chosen has implications for the observational methodology. Methodological considerations include the timing of instructional measures, control groups, external and internal criteria, process of selecting measures.

Burgoyne, J. G. & Singh, R. (1977). Evaluation Of Training & Education. *Journal of European & Industrial Training*, 1,1, 17-21.
View evaluation in the context of the training/education process which sets in motion a chain of consequences made up of cause-effect links. Evaluation can be seen either as collecting data about consequences as an end in itself or as part of larger process of the management of education and training to make informed decisions. Discuss micro and macro evaluation (Type D and Type B) in terms of immediate vs. remote objectives and level of decision.

Byham, W. C. (1982). How Assessment Centres Are Used To Evaluate Training's Effectiveness. *Training*, 19,2, 32-38.
Presents case studies of four evaluations of reaming using assessment centres to evaluate.

Caliguri, J. (1984). The Evaluators Journal: A Qualitative Supplement To Program Evaluation. *Evaluation News*, 5,4, 54 58.
Discusses the use of the journal method to evaluate.

Clement, R. W. & Aranda, E. K. (1982). Evaluating Management Training: A Contingency Approach. *Training & Development Journal*, 36,8, 39-43.
Evaluation must consider variables other than just the training course, e.g. organisational setting within which manager attempts to use training, unique characteristics of manager to be trained, nature of the organisational problem to be solved by training. Propose a Contingency Framework for evaluation of management training.

Covert, R. W. (1984). A Checklist For Developing Questionnaires. *Evaluation News*, 15,4, 74-78.
Practical guidance in developing evaluation questionnaires.

Cummings, O. & Nowakowski, A. (1984a). Course Evaluation Procedures In Professional Education. *Evaluation News*, 5,4, 28-32.
Advocate formative evaluation during course development stage. Present a useful three dimensional evaluation model.

Cummings, O. & Nowakowski, A. (1984b). Microcomputer Training Impact Study. *Evaluation News*, 5,4, 43-47.
Discuss how a large accounting firm evaluated a micro computer course.

Del Gaizo, E. (1984). Proof That Supervisory Training Works. *Training & Development Journal*, 38,3, 30-31.
Using Kirkpatrick's model, he gives some guidelines for data collection at each level.

Deming, B. S. & Phillips, J. A. (1974). Systematic Curriculum Evaluation: A Means And Methodology. *Theory Into Practice*, 13,1, 41-45.
Argue that evaluation has often been no more than an application of 'conventional wisdom' which involved describing philosophic underpinnings, intents, process and product of program, checking internal consistency, and applying appropriate external criteria of judgement.

Dhanens, V. (1984). Evaluation Of Instructor Performance. *Evaluation News*, 5,4, 37-40.
Presents a method for instructor evaluation.

Dopyera, J. & Pitone, L. (1983). Decision Points In Planning The Evaluation of Training. *Training & Development Journal*, 37,5, 66-71.
Argues for planned strategy of evaluation involving 8 decision points: (a) should evaluation be done - is it worth time and effort? (b) what purpose? (c) what will be measured? (d) how comprehensive? (e) who has authority and responsibility? (f) source of data? (g) how will data be collected and compiled? (h) how analysed and presented?

Duncan, W. J. (1984). Planning and Evaluating Management Education and Development: Why So Little Attention to Such Basic Concerns? *Journal of Management Development*, 2,4, 57-68.
Considers general trends in management education, focusing on poorly defined goals and lack of evaluation.

Easterby-Smith, M. (1981). The Evaluation of Management Education & Development: An Overview. *Personnel Review*, 10,2, 28-36.
Critically reviews current practices in training course evaluation (finding it to be mainly a ritual); offers reasons for non-evaluation and suggests asking participants and their bosses to complete short evaluation questionnaires before the course, at the end of the course, and some time later as a review of course effects. Contends that such a procedure has potential for aiding the learning process.

Easterby-Smith, M. & Tanton, M. (1985). Turning Course Evaluation From an End to a Means. *Personnel Management*, April, 25-27.
Look at parallel developments in educational evaluation which have relevance for evaluation of management training in three stages: (a) Cost Benefit Analysis prevalent in the 1960s. (b) Importance of context and impact of organisational value systems. (c) Aid to Decision Making. Defines evaluation as including any intervention aimed at providing feedback about the processes and nature of human development, the organisational systems and programs intended to facilitate it and the wider organisational context within which it occurs.

Eckenboy, C. (1983). Evaluating Training Effectiveness: A Form That Seems To Work. *Training*, 20,7, 56-59.
  Presents a simple diagnostic tool to identify blatant deficiencies as well as to pinpoint specific weak areas in terms of content, presentation, and applicability. Content and instruction are combined to give a measure of 'program total'. A sample form and how to calculate scores is provided.

Elkins, A. J. (1977). Setting Objectives. *Lifelong Learning*, 46,10, 22-23.
  Considers difficulties in evaluating programs where evaluation is linked to set objectives, particularly in adult education where reamer-determined and instructor-determined objectives may not be congruent.

Foxon, M. J. (1986). Evaluation of Training: The Art Of The Impossible. *Training Officer*, 22, 5, 133-137.
  Singles out four main reasons to evaluate: check if training led to relevant learning; check if transfer occurred; check if skills/knowledge have become integral part of job performance; assess cost effectiveness.

Freeman, A. (1978). Evaluation(sic) The Effectiveness Of Training Programs. *Training & Development In Australia*, 5,4, 20-21.
  Effective evaluation requires behavioural objectives, the involvement of supervisors in the training program, and the assessment by someone other than the trainer.

Galagan, P. (1983). The Numbers Game: Puffing Value On Human Resource Development. *Training & Development Journal*, 37,8, 48-51.
  Questions whether any of the simple or complex methods do measure HRD in a meaningful way. Proposes a matrix focussed on verification, relevance and diagnosis at four levels (entry capability; end of course performance; mastery of job performance; organisational performance).

Galvin, J. (1983). What Can Trainers Learn From Educators About Evaluating Management Training? *Training & Development Journal*, 37,8, 52-27.
  Applies Stufflebeam's CIPP model to management training and refers to an ASTD survey comparing trainer attitudes to Kirkpatrick's and the CIPP models.

Glass, G. V. & Ellett, E. S. (1980). Evaluation Research. *Annual Review of Psychology*, 31, 211-228.
  Compare seven alternative conceptions of evaluation to a set of standards (logic, science, ethics). Evaluation may be seen as applied science, as systems management, as decision theory, as assessment of progress to goals, as Jurisprudence, as description or portrayal, and rational empiricism. Best design is a unique compromise between the fundamental purpose of evaluation and the possibilities afforded by situation.

Goldstein, I. (1978). The Pursuit Of Validity In The Evaluation Of Training Programs. *Human Factors*, 20,2, 131-144.
  Discusses three validity issues: did the training make a difference? (internal validity); did they ream? (training validity); are they transferring the learning? (performance validity).

Goldstein, I. (1980). Training In Work Organisations. *Annual Review of Psychology*, 31, 229-272.
  Describes the stages which evaluation efforts have gone through (a) anecdotal, training reactions, (b) strict adherence to experimental/academic approach; (c) consideration of validity issues and design methodology; (d) recognition that program and evaluation interact with the organisation. Critical that evaluation skill centre on Kirkpatrick's model and claims that "the failure to develop methodologies for systematic observation of behaviour is a serious fault".

Grenough, J. & Dixon, R. (1982). Using 'Utilization' To Measure Training Results. *Training*, 19,2, 40-42.
  Suggest a strategic evaluation model to generate future oriented management information which is designed to identify whether or not trainees are using their experience. Evaluation should identify what results training should have produced, what results occurred, how worthwhile results are, and how results will be used.

Guyot, J. F. (1978). Management Training & Post-Industrial Apologetics. *California Management Review*, 20,4, 84-93.
 Claims that much of the current research relating to training simply begs the question by focusing not on the results of training but on the assessment of training 'needs'. Points out that carefully measured benefits are often ones which do not count for much, whereas some of the real benefits are not convincingly connected to the training experience because they are so poorly measured by available standards.

Harper, E. (1985). Evaluation As A Client Service. *Journal of European & Industrial Training*, 9,4, 9-11.
 Sees three stages to evaluation process: (a) investigation of context, (b) implementation, (c) reporting. Advocates a comprehensive formative and summative approach to the evaluation of training.

Harper, E. & Bell, C. (1982). Developing Training Materials: An Evaluation-Production Model. *Journal of European & Industrial Training*, 6,4, 24-26.
 Present their E-P model with three phases: (a) needs analysis - preparatory evaluation; (b) quality control function on first draft - formative; (c) summative evaluation - the real life evaluation.

Harries, J. M. (1981). Evaluating A Management Development Skills Program in Local Government. *Journal of European & Industrial Training*, 5,1, 2-4.
 Present an in-depth evaluation of a management development program using an interview technique and showing what sort of data can be gained from a fairly simple method.

Hawes, M. & Bailey, J. (1985). How A Validation Study of Engineering Courses Was Conducted. *Training & Development*, 4,1, 20-24.
 Outlines a case study in which an evaluation based on semi-structured interviews and questionnaires assessed whether the course produced the intended outcomes or should be improved.

Jerrell, J. (1984). Evaluation Experience in Business Settings. *Evaluation News*, 5,4, 15-17.
 Relates educational evaluation to HRD evaluation issues.

Kane, J. S. (1976). The Evaluation of Organisational Training Programs. *Journal of European Training*, 5,6, 289-338.
 Gives a detailed treatment of the major factors involved in evaluation including validity, quasi-experimental and non- experimental designs.

Kaye, M. (1985). The Myth of Program Evaluation. *Training & Development In Australia*, 12,4, 11.
 Argues that learning and performance, not the program, should be evaluated.

Keenan, K. (1983). Evaluation of Training. *Training Officer*, 19,2, 53-55.
 Develops the four processes approach of Warr, Bird and Rackham (1970) viz. context (TNA), input (resources), reaction (in order to improve program), and outcome (change in knowledge, skill, attitudes; change on job; overall organisational change).

Kelley, A. I., Orgel, K. F. and Baier, D. M. (1984). Evaluation: The Bottom Line is Closer Than You Think. *Training & Development Journal*, 38,8, 32-37.
 Consider evaluation critical to the economic survival of the T&D function and ultimately of the organisation and suggest collection of data measuring profit-relevant behaviours of trainees. Use of graphic analysis of pre/post testing outlined.

Kirkhart, K. (1981). Defining Evaluator competencies: New Light On An Old Issue. *Evaluation News*, 2,2, 188-191.
 Discusses the US Standards of utility, feasibility, propriety and accuracy and develops his own 8 categories of evaluation skills.

Kirkpatrick, D. (1983). Four Steps to Measuring Training Effectiveness. *Personnel Administrator*, 28,11, 19-25.
 Presents his four stage model.

Kirkpatrick, D. L (1977). Evaluating Training Programs: Evidence vs. Proof. *Training & Development Journal*, 31,11, 9-12.
Outlines features of his four stage model and highlights the difference between evidence and proof in evaluation.

Kruger, M. & Smith, K. (1986). Evaluating Management Training. *Training & Development in Australia*, 13,2, 20-22.
Present a case study in which evaluation is used to determine the impact of management training on student performance using pre/post testing. Agree that evaluation of interpersonal skills is difficult other than at the reactions level.

Lange, R. R. (1974). A Search For Utility In New Evaluation Thought. *Theory Into Practice*, 13,1, 22-30.
Examines the contributions of the major educational evaluators, including Scriven, Provus, Stake, Guba and Stufflebeam, but criticises their failure to provide specific methodologies which practitioners could implement.

Larson, R. E. (1985). The Value In Education. *Training*, 22,1, 92.
Evaluation is necessary to survive cost cutting and involves seething measurable objectives establishing and implementing cost effective plan to meet objectives, and measuring the degree to which objectives are met.

McCullough, J. M. (1984). To Measure In A Vacuum. *Training & Development Journal*, 38,6, 68-70.
Explains the Deficiency Analysis Review Technique (DART) which justifies training by quantifying the cost of not doing it. Suggests evaluating the need for/benefits from training before the design stage.

Mezoff, B. (1981). How to get Accurate Self Reports of Training Outcomes. *Training & Development Journal*, 35,9, 57-61.
Advances reasons why some evaluations show no improved level of learning when it has occurred. Identifies the problem of response shift bias when pre and post teasing is used as the main evaluation technique.

Minick, R. & Medlin, S. (1983). Anticipatory Evaluations In HRD Programming. *Training & Development Journal*, 37,5, 89-94.
Propose a model incorporating: (a) anticipatory evaluation which compares organisational needs to program objectives and assesses trial runs; (b) program evaluation which involves program effort evaluation (implementation as agreed) and program effect evaluation (objectives met); (c) organisational impact evaluation which is predictive evaluation (showing impact on organisational effectiveness).

Mmobuosi, I. B. (1985). An Alternative Approach To The Evaluation of Management Training: The Use of Protocol Analysis Method. *Management Education & Development*, 16,3, 262-268.
Offers an alternative to positivist evaluation methods and proposes phenomenological methods which use people's statements and behaviours to interpret their learning and so form the basis of the evaluation.

Morris, M. (1984). The Evaluation of Training. *Industrial & Commercial Training*, 16,2, 9-16.
Proposes a model incorporating eleven steps, but the first five are really an audit of the training function in the context of the organisation. Gives various techniques to be used at reaction, learning, and behaviour levels.

Putman, A. O. (1980). Pragmatic Evaluation. *Training & Development Journal*, 34,10, 36-40.
Refutes the academic evaluation approach which is overly concerned with truth and considers this paradigm inappropriate to HRD. HRD evaluation should be future (rather than past) oriented, provide reasonable evidence rather than irrefutable proof and provide an information base to make future decisions.

Rackham, N. (1973). Recent Thoughts On Evaluation. *Industrial & Commercial Training*, 5,10, 454-461.
 Discusses his attempts to develop predictive rather than descriptive evaluation methods by distinguishing between short cycle evaluation done during the course (either informally in the session or after each session), and long cycle evaluation (where evaluation and redesign is at the end of the course).

Rae, W. L. (1985). How Valid Is Validation? *Industrial & Commercial Training,* 31,1, 15-20.
 Distinguishes between validation, evaluation and assessment. Argues in favour of both subjective and objective measures and expands on a number of techniques including pre/post testing.

Salinger, R. & Deming, B. (1982). Practical Strategies For Evaluating Training. *Training & Development Journal*, 36,8, 20-29.
 Suggest evaluation should answer the question "What do you want to know about training?" and identify 6 strategies to address the purpose of evaluation.

Schearer, R. W. (1976). The Course Was Beaut ... But What Happens Now? *Training & Development In Australia*, 3,3, 8-13.
 Discusses the need to measure course effectiveness but confuses evaluation and transfer issues.

Siedman, B. (1979). Missing From The Curriculum: The Other Side of Program Evaluation. *Evaluation News*, No.12, Sept, 22-23.
 Discusses training evaluation in the context of the organisation and identifies some of the competencies needed.

Smith, M. E. (1980). Evaluating Training Operations And Programs. *Training & Development Journal*, 34,10, 70-78.
 Develops an evaluation matrix expanding Kirkpatrick's model, and proposes a 7 phase process to integrate evaluation into the entire training process.

Snyder, R., Raben, C. and Farr, J. (1980). A Model For The Systematic Evaluation of Human Resource Development Programs. *Academy of Management Review*, 5,3, 431-444.
 Suggest that evaluators should actively avoid either/or inquisition, recognise that a measure of control is gained by those to whom feedback is given, and develop a framework for the conceptualisation of the evaluation process. Propose a systematic model of evaluation for HRD which is an adaptation of Stufflebeam's CIPP model.

Stake, R. E. (1982). How Sharp Should The Evaluators Teeth Be? *Evaluation News*, 3,3, 79-80.
 Discusses the competencies required by an evaluator.

Stevenson, G. (1980). Evaluating Training Daily. *Training & Development Journal*, 34,5, 120-122.
 Considers need for ongoing evaluation to tailor a program and suggests daily evaluation meetings of 30 minutes duration with a selection of participants.

Swierczek, F. & Carmichael, L. (1985). The Quantity And Quality Of Evaluating Training. *Training & Development Journal*, 39,1, 95-99.
 Discuss Kirkpatrick's model in the context of management training evaluation in order to improve the program, give feedback to planners, managers, and trainees, and to assess skill development.

Thompson, J. (1978). How To Develop A More Systematic Evaluation Strategy. *Training & Development Journal*, 32,7, 88-93.
 Trainers must consider why they want to evaluate if they are to develop a strategy which will provide an orderly approach to compare 'what is' with 'what is wanted'.

Trapnell, G. (1984). Putting The Evaluation Puzzle Together. *Training & Development Journal*, 38,5, 90-93.
 Considers the purpose of evaluation as assisting in design and replication of successful training programs, and determining reasons for failure.

Tyson, L. A. & Birnbrauer, S. (1985). High-quality Evaluation. *Training & Development Journal*, 39,9, 33-37.
> Describe evaluation as a system of quality control for training and HR and identify five steps involving evaluation mission statement, evaluator selection process, evaluator's role, evaluation methods and procedures, and training of evaluator.

Williams, G. (1976). The Validity Of Methods Of Evaluating Learning. *Journal of European Industrial Training*, 5,1, 12-20.
> Discusses content, criterion related and construct validity, and highlights the difficulties in evaluating according to objectives. Concludes that the higher levels of objectives are the ones which really count but which are difficult or impossible to evaluate with a reasonable degree of validity.

Wittingslow, G. E. (1986). Making Training Effectiveness Work. *Training & Development in Australia*, 13,4, 8-9.
> Claims the distinction between validation and evaluation is blurred and that much of the discussion of evaluation is about validation. Introduces his technique of single case research design, develops the methodology and discusses applications in context of Kirkpatrick's model.

Woodington, D. (1980). Some Impressions of Evaluation of Training in Industry. *Phi Delta Kappan*, 61,5, 326-8.
> Outlines five impressions he has of evaluation: no clear realisation that a training program is an instructional system; no clear perception of what constitutes evaluation; nature of the organisation will influence evaluation strategies; evaluation programs differ depending whether they are in-house or externally done; lack of personnel trained in evaluation methodology.

Zenger, J. & Hargis, K. (1982). Assessing Training Results: Its Time To Take The Plunge. *Training & Development Journal*, 36,1, 11-16.
> Practitioners need to consider issues of rigour, relevance and economy when evaluating. Apply these three criteria five types of evaluation, and argue for a defined percent of the total training budget to be allocated to evaluation.

Author: Marguerite Foxon is National Director of Professional Education for Coopers & Lybrand. She has had a long standing interest in evaluation and can be contacted at Coopers & Lybrand, GPO Box 2650, Sydney NSW 2001. (Phone 285 7777)