RESEARCH INTO

EFFECTIVE EVALUATION TECHNIQUES FOR

TRAINING AND LEARNING PROGRAMMES

REPORT TO ALL WAYS LEARNING

AUGUST 2002

Annabel Jackson Associates
52 Lyncombe Hill
Bath BA2 4PJ
Somerset
Tel: 01225-446614
Fax: 01225-446627
Email: ajataja@aol.com

CONTENTS

1   INTRODUCTION

1.1   The Brief

All Ways Learning is a new organisation to support and promote continuing professional development among arts managers, be they working within an organisation or self-managed artists. It operates as a consortium of partners with an interest in continuing professional development: two regional arts boards, local authorities, the arts marketing agencies in the southern region, Arts and Business and the University of Brighton.

The purpose of this research is to explore alternative methods of evaluating learning activities across the range of continuous professional development, from short courses to mentoring, from action learning sets to job swaps.

We use the term 'course' throughout this guide as a short hand for any learning opportunity - whether formal or informal, self-guided or externally operated.

1.2   Methodology

There were three stages to the research:

➢ Identifying literature on evaluation of training and learning.

➢ Reading literature on evaluation of training and learning.

➢ Writing up ideas and literature on evaluation of training and learning.

Evaluation theory is a deep and well considered field in which many of the challenges for evaluating training have been discussed and addressed. This paper drew on the general evaluation literature as well as reviewing over a hundred specific references on training and learning evaluation.

1.3   Learning and Training

Learning can have seven broad purposes (table 1).

Table 1: Purposes of learning

1. **Current job purposes**. To give people the skills and knowledge they need to perform more effectively in their current jobs.
2. **Advancement and promotion**. To give people the skills and knowledge they need to achieve promotion and other career advancement goals.
3. **Organisational capacity**. To build the organisation's capacity to perform in an uncertain future.
4. **Orientation and acculturation**. To give people the knowledge and understanding they need to become oriented with, and identified with their organisation.
5. **Employee capacity**. To increase people's capacity to handle stress, work healthily, cope with change and increase their personal resilience.
6. **Leadership capacity**. To develop leaders in the organisation.
7. **Personal benefit**. To provide people with learning opportunities to gratify their interests and develop skills and knowledge they personally wish to acquire but are not needed for current or future performance.

Source: Brinkerhoff & Brown (1997).

Evaluation can be usefully applied across the whole spectrum of learning opportunities: from informal self-study at one end of the spectrum, through one-off workshops, to whole programmes of formal training courses.

Human resource development programmes, especially management development programmes, often fail to add value to the organisation because (Berry; 1990):

➢ Programmes are not linked to specific strategies, challenges or problems in the organisation.

➢ Programmes are designed to create awareness and understanding, but not competence.

➢ Programmes focus on individuals not operating units.

➢ Participants attend programmes for reasons other than personal or organisational need.

➢ Programmes fail to help participants confront reality.

Evaluation can help to increase the value of learning and training programmes.

1.4   Purposes of Evaluation

"Evaluation is the process of determining the merit, worth[1] and value of things, and evaluations are the products of that process." (Scriven; 1991)

Scriven (1991) sees evaluation as a trans-discipline, wider than one area of applied social science. It provides basic tools that span subject areas, rather in the manner of logic, design, and statistics. Evaluation combines two processes. Compiling, analysing, and simplifying or standardising data is only the first step in evaluation. The second step involves the imposition of values or standards. Scriven sees applications such as programme, personnel, product and material evaluation as branches of the core discipline.

The need for evaluation is often dictated externally by funders. However, evaluation is most useful when designed to guide internal processes. There are six broad benefits of carrying out training evaluation:

➢ **Reflection**. Self-evaluation can be part of the learning process itself, helping to deepen and embed learning, rewarding the learner with knowledge of results as well as identifying any need for remedial training.

➢ **Integration**. Evaluation can clarify expectations of training and thereby help to ensure that the different stakeholders involved have consistent expectations.

➢ **Greater effectiveness**. Evaluation can provide valuable insight into ways, big and small, in which training can better meet its objectives.

➢ **Greater efficiency**. Evaluation can help reduce the waste in training programmes.

---

[1] Merit means the professional merit; the extent to which the individual candidate measures up to the accepted standards of their profession. Worth is the value of an employee to the institution, which arises from considerations other than their performance of their duties.

➢ **Accountability**. Evaluation can be used to justify expenditure on training and to demonstrate the benefits of investment in training.

➢ **Publicity**. Evaluation can provide useful material to explain and market training. This can raise the profile for training and increase the commitment to its work.

1.5 Evaluation Issues

Evaluation can consider a number of different questions about training and learning:

➢ Do the course objectives match the organisation's learning needs?

➢ Does the course content matches the course objectives?

➢ Do the right trainees attend the right training?

➢ Do sufficient trainees attend the course?

➢ Do trainees have a positive reaction to the training or learning?

➢ Do trainees acquire the skills, knowledge or attitudes intended?

➢ Are there any barriers preventing trainees from applying their learning in the work environment?

➢ Do trainees' apply their learning in the workplace?

➢ Is the trainees' change in performance substantial enough to affect organisational performance?

➢ Is impact achieved at the minimum cost?

➢ Is the impact sustained?

➢ Is the training organised as effective as comparable learning opportunities?

As well as these practical purposes, training can have a number of symbolic or ritualistic purposes such as rewarding people with a break from work, providing a rite of passage or rite of renewal, demonstrating commitment to members of staff. Evaluation is less valuable for these types of training (McEvon and Buller; 1990).

The series of questions can be seen as a value chain, where each stage links to the next. Linkage is not automatic, however. Evaluation helps to see at which stage the chain of assumptions breaks down.

Training is moving to a view of the trainee as an active participant in the learning process. Evaluation needs to acknowledge this empowerment and shift from a top down determination of evaluation criteria to use more reflective methods that treat the learner as self-directing (Geertshuis; 2002).

2 DIFFERENT LEVELS OF EVALUATING TRAINING AND LEARNING

2.1 Introduction

There are a number of different models of the stages in training evaluation. An early model, which still has value, is that by Kirkpatrick. He identifies four levels to evaluation:

➢ **Reaction**. This measures trainees' views on the training, a version of customer satisfaction. Happy sheets, the most common form of evaluation in training, are a measure of reaction.

➢ **Learning**. This is a measure of the knowledge acquired, the skills improved or the attitudes changed due to training.

➢ **Behaviour**. This measures whether the trainee puts the learning into practice in their work.

➢ **Results**. This looks beyond the individual to see if the training had any effects at the level of the organisation.

This guide adds a further stage, which is inferred but not treated separately in Kirkpatrick's work:

➢ **Cost-effectiveness of training**. This compares the costs and the benefits of training.

Various figures in the literature (e.g. Conway; 2002) suggest that 90% of organisations measure reaction, but only 20% measure learning, 10% measure behaviour and fewer than 5% measure results. Ideally, the five levels of evaluation should all be carried out. There are three reasons for this:

➢ **The value dilemma**. There is a dilemma in evaluation that the most important issues are generally the most difficult to measure. For training, reaction is the easiest level to measure, but behaviour, results and cost-effectiveness are the most valuable subjects. As we will see, behaviour, results and cost-effectiveness are more difficult to measure because of lags, difficulty in attributing cause, and problems in quantifying intangibles.

➢ **Different audiences**. Reaction is focused on the consumers, the people who attend the training. Results and cost-effectiveness are usually of more interest to the clients, the people who fund, support and approve training.

➢ **Following a chain of logic**. Measuring each stage in the process allows the evaluator to see at which point assumptions break down. Training does not automatically lead to organisational benefit.

Generally speaking, reaction would be measured frequently, for example, for each training event. Results and cost-effectiveness would be measured less frequently, for example, taking a sample of events across an entire training programme.

2.2 Evaluation of Reaction

2.2.1 Issues

The vast majority of training courses only use happy sheets. The problem with happy sheets is that they mainly measure reaction without covering the other five levels of evaluation. They cannot give insight into whether the training has met its aims in the workplace.

There are a number of specific criticisms of happy sheets:

➢ **Training can receive a positive rating but still fail to meet its objectives**. Challenging training can be uncomfortable for trainees but still successful in terms of learning objectives. Critics of happy sheets complain that emotional or exciting training sessions, delivered with more panache than substance, can 'outsmile' a well-constructed and relevant course presented in a lower key, or a valuable but difficult course.

➢ **Responses can be affected by the feelings from the course**. Some courses generate an immediate feeling of euphoria that is diminished when trainees attempt to apply what they have learnt. Trainees might confuse "like" with "worth".

➢ **Reaction is subjective**. The use of scoring can give a false impression of precision. In practice, participants' responses might reflect their interaction with the trainer: their desire to please or punish the trainer.

➢ **Responses can be superficial**. Forms might be completed with relatively little thought, especially if tick boxes are provided without open questions.

➢ **Scoring tendency.** There is ample evidence that different people have a pattern to their rating, which can be lenient, severe or in the middle (central tendency). Differences reflect in part the ethnicity, sex, level of education, status and experience of the scorer. Furthermore, the responses to different questions can be highly correlated. This 'halo effect' occurs where a trainee who is strongly dissatisfied with one aspect of the training rates all questions lowly.

➢ **Questionnaires tend to be badly written.** Most trainers are not evaluators and do not know the principles for writing good questionnaires.

However, measuring reaction also has some benefits. The strengths are:

➢ **Happy sheets show that the trainer is interested in the trainees' response.** Happy sheets are part of the dialogue and relationship between the trainer and the trainee.

➢ **Happy sheets give fast feedback**. They can help identify problems immediately while information is fresh in participants' minds.

➢ **Trainees are a captive audience**. Asking for happy sheets to be filled in during the training ensures a high response rate.

➢ **Happy sheets are simple to administer**. They are also cheap.

Some of the criticisms of reaction reflect in part a lack of clarity about the term. Four distinct elements of reaction have been identified empirically: enjoyment, perceived usefulness, perceived difficulty and motivation to apply the material learned (Warr &

Catriona; 1999). Subsequent job performance does appear to correlate with motivation to apply the material and with perceived usefulness, but not with enjoyment.

On balance, there is an argument for including happy sheets, so long as their limitations are understood. They do not give a full picture of the effectiveness of the training. The problem with badly written questionnaires is easily remedied since the principles of writing good questionnaires are very simple to learn and apply.

2.2.2   Good Practice

There are some golden rules about writing questionnaires:

➢ **Only ask about issues that are relevant**. Before you carry out the training you should have a clear statement of your objectives. Clarify the assumptions you are making in delivering the training. Do not ask about topics that are not important or not capable of change. Focus on the more meaningful elements of reaction: for example, ask whether participants' felt that the case studies were relevant to their art form or particular circumstances, not whether the participant liked the case studies.

Table 2: Pros and cons about asking about different issues in happy sheets

	Pros	Cons
The environment	Problems with the environment can colour trainees' responses to the training. Evaluating the environment helps to identify the nature of the problem.	The environment is not in itself relevant to measuring learning.
The training	Data can provide immediate information to improve delivery.	Trainees have uneven comparative histories. What one person considers outstanding might appear ordinary to another. Ratings are sensitive to mood. Data are retrospective. Ratings might be the result of selective memory.
The trainer	If the trainees do not make these judgements, there might be no other source. Poor material can be adversely affected by a poor trainer. Evaluating the trainer helps to identify where the	Trainers might change their approach if they know they will be rated. Ratings might be coloured by personal factors. Trainees are unlikely to be professionally

	weakness lies.	qualified to comment on trainers. The evaluation is one sided if it asks about the trainer's performance but not about the trainees' efforts to learn.
Learning	Happy sheets can test memory and motivation to learn.	Learning is more apparent once trainees return to their workplace.

> **Ensure that the questions are clear and well written**. Table 3 lists the principles for writing questions.

Table 3: Principles for writing questions

1. **Use short questions**. They are easier to understand.
2. **Do not use multiple questions**. A multiple question is one that asks about two aspects at the same time. For example, the question "How do you rate the visual aids and the handouts?" is difficult to answer if the trainee had a positive impression of the visual aids but did not like the handouts.
3. **Make questions as narrow and precise as possible**. Respondents have difficulty rating broad questions because their answer is often "it depends". The Fenman Evaluation Toolkit contains many questions which are too broad to be valid. For example, one of their reaction questionnaires asks "Generally speaking, how much practical value have you gained from the workshop?" This gives the respondent insufficient information to measure the amount of practical value: this could be interpreted in terms of the number of practical points, the frequency with which one practical point could be applied, the ease with which the information could be applied, or the financial and organisational benefits for the information etc.
4. **Explain the criteria by which judgements are to be made**. For example, the question "What did you think of the visual aids used" (score from 7 good to 1 poor) is confusing because it could refer to the clarity, relevance, or number of the visual aids.
5. **Do not use loaded language**. For example, the question "How well do you rate the visual aids used?" is leading the respondent because it implies that the answers are simple levels of wellness.
6. **Do not ask questions that presume some behaviour from the participant**. The respondent will not know how to answer that the question does not apply.
7. **Avoid double negatives**. This is where the respondent is asked to disagree with a negative statement. Where possible, state questions or comments in the positive.
8. **Do not use jargon**. Trainees might not all interpret it in the same way.
9. **Do not ask hypothetical questions**. For example, the question "How useful do you think the handouts are going to be?" is asking the trainee to predict the future, which is not likely to produce a valid response.
10. **Take account of cultural differences**. Cultures vary in how authority is

> viewed, the value of verbal versus written skills, language skills and pace, cultural values as they relate to judging self and others, demographic and psychographic characteristics (Goss and Kaska; 1998).
> 11. **Keep scoring as simple as possible**. The questionnaire should not mix questions that need agree/disagree, good/poor, useful/not useful, high/low answers. Mixing the scoring lengthens the questionnaire and can confuse or annoy the respondent.

➢ **Include open questions**. One way of improving the data from happy sheets is to have an open question at the end of each rating scale asking why they have given this rating. This encourages the respondent to think about the question more carefully before they answer, and also gives data to help interpret the responses.

➢ **Use a rating scale with an even number of points**. This forces the respondent to give a response rather than ticking the middle point. A four or six point scale gives sufficient flexibility for a response without confusing the respondent. A two point scale is too narrow. The most positive score should be denoted by the highest number.

➢ **Put demographic questions at the end of the questionnaire**. Once respondents have spent time answering the other survey questions they are likely to complete the demographic section at the end because of "completion tendency". When controversial demographic questions are placed at the beginning of a form respondents might react negatively and not complete the other questions.

➢ **Ask for background information on the respondent and their organisation.** There is a view that happy sheets should be confidential. However, if your evaluation is to extend beyond happy sheets, for example, to include follow up work on whether they applied their learning, then you need to obtain background information.

➢ **Ensure that the questionnaires are well presented**. A sloppy appearance gives the impression that the evaluation is not given a high priority. Have someone proof read the questionnaire to ensure it has no spelling or grammatical errors.

➢ **Pilot the questionnaire**. The person writing the questionnaire might have a different perspective or use of language to those who will fill it in. Piloting allows the author of the questionnaire to see it from the recipients' point of view. Apply the questionnaire to a group of people similar to the trainees and examine the answers to see if any of the questions is ambiguous. Members of the pilot group can also be asked for their comments on the questionnaire: the order of questions, and any gaps.

Table 4: Examples of possible issues for a happy sheet

Objectives	Clarity of objectives Balance of objectives Relevance to personal objectives Relevance to organisational objectives Prior knowledge
Structure	Logic Step by step progression Explanation/illustration Recapping
Time	Convenience of the date Convenience of the time Pace Depth Overall length
Environment	Location Accessibility Food Comfort Layout Library
Training methods	Time for absorbing knowledge Time for interaction with other participants Assignments Assessment Balance Preferences
Training materials	Clarity Depth Insight
Trainer	Qualification Skills Attitude to participants Preparation Use of visual aids Interest
Content	Appropriateness Coverage Gaps
Satisfaction	Expectations Satisfaction of expectations Value for money Whether the trainee would recommend the course to others
General questions	Most and least useful sessions Improvements to the course Actions planning Motivation to apply the material

There are some simple rules for administering the forms:

➢ **Allow sufficient time for trainees to complete questionnaires.** Hand out the forms slightly before the end of the event. If questionnaires are handed out at the end, trainees might be in a hurry to leave, or already thinking about the journey home. They will not give the evaluation their full attention.

➢ **Obtain a 100 per cent response rate.** Do not ask trainees to take forms home for completion. This reduces the number of completed forms obtained and introduces a possible area of bias.

➢ **Treat the questionnaires seriously**. Explain the purpose and value of the evaluation, and tell trainees what will happen to the information. Emphasise that their responses will help to improve the training.

➢ **Do not try and influence the responses**. Giving examples of what other trainees have said, looking at the forms as people are writing them or praising trainees, are all approaches which could reduce the willingness of the respondent to be honest.

➢ **Have a neutral person collect the forms**. Trainers should not collect the evaluation forms for their own courses.

Most training courses use happy forms. Few then take the step to analyse the data. The data should be analysed on a spreadsheet. This should only take an hour or so for an average course. Analyse the data:

➢ **Identify the strengths and weaknesses of the learning process.** Compute an average rating for each question. Examine the scores for each question rather than calculating one rating across all questions. Averaging scores across questions is implicitly assuming that all questions are equally important, which might not be correct. Use this information to analyse the strengths and weaknesses of the course.

➢ **Monitor customer satisfaction.** Calculate the percentage of participants who say that they are satisfied or very satisfied with the course.

➢ **Develop norms and standards.** Set a target for levels of satisfaction with different aspects of the course.

➢ **Evaluate instructors.** Compare average scores between different trainers, taking into account factors, such as the size of group, which might affect comparisons.

➢ **Identify needed improvements.** Answers to open questions should be grouped according to apparent themes so as to provide recommendations about improvements.

➢ **Use material for marketing.** Quotations from participants can help explain the course to those who might be interested. Quotations should not give an inflated picture of the course, since this will simply attract unsuitable people and raise expectations, both of which could depress satisfaction. Information should also be used to help understand the sorts of people who do not benefit from the course.

➢ **Compare subgroups in the survey.** Analyse more and less experienced groups separately to see if the training addressed the needs of both.

2.3   Evaluation of Learning

2.3.1   Issues

This third level of evaluation measures whether the trainees have acquired the knowledge, skills or attitudes intended by the training. There are three broad ways of measuring learning:

➢ Paper tests or exercises.

➢ Simulations, role plays or demonstrations.

➢ Self-reports such as a learning log, portfolio or critical incident report.

These are each appropriate in different circumstances. Paper tests are particularly useful for measuring knowledge. Simulations can be helpful for testing skills. Written reports and follow up interviews can be useful for measuring knowledge, skills and attitudes.

2.3.1.1   Tests

Tests can measure knowledge in four areas: recall (memory), comprehension (interpretation), application (putting principles into practice) and analysis (patterning of information). Tests have several possible benefits:

➢ **Reliability**. Standardisation reduces the variables covered by measurement and increases the consistency of assessment.

➢ **Ease of use**. Tests are easy to administer and simple to analyse.

➢ **Motivation**. Anticipation of a test can motivate trainees to study more.

The criticisms of tests are:

➢ **Cultural bias**. Standardisation reduces the scope to take account of the personal circumstances of individual trainees.

➢ **Validity.** Tests take a narrow view of learning. Results are also affected by test-taking ability, which does not necessarily correlate with ability. As with other standardised approaches, tests trade off validity for reliability (table 5): the more standardised and simplified the procedure, the easier it is to repeat (reliability), but the less representative of the area of interest (validity).

➢ **Expense.** Costs include devising or purchasing tests and exercises, training assessors, hiring accommodation, and organising travelling for candidates.

There tends to be a trade-off between using standardised tests that have high validity (but do not match the objectives or syllabus of the training delivered), and customised tests, which reflect the philosophy and aims of the course (but have not been validated statistically) (McClelland; 1994). In practice, off-the shelf tests are unlikely to be available.

The principles of good practice in designing and applying tests are:

➢ **Representativeness**. Ensure the test covers as much of the curriculum for the course as possible.

➢ **Planning**. Each part of the test should be thoroughly planning, including the timing, the preparation of the participant, the collection of necessary materials and tools, and the evaluation of results.

➢ **Consistency**. All participants should be given the same instructions and should work under the same conditions.

➢ **Avoidance of trick answers**. Information that might lead participants astray should not be included.

➢ **Quality control**. Tests should be piloted. Calculate the difficulty of each question - the percentage of trainees who produced the correct answer. Questions with very low percentages answering correctly might be too difficult (Sullivan et al.; 1993). Calculate also the discrimination of each question. Divide the respondents into two groups. Looking at the overall results across the questions, group the top half as high scorers and the bottom half as low scorers. Calculate the number of times the high group answered each question correctly minus the number of times the low group answered correctly divided by the total number in the group. Ideally questions are answered correctly more often by high scorers. If this is not the case, the question might be ambiguous or misleading.

➢ **Appraisal**. Procedures should be developed for objective evaluation. Acceptable standards should be developed.

Table 5: Types of validity and reliability

Validity	**An instrument's ability to measure the characteristic or ability it purports to measure: whether a statistically significant relationship exists between a predictor and a measure of successful performance.**
Face validity	The credibility of the measure to those to whom it is applied.
Content validity	Whether the items or questions in an instrument are a well balanced sample of the content domain to be measured.
Construct validity	Whether there is a relationship between the variables that are measured and the abstract constructs which they are intended to represent.
Predictive validity	Whether the results from one set of observational data are predictive of another set of data such as job performance.
Concurrent validity	Whether the test correlates with other, well validated measures of the same topic administered at the same time.
Reliability	**The consistency in an instrument's performance: whether differences in measurements are the result of changes in the underlying variable rather than measurement error.**
Test-retest consistency	Whether the results produced from applying the instrument on different occasions are highly correlated.
Internal consistency	Whether the results from different parts of the same instrument are highly correlated.
Inter-rater agreement	Whether the ratings produced by different raters are highly correlated.

Source: Oppenheim (1992)

Tests have become unfashionable in Britain because of the movement against seeing learning as purely about knowledge acquisition. However, there are cases where training is clearly about imparting knowledge – an example might be in courses on the legal duties of directors or health and safety - and tests would therefore be the ideal evaluation method. Tests can also be designed as checklists for participants to revise their knowledge at points in the future.

2.3.1.2 Simulations, Role Plays or Demonstrations

Simulations, role plays or demonstrations have several advantages:

➢ **Validity**. Exercises measure more complex skills than are traditionally covered by exams.

➢ **Reliability.** Systematising the tasks improves the consistency of the data.

➢ **Flexibility**. Simulations can present circumstances that might take time to occur in the workplace.

The disadvantages of simulations and role plays are:

➢ **Artificiality**. Some people find simulations and role plays too artificial and therefore do not give their full commitment. Learners can sometimes adopt stereotypes rather than displaying their full emotions or attitudes.

➢ **Administration**. Success can depend on the skills of the trainer, and the co-operation of other trainees, in organising the simulation or role play.

Performance in a simulation does not always equate to performance in the workplace. For example, simulations can overlook barriers to application existing in the workplace.

2.3.1.3 Learning Logs, Portfolios and Action Plans

Learning logs (table 6), portfolios and action plans (table 8) are used for assessment of NVQs and other examinations, as well as for professional training.

Table 6: A learning log

| Course ……………… Date …………… |
| Name …………………………………………. |
| Things I want to discuss |
| Things I want to know more about |
| Things I want to remember |
| Things to do |

Source: Rae (1999)

A portfolio is:

> "The structured, documentary history of a set of coaching or mentored acts of teaching, substantiated by samples of student portfolios, and fully realised only through reflective writing, deliberation, and conversation." (Shulman; 1998)

Portfolios typically include elements of self and peer assessment, work observation and mentoring (table 7).

Table 7: Possible elements in a portfolio

A learning log or diary
Personal development plans and action plans
Written responses to questions or assignments
Records of observations or simulations
Statements of competence
Videos or audio tapes of work
Surveys of beneficiaries

Portfolios present the following potential benefits:

➢ **Construct validity**. Portfolios focus on authentic performance and knowledge in use.

➢ **Content validity**. Portfolios include samples drawn from many activities in many settings over a longer period of duration than most on-course or one-test samples.

➢ **Reflection**. Talking about trainee teacher assessment, Shulman (1998) emphasises that "a portfolio is a theoretical act". Portfolios encourage trainees to examine their assumptions and beliefs about training, to link personal and professional views, theories in use (behaviour) and espoused theories (beliefs and attitudes).

➢ **Documentation**. Production of clear evidence is at the heart of this method of assessment.

➢ **Initiative**. Portfolios lead trainees to internalise standards for excellence. This shifts responsibility for documenting proficiency to the trainee.

➢ **Flexibility**. Portfolios can be adapted to the cultural or socio-economic backgrounds of the trainee, to all different levels and kinds of training.

➢ **Celebration**. Portfolios emphasise the strength of trainees, not their weaknesses.

Guidelines for organising portfolios are (Peterson; 2000):

➢ **Clear purpose.** Portfolios should be linked to learning objectives, for example, they could be structured according to items in an action plan (table 9).

➢ **Portfolio construction handbook.** Trainees should be given a clear guide to constructing the portfolio. This should specify the types and number of work samples required, the length and structure of written commentaries, the deadlines for submitting materials for review by mentors and supervisors, the evaluation criteria and processes, feedback and assessment processes.

➢ **Explicit evaluation criteria**. Portfolios should be structured around individual goals and organisational standards. Standards should provide clear targets for performance.

➢ **Reflection**. The contents of a portfolio should be framed by captions and written commentaries that explain and reflect on the contents of the portfolio.

Table 8: Action plan elements

PERSONAL ACTION PLAN	ISSUES
What	Personal action objectives Priorities among these objectives
How	Resources or support needed for implementation Appropriate tactics for implementing the plan The effect of the plan on others Any possible obstacle to implementation Any special conditions that will help implementation
When	Targets to measure satisfaction of objectives
Follow up interview	Items of the plan implemented Degree of success from this Items not implemented Reason for not implementing these Any planned actions that did not achieve the desired results Reasons for this failing Plans for further action on successful and unsuccessful points

2.3.2   Good Practice

Measurement of learning should:

➢ **Clarify learning objectives**. Consider whether the purpose of the training is to maximise the learning that has occurred during the programme or just to ensure that trainees have reached target standards of competence (Sacket and Mullen; 1993).

➢ **Measure change over time**. Except where trainees are learning an entirely new skill, measurement is more valid if taken before and after the training.

➢ **Try and measure the learning objectively**. Self assessment can lead to inflated judgements before the training. Part of the benefit of training can be in trainees realising how little they knew before the course.

➢ **Use a control group, if feasible**. Control groups are discussed in the next section.

➢ **Ensure instructions are clear**. The rules for writing questions given in the previous section are applicable here.

➢ **Analyse the results statistically**. Differences between before and after scores might be too small to be significant.

_____

2.4   Evaluation of Behavioural Change

2.4.1   Issues

Behavioural change refers to the trainees' job performance after training, the extent to which their learning is transferred into action. Measuring behavioural change is more complicated than measuring learning and reaction.

Making the link between training events and behaviour on the job is conceptually difficult:

➢ **The desired behaviour might be difficult to define.** Especially in management training, there might be a variety of ways in which learning could be manifested.

➢ **Trainees might not have the opportunity to use their learning at work**. People cannot change their behaviour if there are no opportunities for them to do so. And even if a trainee has an opportunity to apply the learning, he or she might not do so immediately. Even if the trainee applies the learning after the training, they might not do so consistently.

➢ **There might be obstacles at work that prevent trainees from using their new learning**. These could include resistance from co-workers, difficulty in changing established systems, or lack of security.

➢ **Trainees might have encountered other learning opportunities as well as the training course**. This is described as the confounding variables effect.

➢ **Changes might be difficult to quantify**. Benefits might be intangible such as an increase in leadership or team building.

➢ **Hawthorne effect**. Trainees might improve their performance because of greater self-consciousness, but this change might be temporary.

Measurement of behavioural change is usually carried out using questionnaires or interviews, either of the participants themselves, or of those in a position to observe their behaviour (table 9).

Table 9: Sources of data for behavioural measurement

Follow up surveys
Follow up interviews
Interviews with supervisors, subordinates or managers
Observation on the job
Follow up focus groups
Programme assignments
Action planning
Performance contracting
Performance monitoring

Source: Phillips (1997a)

Self-assessment can be made more objective by providing a structure of relatively concrete questions (table 10).

_____

Table 10: Possible issues for behavioural measurement

Their view of how far training objectives have been met

Progress with participants' action plans

Rating of relevance of the training elements to their work

Whether participants have used the course materials since the course

Whether they have applied specific action points from the course, and how often

Any specific accomplishments related to the course

How the course has effected general indicators such as quality or time of work

A general question: what is the participant doing differently since they attended the course

Any barriers to applying the skills, knowledge or attitudes from the course

Any enablers helping them to apply the skills, knowledge or attitudes from the course

Information on participants' learning can be obtained from interviewing supervisors, managers, colleagues, or customers about their observations. The advantages of using follow up interviews from third parties in this way are:

➢ **Co-opting support for the training**. Involving other workers can increase their awareness of the training. This can be significant because in some types of training, opposition from managers or co-workers can be a major reason that trainees cannot put their learning into practice.

➢ **Reliability**. The availability of a large number of interviewees increases the reliability of data because individual biases are diluted.

The weaknesses of using feedback from observation are:

➢ **Low content validity**. If observation is carried out rarely or for a short time the picture of trainee performance that it gives will be unrepresentative.

➢ **Low inter-rater reliability**. Assessments can be biased by the raters' own views and their existing relationship with the trainee.

➢ **Partiality.** Only part of the trainees' performance is visible.

Observational techniques can be broadly classified into (Stodolsky; 1990):

➢ **Open systems**. The observation record contains a description of a sequence of behaviours in ordinary language or mechanically reproducible form. An effort is made to fully describe all behaviour as it occurs and to avoid interpretation or selection.

➢ **Closed systems**. The observation record focuses on specific types or aspects of behaviour. Closed systems include category and sign systems, as well as behaviour checklists and rating scales.

➢ **Low inference**. Specific and easily identifiable behaviours are codified. Low inference techniques are aimed at increasing objectivity. They give priority to reliability.

➢ **High inference**. Behaviour is observed and deductions drawn from it. These deductions are presented in scores or ratings. Qualitative judgements are inherent in the scoring or rating. High inference techniques are aimed at capturing the complexity of performance. They give priority to validity.

Many of the weaknesses of observation listed above apply to the use of rating scales. Ratings are problematic where they are closed systems but also high inference. Ratings scales that are applied after open observations are more transparent and easier to validate.

Guidelines for effective observation are:

➢ **Observers should be fully prepared**. Ideally observers should be trained for the evaluation and given a chance to practice their observation skills.

➢ **Observations should be systematic**. Observers should have a clear picture of what behaviour they are expected and how it will be manifested.

➢ **Observers should know how to interpret and report what they see**. Observers should know how to summarise behaviour and report results in a meaningful manner.

➢ **The observer's influence should be minimised**. Except for mystery shopping observations, it is impossible to completely isolate the overall effect of an observer. To the extent possible, the observer should blend into the work environment.

2.4.2   Good Practice

When individual or organisational performance increased substantially after a major training programme, the two events appear to be linked. While the change in performance might be related to the training programme, other factors have usually contributed to the improvement as well. As a first stage, all key factors that might have contributed to the performance improvement should be identified. There are then a number of different ways of isolating the effect of the training (table 11). These fall into three groups:

➢ **Control groups**. The most precise method to isolate the impact of training is to use a control group in an experimental design process. This involves comparing the performance scores between two similar groups, only one of which receives the training. True experimental design demands that the members of the experimental group (the group that attends training) are selected randomly, which is often impractical especially in training that is not in-house. Performance of both groups is taken after the training and the difference between the groups taken as the amount of improvement that is directly related to the training programme. There are a number of problems of using control groups, such as selecting the control group, and the possibility of contamination (that those trained teach those who did not attend the course).

➢ **Trend-line analysis and forecasting methods**. A trend-line is drawn using previous performance as a base and extending the line into the future. The performance after training is compared with the trend line. Any improvement in performance above the trend line can be broadly attributed to training. Forecasting

methods take the same general approach, except that they use statistical techniques to produce the projected line.

➢ **Interviews with participants or their colleagues**. Information can be obtained by asking participants and their colleagues about their view of the percentage of the improvement that can be attributed to training. This question should be backed up by probing about the reason for the answer and the respondent's confidence in their estimate. Interviewees should also be asked about other factors that could have contributed to improvements in performance. Before analysis, any extreme responses should be omitted.

Table 11: Methods for isolating the effects of the programme

QUANTITATIVE METHODS
Use of a control group
Trend line analysis of performance data
Use of forecasting methods of performance data
Comparison with previous studies
Calculating the effect of other factors
QUALITATIVE METHODS
Participants' estimate of programme impact
Supervisors' estimate of programme impact
Managers' estimate of programme impact
Subordinates' estimate of programme impact
Customers' estimate of impact

Source: Phillips (1997a)

There are two other issues to consider in evaluating behavioural change:

➢ **Allow time for the change in behaviour to take place**. This could be three to six months.

➢ **Estimate the opportunity for improvement**. Use current differences between high and low performers to estimate the opportunity for gain.

2.5   Evaluation of Organisational Change

2.5.1   Issues

Organisational change generally falls into four categories (Phillips; 1997):

➢ Output increases.

➢ Time savings.

➢ Quality improvements.

➢ Cost savings.

Some areas of training (e.g. developing skills) are easier to measure than others (e.g. changing attitudes or preventing problems). Table 12 lists possible impacts of management training in an arts organisation.

Table 12: Possible organisational effects of management training in an arts organisation

AREAS	POSSIBLE MEASURES
Increase in strategic planning	Planning further ahead Spending more time on long term issues
Increase in artistic quality	More innovation More time for research and development Better reviews
Improvement in staff morale	Higher employee satisfaction Lower absenteeism Lower staff turnover
Improvement in financial position	Funding raised Increase in earned income Increase in sponsorship Broader base of funding
Increase in public benefit	Larger audience More young people involved More socially excluded involved More targeting of disadvantaged More educational work More services to other arts organisations Greater benefit to artists Greater regeneration effects

The main problem of measuring organisational change is to make the link back to training. As Brinkerhoff (1997) explains: "The complexity of the performance equation makes it abundantly clear than any effort to assess the impact of training, where that impact is construed to mean organisational results and benefits, must cope head on with the fact that training – the acquisition of capability through increased competence – is only a partial player, and most likely a bit player at that, in the overall drama of effective performance and business results."

2.5.2   Good Practice

Good practice is to:

➢ **Isolate the effects of the training**. See the previous section. Interpret information with care taking account of other factors beyond the training that could have caused the change.

➢ **Ensure a representative picture**. Ensure the analysis period is long enough to discount temporary fluctuations in performance.

➢ **Narrow the boundaries of analysis**. Consider the impact on a unit of an organisation, such as a team or department, rather than the whole organisation.

2.6   Evaluation of Cost-Effectiveness

2.6.1   Issues

There is an increasing pressure on training functions to measure the cost-effectiveness of their work (Phillips; 1997, 1997a. Campbell; 1994, 1995). There are three broad reasons for this trend:

➢ **To increase the strategic value of training (its effectiveness)**. This is about identifying the major drivers that learning can impact, so that the highest leverage interventions can be selected, managed and measured. Conway (2002) comments that "Too often we focus on learning measurement without considering the business value of learning content."

➢ **To increase the leverage of training (its efficiency)**. This is about maximising the average competency level achieved for the minimum delivery cost per participant.

➢ **To protect training budgets during lean times**. Training has traditionally been seen as a non-urgent function that can be cut without organisational effect. Training is in a stronger position if it reports costs alongside benefits.

➢ **To change the way training is seen**. Figures such as ROI emphasise that training is an investment, not a cost.

The costs associated with training are relatively easy to identify and separate out (table 13):

Table 13: Categories of training costs

Development and programme costs	Fees and expenses of trainers Cost of developing the course, spread over the number of times the course is run Cost of preparing instructional materials, spread over the number of times the course is run Cost of producing instructional materials Cost of purchasing and maintaining or hiring training equipment
Administration costs	Cost of organising the course Cost of any publicity for the course Cost of hire of training space Cost of materials and supplies Cost of purchasing or hiring training equipment, spreading over its life
Trainee costs	Cost of trainees' wages and benefits for the days of preparation and attendance Cost of any cover for employees' work during the training for the days of preparation and attendance Travel, lodging and subsistence costs Manager's time spent on coaching, preparing and assessing
Evaluation costs	Cost of preparing questionnaires Any printing costs for questionnaires Cost of analysing data and questionnaires Cost of disseminating results

There are three broad forms of analysis possible (Campbell; 1994, 1995):

➢ **Cost-benefit ratio**. This compares the annual economic benefits of the training programme to its costs. The advantage of this method is that it does not invite comparison with other categories of investment.

➢ **Return on investment**. This is the rate at which the training pays back its cost. It is calculated as the increase in revenue and the operational savings divided by the full cost of training. This formula is ostensibly the same as that used for other categories of investment.

➢ **Payback period**. The annual cast proceeds (savings) are equated to the original cash outlay required by the investment to arrive at some multiple of cash proceeds equal to the original investment. Measurement is usually in terms of years and months. The payback period is not widely used in investment because it ignores the time value of money.

Table 14: Formulae for financial calculations

CBR =   Programme benefits

          Programme costs

ROI =   Net programme benefits   x 100

          Programme costs

Financial evaluation of training, especially calculation of return on investment, has the following weaknesses:

➢ **Confusion of objectives**. The ROI approach suggests that a positive ROI (a return greater than the costs of training) is equal to success. This makes the erroneous assumption that the purpose of training is inevitably to save money. In practice training could save money in some area of operations, but these might have no strategic or business value.

➢ **Low credibility**. In practice, return on investment is difficult to apply because most benefits from training cannot be accurately quantified in monetary terms.

➢ **Isolation**. Separating out the effect of training is difficult. Furthermore some evaluators would say it is undesirable (Brinkerhoff; 1997), since training works best when integrated with other organisational development activities.

2.6.2   Good Practice

The principles of good practice are:

➢ **Take a conservative approach to calculating both benefits and costs**. Studies on management type training have often produced ROI values of 100-700%, which reduces their credibility.

> **Link analysis to action.** Findings are more valuable as a guide to possible improvements, not as a global judgement about the value of the training. For example, impact might be lower than possible because the wrong people attended the course, people dropped out, or there were barriers to application of knowledge in the workplace.

> **Measure intangible benefits**. Rather than attempting to convert intangible benefits into financial values, it is better to report measures of cost-effectiveness alongside a discussion of the intangible benefits of training.

> **Do not equate the cost-effectiveness of training with the cost-effectiveness of the training function**. Many of the factors affecting application and impact are beyond the control of the training function.

Table 15: Potential intangible benefits

Increased job satisfaction
Increased commitment to the organisation
Lower staff turnover
Improved team work
Greater flexibility of work practices
Improved customer service
Increased personal confidence
Reduced complaints
Reduced conflict
Reduced stress

Based on Phillips (1997) and Simmonds (1998)

3   CONCLUSIONS

Training can have far reaching effects on individuals and organisations. Multi-level methods of evaluation are an attempt to capture the true depth and range of effects. Our guide suggests five levels to evaluation:

> **Reaction**. This measures trainees' views on the training, a version of customer satisfaction. Happy sheets, the most common form of evaluation in training, are a measure of reaction.

> **Learning**. This is a measure of the knowledge acquired, the skills improved or the attitudes changed due to training.

> **Behaviour**. This measures whether the trainee puts the learning into practice in their work.

> **Results**. This looks beyond the individual to see if the training had any effects at the level of the organisation.

> **Cost-effectiveness of training**. This compares the costs and the benefits of training.

Widening evaluation in this way offers many benefits: great understanding of the mechanics and potential effects of learning, greater ability to identify and choose between alternatives in training design, motivation of learners and trainers, and advocacy for the training function.

However, training evaluation is not simple. Isolating the effects of training requires some understanding of the principles of experimental methodology, such as the use of control groups. Furthermore, the technical need to separate training from other organisational improvement programmes should not distract from the need for the two to be intertwined in day to day functioning so that training is serving the strategic needs of the individual and the organisation.

Moving from a simple application of happy sheets to a multi-level approach, as this paper advocates, is time consuming. There are ways of controlling the expense (table 16). At heart, many of these represent an integration of evaluation with training activity. Evaluation flourishes best where it is intrinsic to the functioning of the organisation rather than a bolt-on. The philosophy of evaluation is entirely compatible with that of training. It is a learning process about and within a learning process. These two give a positive cycle of improvement.

Table 16: Cost saving approaches

Focus on key drivers of performance
Build evaluation into the learning process (needs assessment, appraisal etc.)
Plan for evaluation early in the process
Carry out higher levels of evaluation less often and on a sample basis
Share the responsibilities for evaluation
Require participants to conduct the major steps
Use estimates for attribution (data compiled to isolate the effect of the training)
Streamline the reporting process
Systematise evaluation procedures

Based on Burkett (2002)

The principles of good practice are relatively consistent across the different levels and context of training:

➢ **Clarity of purpose**. The criteria for evaluation and the intended use of the data collected must be explicit and consistent.

➢ **Quality control**. Evaluators need to take a critical eye to the data from which they are expected to draw their conclusions.

➢ **Triangulation**. Each evaluation methods have strengths and weaknesses. Using a mix of complementary methods helps to overcome the weaknesses of the methods and gives a more rounded picture.

➢ **Separation of data collection from analysis**. The processes of drawing conclusions and weighting different results need to be made consciously. Each stage in logic should be recorded so that judgements are explained.

➢ **Encouragement of self-evaluation**. It is important to see the trainee as an active participant in the learning process.

➢ **Constructive feedback**. Communicating the results of evaluation is a central element of the work.

REFERENCES AND BIBLIOGRAPHY

Abernathy, Donna J. (1999) "Thinking Outside the Evaluation Box", *Training and Development*, Vol. 53, Issue 2, pp. 18-24.

Athanasou, James, A. (1998) "A Framework for Evaluating the Effectiveness of Technology-Assisted Learning", *Industrial and Commercial Training*, Vol. 30, No. 3, pp. 96-103.

Berry, J. K. (1990) "Linking Management Development to Business Strategies", *Training and Development*, August, pp. 20-21.

Bertnthal, Paul R. (1995) "Evaluation that goes the Distance", *Training and Development*, Vol. 49, Issue 9, pp. 41-46.

Boverie, Patricia, Sanchez Mulcahy, Deanna, and Zondlo, John A. (1995) "Evaluating the Effectiveness of Training Programs", in Pfeiffer, J. W. (Eds.) *Developing Human Resources*. San Diego: Pfeiffer & Company.

Bramley, (2000) *Evaluating Training*. Woodstock: Beekman Publishers.

Brinkerhoff, R. O. (1997) "Clarifying and Directing Impact Evaluation", in Brown, Stephen M. and Seidner, Constance J. (Editors)(1998) *Evaluating Corporate Training: Models and Issues*. Evaluation in Education and Human Services.

Brinkerhoff, R. O. and Brown, V. (1997) *How to Evaluate Training*. Beverly Hills, Ca: Sage.

Brown, Stephen (1997) "Changing Times and Changing Methods of Evaluating Training", km@coproateeducation.biz.

Brown, Stephen M. and Seidner, Constance J. (Eds.)(1998) *Evaluating Corporate Training: Models and Issues*. Evaluation in Education and Human Services.

Burkett, Holly (2002) "ROI on a Shoestring" In *Training Measurement Workbook*, International Quality and Productivity Center.

Burton-Goss, Sadie and Kaska, Michael (1998) "Cultural Dimensions of Evaluation", in Brown, Stephen M. and Seidner, Constance J. (Eds.)(1998) *Evaluating Corporate Training: Models and Issues*. Evaluation in Education and Human Services.

Campbell, Clifton P. (1994) "A Primer on Determining the Cost Effectiveness of Training – Part 1", *Industrial and Commercial Training*, Vol 26, No. 11, pp. 32-38.

Campbell, Clifton P. (1995) "A Primer on Determining the Cost Effectiveness of Training – Part 2", *Industrial and Commercial Training*, Vol 27, No. 1, pp. 17-25.

Campbell, Dorothy M., Melenyzer, Beverly J., Nettles, Diane, H. and Wyman, Richard M. (2000) *Portfolio and Performance Assessment in Teacher Education*. London: Allyn and Bacon.

Conway, Mal (2002) "Training Results Measurement: A Methodology for Measuring and Driving Business Results from Learning". In *Training Measurement Workbook*, International Quality and Productivity Center.

Costa, Arthur L. and Kallick, Bena (1995) *Assessment in the Learning Organization: Shifting the Paradigm*. Alexandria, Virginia: Association for Supervision and Curriculum Development.

Dawson, Ruth (1995) "Fill This In Before You Go: Under-Utilized Evaluation Sheets", *Journal of European Industrial Training*, Vol. 19, No. 2, pp.3-7.

Dyer, Steven (1994) "Kirkpatrick's Mirror", *Journal of European Industrial Training*, Vol. 18, No. 5, pp.31-32.

Foxon, Marguerite (1989) "Evaluation of Training and Development Programmes: a Review of the Literature", *Australian Journal of Educational Technology*, Issue 5, No. 2, pp. 89-104.

Geertshuis, Susan, Holmes, Mary, Geertshuis, Harry, Clancy, David, and Bristol, Amanda (2002) "Evaluation of Workplace Learning*", Journal of Workplace Learning*, Vol. 14, No. 1, pp. 11-18.

Gibb, Stephen (1994) "Evaluating Mentoring", *Education + Training*, Vol. 36, No. 5, pp. 32-39.

Gilleard, Jenni (1997) "Empowering Self-Study Learning Through Evaluation", *Empowerment in Organizations*, Vol. 5, No. 1, pp. 46-53.

Guskey, Thomas R. (2000) *Evaluating Professional Development*. Thousand Oaks, California: Corwin Press.

International Quality and Productivity Center (2002) *Training Measurement Conference Workbook*.

Jones, John E. (1999) "Don't Smile About Smile Sheets", *Training and Development Journal*, December issue.

Kimball, Lisa (1998) "Easier Evaluation with Web-based Tools", Training and Development, Vol. 52, Issue 4, pp. 54-56.

Kirkpatrick, Donald L. (1994) *Evaluating Training Programs. The Four Levels*. Berrett-Koehler Publishers.

Kirkpatrick, Donald L. (1998) *Another Look at Evaluating Training Programmes. Fifty Articles Reprinted from Training Magazines*. ASTD, Alexandria.

Lewis, Philip and Thornhill, Adrian (1994) "The Evaluation of Training: An Organisational Culture Approach", *Journal of European Industrial Training*, Vol. 18, No. 8, pp. 25-32.

Lyons, N. (1998) *With Portfolio in Hand: Validating the New Teacher Professionalism*. New York: Teacher College Press.

Mann, Sandi and Robertson, Ivan T. (1996) "What Should Training Evaluations Evaluate?" *Journal of European Industrial Training*, Vol. 20, No. 9, pp. 14-20.

Marsden, Jane M. (1991) "Evaluation: Towards a Definition and Statement of Purpose", *Australian Journal of Educational Technology*, Issue 7, No. 1, pp. 31-38.

Martin-Kneip, Giselle O. (1999) *Capturing the Wisdom of Practice. Professional Portfolios for Educators*. Alexandria, Virginia: Association for Supervision and Curriculum Development.

McClelland, Sam (1994a) "A Model for Designing Objective-Oriented Training Evaluations", *Industrial and Commercial Training*, Vol. 26, No. 1, pp. 3-9.

McClelland, Sam (1994b) "Training Needs Assessment Data-Gathering Methods: Part One, Survey Questionnaires", *Journal of European Industrial Training*, Vol. 18, No. 1, pp.22-26.

McEvoy, Glenn M. and Buller, Paul F. (1990) "Five Uneasy Pieces in the Training Evaluation Puzzle", in Kirkpatrick, Donald L. (1998) *Another Look at Evaluating Training Programmes. Fifty Articles Reprinted from Training Magazines*. ASTD, Alexandria.

McGee, Phillip (2002) *The Productivity Model: A Conceptual Framework*. At www.internetraining.com/Productivity/Productivity.htm

Medsker, Karen and Roberts, Don (1992) "Evaluating Training: There is no "Cookbook" Approach, *ASTD Tool Kit*. http://www.orau.gov/pbm/training.html

Michalski, Greg V. and Cousins, J. Bradley (2001) "Multiple Perspectives on Training Evaluation: Probing Stakeholder Perceptions in a Global Network Development Firm", *The American Journal of Evaluation*, Vol. 22, No. 1, pp. 37-54.

Mulder, Martin (2001) "Customer Satisfaction with Training Programs", *Journal of European Industrial Training*, Vol. 25, No. 6, pp. 321-331.

Oakley, Karla (1998) "The Performance Assessment System: a Portfolio Assessment Model for Evaluating Beginning Teachers." *Journal of Personnel Evaluation in Education*, Vol. 11, pp. 323-341.

Peterson, Kenneth D. (2000) *Teacher Evaluation: A Comprehensive Guide to New Directions and Practices*. Thousand Oaks: Corwin.

Phillips, Jack J. (1997a) *Return on Investment in Training and Performance Improvement Programs*. Woburn: Butterworth Heinemann.

Phillips, Jack. J. (1997) *Handbook of Training Evaluation and Measurement Methods*, Third Edition. Houston: Gulf Publishing Company.

Rae, Leslie (1999) *Using Evaluation in Training and Development*. Kogan Page.

Redshaw, Bernard (2000) "Evaluating Organisational Effectiveness", *Industrial and Commercial Training*, Vol. 32, No. 7, pp. 245-248.

Sackett, Paul, R. and Mullen, Ellen J. (1993) "Beyond Formal Experimental Design: Towards an Expanded View of the Training Evaluation Process", *Personnel Psychology*, Vol. 46, Issue 3, pp. 613-628.

Sadri, Golnaz and Snyder, Peggy J. (1995) "Methodological Issues in Assessing Training Effectiveness*", Journal of Managerial Psychology*, Vol. 10, No. 4, pp. 30-32.

Scriven, Michael (1991) *Evaluation Thesaurus*. Fourth Edition. London: Sage.

Shapiro, Lester T. (1995) *Training Effectiveness Handbook*. McGraw-Hill.

Shulman, L. (1998) "Teacher Portfolios: A Theoretical Activity." in Lyons, N. (1998) *With Portfolio in Hand: Validating the New Teacher Professionalism*. New York: Teacher College Press.

Simmonds, David (2001) *The Evaluation Toolkit*, Ely: Fenman.

Sims, Ronald, R. (1993) "Evaluating Public Sector Training Programs", *Public Personnel Management*, Vol. 22, Issue 4, pp. 591-616.

Sullivan, Richard L., Wircenski, Jerry L. and Major, Mary Jo (1993) "Analysing Knowledge-based Tests", in Kirkpatrick, Donald L. (1998) *Another Look at Evaluating Training Programmes. Fifty Articles Reprinted from Training Magazines*. ASTD, Alexandria.

Warr, Peter, and Allan, Catriona (1999) "Predicting Three Levels of Training Outcomes", *Journal of Occupational and Organizational Psychology*", Vol. 72, Issue 3, pp. 351-376.

Yang, Hyuckseung, and Sackett, Paul R. (1996) "Statistical Power and Cost in Training Evaluation: Some New Considerations", *Personnel Psychology*, Vol. 49, Issue 3, pp. 651-669.