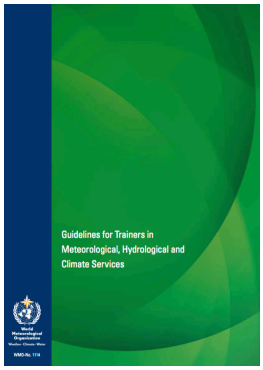
Chapter 8: Learning Evaluation



NOTE: This is an extract of the document “Guidelines for Trainers in Meteorological, Hydrological and Climate Services” (WMO-No. 1114).

To read the full document, please access:

WMO-No 1114

http://www.wmo.int/pages/prog/dra/documents/wmo_1114_en.pdf

8. LEARNING EVALUATION

Competence VI: Assess learning and evaluate the learning process

Competency description
Learning is assessed against the required learning outcomes, and training activities, events and programmes are monitored and evaluated to improve learning processes.

Performance criteria
- Provide clear assessment policies and guidelines;
- Use formative assessment to promote deep learning;
- Assess learning against specified performance outcomes;
- Collate, analyse and use data on all aspects of the training;
- Evaluate training at the required level using established models;
- Improve training on the basis of the results of the evaluation.

Knowledge requirements
To be able to understand, explain and/or critically evaluate:
- Purposes and characteristics of assessment and evaluation;
- Advantages and limitations of assessment strategies;
- How to design good assessment items;
- Application of the Kirkpatrick Model and/or other models of evaluation;
- How to conduct the evaluation process;
- How to evaluate a training department and training programmes.

Personnel who should demonstrate this competency
- Training managers and senior trainers;
- Trainers involved in assessment procedures;
- Managers of staff identified as having a learning need;
- Human resources personnel.

8.1 Introduction

To find out whether it has benefited from investing in a learning activity, an organization needs to evaluate whether the knowledge, skills or behaviour of the participants have changed in a way that enhances job performance or career development. The evaluation process should extend beyond the trainers and learners to involve also the stakeholders, who must support the learning enterprise. Enhancing the learning process should be a shared responsibility.

8.2 Assessment and evaluation

The concepts of assessment and evaluation tend to overlap and, in some languages, there is no difference between them. In the context of learning activities, the terms will be used to mean the following:

– *Assessment*: a way of measuring what individuals have learnt as a result of a learning activity – for example, by testing a learner's knowledge, skill or behaviour – or of determining current knowledge or competency. The assessment should be based upon the objectives of the learning situation or job requirements;

– *Evaluation*: a way of measuring the worth of providing a learning opportunity – for example, by judging whether the learning opportunity met its objectives, made a difference to the organization or was good value for money – with the aim of improving the learning process.

Though there are differences between assessment and evaluation, they are intimately linked because a full evaluation has to include information on what the individuals have learnt.

8.2.1 ***Assessment: basic concepts***

Assessments can be used for a variety of purposes, including recruitment, identifying learning needs, performance management and certifying competency.

An assessment may be categorized according to when it occurs in relation to a learning event:

– *Initial assessment*: identifying the existing knowledge and skills of individuals to decide which type of learning would be most suitable for them. This also provides a basis for assessing the resulting gain in knowledge and skills;

– *Formative assessment*: gathering information during training to provide feedback to both learners and trainers about progress and to identify any alterations that might be required. A formative assessment allows the trainer and learner to monitor progress, thus reducing the chance of failure at the final (i.e. summative) stage;

– *Summative assessment*: determining the knowledge and skills acquired, in terms of the learning outcomes, or the gain in learning since the initial assessment.

An assessment may also be categorized according to an external standard:

– *Criterion-referenced assessment*: compares the learning of individuals with a standard, such as job competencies, or specified learning outcomes (an absolute measure);

– *Norm-referenced assessment*: compares the learning of individuals and ranks the learners (a relative measure). A norm-referenced assessment only works with a large number of learners.

Assessments are used as a basis for making judgements, some of which may be crucial for the learners being assessed; for example, they might control career progression or have a significant influence on job satisfaction. They therefore need to be reliable and valid:

– *Reliability*: the extent to which an assessment process yields consistent results each time it is used in similar circumstances;

– *Validity*: the extent to which an assessment process measures what it claims to measure. Internal validation is concerned with assessing whether learners have met the required learning outcomes; external validation determines with whether the learning outcomes were based on an accurate assessment of the learning needs.

Assessment processes also need to be administratively workable and consistent with any professional, functional and status-related expectations (see Figure 8.1).

Consideration will now be given to the evaluation process. Ways of carrying out assessments will be described in the context of evaluating learning activities.

8.2.2 ***Purpose of evaluation***

Evaluation is an important part of the learning cycle. It can provide information about:

– Ways of improving the delivery of learning activities by asking questions such as “were the learning methods appropriate” and “was the content relevant?”;

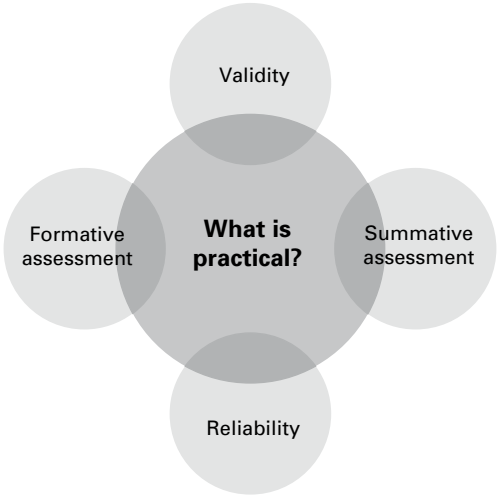


Figure 8.1. The type of assessment chosen and the need for reliability and validity will determine what is practical.

– Whether the organization benefited from its investment by asking questions such as "were learning outcomes satisfied?" and "has job performance improved?".

A full and frank evaluation improves learning activities, engages learners and their line managers, enhances the credibility of training, and provides evidence of the value of learning events and programmes for the organization.

According to Mark Easterby-Smith,[14] the purpose of evaluation can be specified as follows:

– *Proving*: demonstrating to stakeholders that positive outcomes have resulted from training and development;

– *Improving*: identifying how training processes can be improved;

– *Learning*: providing feedback to support individual learning;

– *Controlling*: ensuring that training is being delivered according to agreed procedures and requirements.

Evaluation is only effective if the results are communicated to all the stakeholders and any concerns are promptly acted upon.

8.3 The Kirkpatrick model

A widely used approach to evaluation is the Kirkpatrick model which proposes four levels of evaluation (see Figure 8.2):

– *Level 1: Reaction*. Are learners satisfied with the learning activities?

– *Level 2: Learning*. Have the required changes in knowledge, skills and behaviour been achieved?

– *Level 3: Transfer*. Have the knowledge, skills and behaviour acquired through the learning activities improved job performance?

---

[14] Easterby-Smith M., 1994: *Evaluating Management Development, Training and Education*. Brookfield, Vermont, USA, Gower.

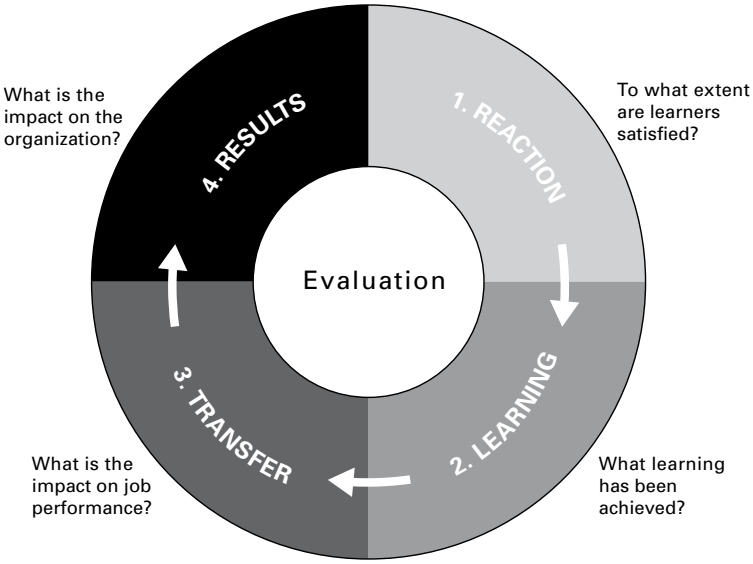


Figure 8.2. The four levels of evaluation of the Kirkpatrick model

– *Level 4: Results*. Has the learning had an impact on the performance of the organization?

An organization can use various techniques to evaluate these four levels. An evaluation process might incorporate techniques that provide information about several levels at once.

Level 1: Reaction

Finding out about the satisfaction of learners gives the trainer feedback on how to improve the learning process, administration, facilities and domestic arrangements. Seeking their views, gives learners a role in the learning process and in making it better.

Feedback may not reflect the true feelings of some learners: they may, for example, give a high rating to avoid offending anyone, or a low rating because they were unwilling participants. Ratings can be subjective and influenced by a variety of factors, for example, an opinion about one part of an activity can rub off onto others.

Reaction questionnaires (sometimes referred to as "happy sheets") are the most common tool for gaining information, especially because they are easy to prepare. They often cover:

– *Content*: was it relevant? Were there omissions or any unrelated material?

– *Learning methods and materials*: were the learning methods and materials suitable?

– *Length, structure and pace*: were the length, structure and pace of the learning event okay?

– *Learning outcomes*: were the learning outcomes clearly defined and satisfied?

– *Trainer's skills*: did the trainers have the required knowledge and skills?

– *Learning transfer*: how much of the learning will be used in the workplace?

– *Facilities and domestic arrangements*: were these satisfactory?

– *Administrative arrangements*: did participants get the information they needed?

In addition, questionnaires usually include an overall rating for the activity. Box 8.1 outlines the main types of question that might be used in questionnaires.

Box 8.1. Questionnaires

Unjustified questions
(a) The session was:

Useful								
False	1	2	3	4	5	6	7	True
Interesting								
False	1	2	3	4	5	6	7	True

(b) I found the session to be:

Useless	1	2	3	4	5	6	7	Useful
Uninteresting	1	2	3	4	5	6	7	Interesting

(c) How useful was the session? Useless, not very useful, useful, very useful, extremely useful (circle one).

Justified questions
As a result of the session:

I learnt little	1	2	3	4	5	6	7	I learnt a lot

If you chose 4, 5, 6 or 7, please state why you gave that rating. If you chose 1, 2 or 3, please state what aspects you did learn.

Open questions
Which part of the session did you find most useful? Why?
Which part of the session did you find least useful? Why?

Blank sheet
Please provide comments about the session or anything related to it. Be as open as you can, and be specific.

It may be worthwhile having separate questionnaires on the learning aspects and on the organization of the event, that is, facilities, domestic and administrative arrangements. Separate questionnaires may help learners differentiate more easily between their levels of satisfaction with those two aspects.

Questionnaires should be easy to understand, attractive in appearance and short. They should include space for comments and suggestions to augment the data from ticked boxes or ratings.

Web-based questionnaires are becoming more common. They provide easily adaptable templates for questionnaires and analysis of results that is mostly automatic.

For a training event or programme that lasts a long time, such as an initial forecasting course, it might be worth issuing questionnaires throughout as well as at the end of the event. The final questionnaire could be distributed a day or so before the end so that the results can be analysed and discussed with the learners before everyone departs. This would clearly demonstrate that the views of the learners are valued.

Apart from questionnaires, information about learners' reaction can be gathered through a general critique session towards the end of a learning event. This allows all participants to put forward their views and the trainer to respond to any comments or suggestions. There is, however, a danger that the session becomes dominated by a few people or just focuses on areas of discontent. Blogs maintained by learners during the activity can also provide useful feedback throughout an event.

Another way of collecting information is to interview – either by telephone/teleconference or face-to-face – some of the learners or a focus group after the activity. The interviewer would use open questions similar to those in the questionnaire and also probing questions.

Once the information has been gathered, the evaluator needs to analyse the results and ensure action is taken in response to the feedback, if necessary. Note that a high level of satisfaction is no guarantee that the required learning has taken place.

Level 2: Learning

The next step is to find out if changes in knowledge, skills and behaviour have been achieved. Recall that jobs require the following kinds of learning:

– *Knowledge*: the information and understanding someone needs to perform the job;

– *Skill*: what someone has to be able to do on the job;

– *Behaviour*: how people should conduct themselves on the job.

These should be evaluated differently. It is also worth remembering that knowledge and skills requirements can each cover a broad spectrum. For example:

– Knowledge can be recalled, applied in new situations or used to solve problems;

– Skills can be rehearsed, applied in a new context or used in a creative way.

When developing ways of assessing learning, it is necessary to decide what is relevant and to ensure that the assessment covers either all or a reasonable selection of the learning outcomes.

Some of the issues that should be considered when assessing learning are listed in Box 8.2.

Box 8.2. What trainers should ask about any assessment

– What level of assessment is really required?
– What do I want to know and what will I do with it?
– How much time and effort will be involved? Is it worth that effort?
– Who is the information for and why do they need it?
– Am I using the correct kind of assessment?
– Are my questions or instructions simple, direct and unambiguous?
– Do my questions cover an appropriate range of topics?
– Has the assessment been peer-reviewed?

Knowledge

The level of knowledge is usually determined by the use of tests. These fall into two broad categories:

– *Objective tests* contain yes/no or multi-choice questions that test knowledge of facts, regulations and procedures most easily (see Box 8.3). But objective questions can also be designed to measure aspects of learning at higher levels of Bloom's taxonomy (see Table 4.1), such as application, analysis and evaluation, though this is not as straightforward as assessing knowledge. Objective tests can often be web-based;

– *Subjective tests* include open questions, essays, oral questioning, interviews or case studies, which test knowledge and application of complex concepts. Marking these tests is, however, time consuming even if there is a well-defined marking scheme.

Box 8.3. Examples of objective tests of knowledge and skill

Multiple-choice questions
Which parameter is conserved during unsaturated adiabatic ascent?
(a) Dew point temperature
(b) Potential temperature
(c) Temperature
(d) Relative humidity

Which parameters are conserved during unsaturated adiabatic ascent?
(a) Dew point temperature
(b) Potential temperature
(c) Temperature
(d) Relative humidity

Analyse the accompanying satellite image and Numerical Weather Prediction (NWP) product and select those areas of highest concern for heavy precipitation. (Select all that apply)
(a) Area A
(b) Area B
(c) Area C
(d) Area D
(e) Area E

Advice: use simple, clear language, avoid negatives and words that give a lead, ensure that choices are of a similar length, vary the position of the correct answer and keep questions independent.

True/false and binary questions
The humidity mixing ratio is conserved during unsaturated adiabatic ascent.

True False

Is the humidity mixing ratio conserved during unsaturated adiabatic ascent?

Yes No

Advice: use simple, clear language, avoid negatives and words that give a lead, and keep questions independent.

Questions requiring short or selected answers
During unsaturated adiabatic ascent……is conserved.

Snow cover can inhibit fog when the boundary layer moisture is … … … (deep, shallow, or mixed) and radiative processes are … … … (dominant, negligent, or absent).

Advice: keep the blank space towards the end of the question or statement, leave the same space for each answer and, for a numerical answer, indicate the required units.

Skills

The acquisition of skills is usually best assessed through exercises, direct observation or role playing. For example, complex skills associated with synoptic analysis and forecasting could be assessed using exercises based on case studies or real-time data (possibly employing the workstation used by forecasters). Alternatively, a trainer could watch someone carry out a task and ask questions about what is being done and why. A similar approach could be used to assess meteorological observing skills. These examples illustrate that it may be desirable to

combine the assessment of skills, behaviour and underlying knowledge to achieve higher assessment validity.

The assessment of complex skills tends to be time consuming because it is most effective when carried out on a one-to-one basis. Assessors should record evidence of skills as they are demonstrated to ensure some objectivity and to give immediate feedback.

Behaviour

Assessing behaviour is much more difficult than assessing knowledge and skill. The workplace, or a simulation of the workplace, is probably the best location for directly observing and assessing behaviour.

In some cases, role playing can be used, but it needs to be carefully prepared and managed. To observe how someone behaves naturally, the trainer should not leave the participants time to think about what behaviour they are expected to display, so that they are less likely to modify their natural inclinations. Developing a good role-play exercise or simulation is time consuming and resource-intensive.

Another way of assessing behaviour is to use a survey designed to measure likely responses to situations, asking participants to complete it before and after the learning event. Alternatively, interviews could be used. In both cases, however, the assessment is based on a hypothetical situation and written or verbal response rather than observed behaviour in work situations.

Level 3: Transfer

Unfortunately, just because something has been learnt it does not mean that it will be used to improve performance on the job. Organizations should, therefore, assess whether learning is being put into practice. Some of the actions that support the transfer of learning into enhanced job performance are given in Box 8.4.

There are, however, other factors, apart from the level of newly acquired knowledge, skills and behaviour, which can influence the transfer of learning into improved performance:

– *Support*: the level of support, for example, through coaching given by the line manager and the learner's colleagues;

– *Practice*: the time available to practice new skills and behaviour, and to reflect on and assimilate newly acquired knowledge;

– *Culture*: the expectations of the organization and rewards offered for improved performance;

– *Infrastructure*: the availability of information technology or other infrastructural support.

Despite these complicating factors, the assessment of job performance provides valuable information about the overall impact of a learning activity.

There are basically three ways of assessing the impact of a learning activity on job performance:

– *Direct observation*, which could be carried out by the line manager or an external assessor, preferably using a structured form to record evidence. This approach is similar to the one used to assess skills. The assessment could be treated as part of the normal performance management process;

– *Questionnaires*, which could be completed by the learner, the learner's manager or both. They could contain questions about the application on the job of the newly acquired knowledge, skills and behaviour, and related issues, such as their utility or barriers to their use that have hindered job performance. Questionnaires could be part of an organization's routine performance management process;

– *Interviews*, which would normally be face-to-face or over the telephone, with someone from training doing the interviewing and recording the results. Both learners and managers could be interviewed.

Whichever approach is taken, there is a need to collate the results and draw conclusions. For example, if the impact of learning on performance is low for most participants but they have achieved all the learning outcomes, it is likely that the learning outcomes were not suitable and, therefore, have to be revised. The initial learning needs analysis might be flawed.

Box 8.4. Transfer of learning

Here are some actions that support the transfer of learning into enhanced job performance:
- Pre-learning briefing: the learner and the manager discuss what is involved in the learning event;
- Post-learning briefing: the learner and the manager discuss whether the learning outcomes have been satisfied, and agree on how to use the newly-acquired knowledge, skills and behaviour on the job;
- Support: the learner's manager and other colleagues support the implementation of the new learning;
- Feedback: the learner's manager provides regular feedback about progress;
- Final review: after an agreed period of time, the learner and the manager meet to review the impact of the learning activity on job performance and agree on any further actions.

Level 4: Results

Some organizations already have performance indicators in place for assessing the benefits of an investment in learning. For example, some NMHSs may commission customer surveys that can quantify, at least in part, the impact of a learning activity.

National Meteorological and Hydrological Services may have verification statistics that can measure the impact of a learning activity designed to improve a forecast service (such as warnings of heavy rainfall) or to introduce a new observing system (such as the introduction of Doppler radars to improve the prediction of tornadoes). However, the measure might be biased unless there is a way of comparing forecasters who have and those who have not completed the learning activity, and of considering those who work in very different forecast regimes.

Assessing the impact of specific learning activities on the organization is the ultimate way of evaluating those activities, but it is also the hardest. The United Nations has established standards for evaluation (see Box 8.5). These include standards for conducting evaluations, which cover planning, design, implementation and follow-up. If all stages have worked correctly, the assessment of the impact of the training can be trusted.

8.4 Evaluation procedure

Organizations need to decide what they want to get out of the evaluation process. They might ask questions such as:

– Has participation in a train-the-trainer course improved trainers' performance?

Box 8.5. Standards for evaluation in the United Nations system

The United Nations Evaluation Group has defined norms and standards which aim to the professionalization of the evaluation function. They also provide guidance to evaluation offices in preparing their evaluation policies or dealing with other aspects of their operations. The standards cover:
- Institutional framework and management of the evaluation function;
- Competencies and ethics;
- Conduct of evaluations;
- Evaluation reports.

The standards indicate that all those engaged in designing, conducting and managing evaluation activities should aspire to high-quality and ethical work. This means that they should possess core evaluation competencies. Consequently evaluators should:
- Have relevant educational background, and qualification and training in evaluation;
- Have professional work experience relevant to evaluation;
- Have specific technical knowledge of, and be familiar with, the methodology or approach that will be needed for the specific evaluation to be undertaken, as well as certain managerial and personal skills;
- Be sensitive to beliefs, manners and customs and act with integrity and honesty in their relationships with all stakeholders;
- Ensure that their contacts with individuals are characterized by respect;
- Protect the anonymity and confidentiality of information provided by individuals;
- Take responsibility for their performance and their product(s).

For more information on this topic go to: http://www.uneval.org/.

- Did the new learning activities in support of continuing professional development (CPD) meet their objectives?
- Has replacing conventional training courses with distance learning worked?
- Is the on-the-job training being done effectively?
- Should an external organization run some learning events?
- Is training providing value for money?

When conducting an evaluation it is worth going through the steps illustrated in Figure 8.3.

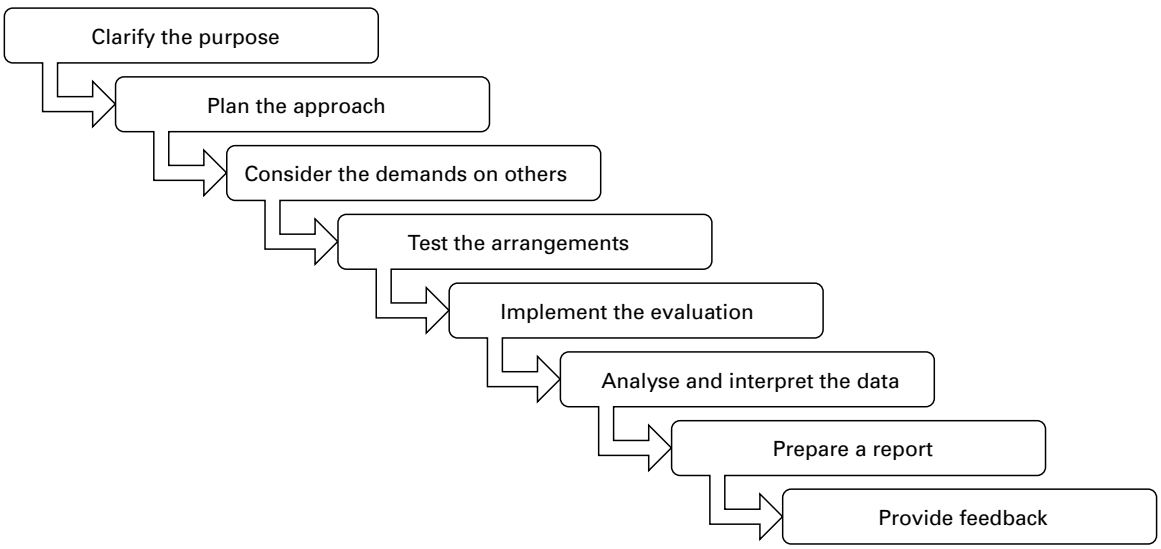


Figure 8.3. The key steps of an evaluation. The emphasis placed on each step will depend upon what an organization expects from the evaluation process.

Now consider in more detail the steps identified in Figure 8.3:

– *Clarify the purpose of the evaluation*: ensure that the person commissioning the evaluation and those carrying it out agree on its purpose and the deliverables. Identify all stakeholders in the evaluation and their needs as well;

– *Plan the approach*: decide what resources are available to do the evaluation, what data is required and how to get the data;

– *Consider the demands on others*: ensure that the size and scope of the evaluation matches the learning activities, and that it places reasonable demands on respondents and those analysing the information;

– *Test the arrangements*: check the evaluation before rolling it out. A small evaluation might need just someone to check whether a questionnaire is logical and understandable. For a more wide-ranging evaluation, it may be necessary to pilot the whole process;

– *Implement the evaluation*: carry it out, monitoring whether there are any problems and adjusting it as required. Do not draw early conclusions and take care with sensitive data;

– *Analyse and interpret the data*: use the data collected to answer the original questions;

– *Prepare a report*: use the analysis to prepare a report that covers the scope and purpose of the evaluation, the methods used, findings and recommendations;

– *Provide feedback*: give feedback about the outcome of the evaluation to all those who have contributed to it.

The evaluation report and recommendations are a key part of the process. The recommendations might be supported by:

– *Quantitative information*: statistical information that can be used to provide an overview of the evaluation results and make comparisons with similar events;

– *Qualitative information*: comments, observations, suggestions and recommendations from learners and other stakeholders.

To get the recommendations implemented, make sure that the report contains only information that is relevant to the recipients and that the recommendations are clear and concise.

Ideally an evaluation should be built into the design of a learning event rather than treated as an add-on. An evaluation is only worthwhile if there is commitment to act on its findings. Otherwise the cost of the evaluation will outweigh its benefits.

8.5 Evaluating training departments

Organizations may wish to evaluate their training department periodically. The question could be: "do the benefits provided by the training department outweigh the costs?"

Such questions can be answered by using a cost/benefit analysis or an investment appraisal which depend upon:

– The cost of the training department – for example, for staff and facilities;

– The benefits in monetary terms of the training department – for example, cost savings, increase in productivity and more income generation.

Usually the costs are relatively easy to identify, but putting the benefits in monetary terms is much more difficult and inevitably subjective.

An alternative approach is to use benchmarking based on a comparison of costs and benefits of the training department with that of another organization. The following gives an indication of the type of information that could be used:

– The proportion of employees participating in learning activities as part of their CPD;

– The average number of days allocated yearly to learning activities per employee;

– The proportion of people working in the training department compared to the total number of employees in the organization;

– The time taken to train new recruits;

– The expenditure on learning activities per employee.

Organizations vary in the way they put together such information. For example, in some organizations expenditure on training activities might cover only formal learning activities whereas in others all types of learning activity are included. Organizations might also vary in the way they account for the cost of learners. Comparisons should only be made when the data provided by another organization is compatible.

8.6 Next step

The information and insights gained from evaluations should be used at the beginning of the next learning cycle, when the process of developing training policies, processes and procedures, and identifying learning needs starts again.

8.7 You and your organization

In order to consolidate the material presented in this chapter, try answering the following questions:

– Why is it important to evaluate training and to what extent is the evaluation of training considered important in your organization?

– In the past, how have you modified your training processes based on evaluation results?

– To what extent are line managers in your organization involved in the evaluation process?

– How is reaction to training assessed in your organization and what processes are there to act upon the resulting information?

– What techniques are used in your organization to assess changes in knowledge, skills and behaviour resulting from training?

– In your organization, to what extent are changes in job behaviour assessed as a result of training?

– To what extent does your organization have a systematic approach to designing, implementing, analysing, reporting and taking action on the evaluation of training?