Part Four – Assessment of Learning

Assessment is a major concern for those who learn, those who teach and those who are responsible for developing courses or training programmes.

- 'How will I be assessed?'; 'Are there exams?'; 'What do I have to do to pass?' are some of the questions students ask before they even start.
- 'How well are my students getting on?', 'Can my trainees really understand the subject and demonstrate the skills?' we ask ourselves as teachers.
- 'How can we be sure that the learners are achieving what the course sets out to do?' ask the programme leaders and curriculum developers.

We are all involved in making judgements about attainment or performance at some time in our professional and personal lives – either as an assessor or as the person who is assessed. Employers, colleagues and clients assess how we do our work; friends, family and magistrates may assess our behaviour; doctors assess our health and teachers assess our learning. As Rowntree (1987: xii) notes:

> *… we spend our lives assessing others, trying to know them and explain them to ourselves – and often influencing them by our consequent decisions. And even in death we cannot escape the assessors – obituary writers for the famous; just family, workmates and friends for the rest of us.*

Assessment plays an important part in the teaching and learning process at all levels in education. Rowntree (1987: 1) suggests that:

> *If we wish to discover the truth about an educational system, we must look into its assessment procedures. What student qualities and achievements are actively valued and rewarded by the system? How are its purposes and intentions realized? To what extent are the hopes and ideals, aims and objectives professed by the system ever truly perceived, valued and striven for by those who make their way within it? The answers to such questions are to be found in what the system requires students to **do** in order to survive and prosper. The spirit and style of student assessment defines the de facto curriculum.*

There have been considerable changes in the purpose and practice of assessment over the last few decades. Huddleston & Unwin (1997: 111) note that:

> *Assessment is now not so much something which is 'done unto' students but which often involves negotiations with students and sometimes with employers as well.*

So in this Part we shall attempt to answer the following questions:

- What is the purpose of assessing learners?
- What is it that we assess?
- What types of assessment methods do we use?
- Who assesses learners?
- What can we do to try to ensure that assessment is fair?
- How do we go about assessing?

6. Basic Issues

We will first discuss why assessment is important, and try to put it into a social, educational and political context. We will then go on to define what we mean by assessment.

The context

Student learning is very often assessment-driven. There is much research evidence that most learners tend to work harder for those aspects of a programme which are being assessed. This means that assessment can determine, to a certain extent, what students learn. However, assessment should be an integral part of the teaching and learning process. It should not be the main focus for learning; nor should it be just an afterthought or simply a test. Assessment provides learners and teachers with feedback and this feedback can be used to improve learning. Assessment is basic to learning and assessment methods should be integrated into a programme of study or training course when it is designed – as reflected in Figure 14 – rather than 'bolted on'.

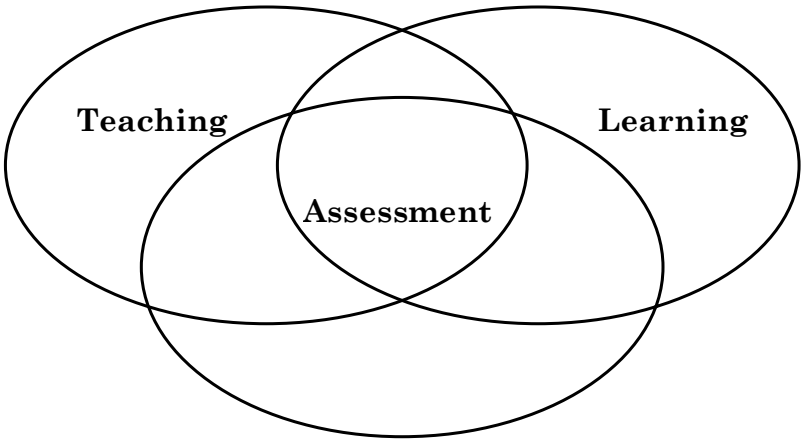

Figure 14 Assessment is an integral part of teaching and learning

(Brown & Knight, 1994: 46)

For this reason, teachers and trainers have always been aware of the importance of assessment. In recent years, however, assessment has also become a political issue. Increasingly, teachers are being required to demonstrate that students have profited from their learning. Organisations involved in teaching or training are expected to prove their 'success' with groups of learners. It is part of a changing culture, particularly since the early 1990s, which has placed an emphasis both on accountability and on a market model of education and training. Many organisations involved in training have externally set targets and, in some cases, funding is linked to the extent to which those externally set targets are met. League tables, which rank examination achievements, are published in the press. They are used, if

appropriate, in organisations' promotional materials. They are also used as evidence to counter or to support arguments about education and educational standards.

For state-funded provision, there has been growing pressure to demonstrate the value of a seemingly large financial investment, to improve 'results', to meet given targets and to raise standards. For private training providers, of course, good achievement and success rates are necessary for commercial survival.

Over the years there have been trends in assessment. Perhaps the most significant recent shift has been towards an emphasis on demonstrating and assessing *competence*, often within the workplace, resulting from the introduction of new types of vocational qualifications in the late 1980s. There has also been a move towards assessing students' transferable, personal skills as well as the content of what they are studying.

Changes in assessment methods can be controversial. The contribution of continuous assessment towards final grades, for example, remains an issue for debate. It becomes topical each year when school, college or university exam results are announced. If results are higher than in the previous year, some will argue that standards must have gone down. They may conclude that the tasks, or the method of assessment, must be less rigorous. Others will argue that teaching has improved and that standards are rising in the same way as in sport, where people are now routinely achieving standards that seemed out of reach a decade ago. These issues will be explored further in later sections.

The meaning of assessment

Assessment can be described most simply as a set of devices – or tools – which are employed to measure learner achievement.

We need to be clear that 'assessment' is not the same as 'evaluation', although some authors use these two words interchangeably. We have taken evaluation to represent a broader concept than assessment. As we have seen in Section 5, teachers will often evaluate a session they teach, after the event. This will involve asking themselves questions such as: 'How did it go?', 'What could I do better next time?'

Teachers also evaluate whole courses or programmes. Often, learners as well as teachers are involved in evaluations. Students may, for example, be asked to complete questionnaires, or to participate in committee or team meetings. An evaluation of a particular training programme usually results in minor changes which are then implemented in time for the next intake of students. But an evaluation often begins by examining students' achievements. If these achievements are not satisfactory then the evaluation may result in changes being made to assessment strategies, to the teaching methods or even to the curriculum content itself.

We shall follow the practice of most educational writers and distinguish between the two terms. 'Assessment', then, is taken to refer to the tools used to measure learner achievement.

Assessment, as Rowntree (1987: 4) suggests:

> *… can be thought of as occurring whenever one person, in some kind of interaction, direct or indirect, with another, is conscious of obtaining and interpreting information about the knowledge and understanding, or abilities and attitudes, of that other person... In education we are mainly conscious of this 'encounter' in the shape of teachers finding out about their students.*

Assessment has been further defined as:

> *a systematic basis for making inferences about the learning and development of students ... the process of defining, selecting, designing, collecting, analyzing, interpreting and using information to increase students' learning and development.*
>
> (Erwin, 1992: 15, quoted in Brown & Knight, 1994: 12)

7. Purposes of Assessment

Assessment is intended primarily to benefit the learner. However, many people, in addition to the learner, have an interest or a 'stake' in the outcome of someone else's learning. These stakeholders may include, for example, potential learners, working colleagues, mentors, tutors, sponsors, institutional managers, 'competing' institutions, employers and funding bodies.

Why do we need to assess?

Assessment is used for a number of purposes, among them to:

- measure the relationship between the teachers' aims and the students' output
- test the progress of students
- diagnose particular weaknesses or highlight strengths
- provide feedback to learners, leading to future improvement
- provide feedback to teachers and trainers
- provide feedback to other stakeholders, as described above
- select students for courses of study or employment
- predict future achievement
- estimate learners' current skills
- form part of a student's profile of abilities
- contribute to some publicly recognised accreditation system
- recognise prior achievement and experience and possibly to lead to credit accumulation and transfer
- demonstrate to students that they have attained some goal or acquired some skill
- motivate the learner.

In order to meet these differing needs, assessment usually has to take different forms, be undertaken at varying times and have its results communicated in various ways. Learners may require feedback early on in a course, for example, to know how well they are doing and to identify any current or potential problems. This type of assessment is known as *formative*.

An employer, on the other hand, may wish to know what a trainee can actually do at the end of a training programme. Students, too, will want this kind of *summative* verdict on completing a course of study. Before enrolling on a course, tutors may wish to identify any weaknesses students have and so they may wish to give learners a *diagnostic test*.

Formative assessment

This type of assessment is used to monitor learning progress during a course or period of training. Its prime purpose is to provide feedback to students and to the

tutor so that achievement or performance can be improved. For teachers and trainers, formative assessment also provides information about how successful they have been in enabling students to learn. Usually, it confirms what the teacher already knows. Occasionally, however, it may indicate the need to change teaching strategies, for example, to provide tasks at different levels for specific individuals or to spend more time with the whole group recapping and consolidating earlier learning.

Much formative assessment is informal, that is, it is based on the tutor's observations of what is happening in the classroom or workshop; but more formal tests of various sorts are also used. Many of these are devised by teachers or trainers in the normal course of their work, although in some competence-based programmes such monitoring devices are specifically designed by the validating body to check on the ongoing achievements of students.

For the trainee or student, formative assessment provides reinforcement or a metaphorical 'pat on the back'. It may also identify specific problems, errors or weaknesses to which the student needs to pay attention.

A key aspect of continuing formative assessment is that it provides a basis for discussion between tutor and learner. Assessment is the aspect of learning or training which is most likely to cause concern for those being assessed. It is important, therefore, that students are fully aware of what is involved, what is expected and how to prepare.

Teachers need to allocate time to discuss grades and performance with students. It is important for this dialogue to take place as assessment without communication is of limited value. This communication is often of an informal nature in the class, workshop or workplace, but it can also take place effectively during scheduled one-to-one tutorial sessions, by letter, over the telephone, via email or electronic or video conferencing.

Formative assessment can provide insights for the tutor into what material needs covering again and can therefore mark the starting point for the next session. The techniques most commonly used for formative assessment typically include oral questioning in class, short answer written tests, essay or assignment tests and assessments of ongoing practical activities undertaken in workshops and classrooms.

Summative assessment

As the name implies, this type of process comes at the end – or summing up – of a programme, course, module or unit. It is typically designed to assess the extent to which learning has been achieved, the quality of students' work and, in some cases, to assign course grades and final certification.

The techniques used for summative assessments are various but may include the following:

- examinations produced by examining bodies

- projects
- teacher-produced achievement tests
- skills/competency assessment
- inspection of diaries, laboratory note-books and work experience reports
- observations of products, portfolios and craft products.

These methods of assessment, amongst others, are discussed more fully in Section 10.

Diagnostic assessment

Diagnostic assessments are used to determine the presence or absence of necessary skills or knowledge. They may also determine the underlying causes of repeated learning difficulties. A diagnostic test may highlight, for example, that a student on a vocational course is struggling with his or her studies because of a lack of adequate numeracy skills. Once identified, appropriate support can be given to this learner.

The differences between formative, summative and diagnostic assessment are not always clear cut. If coursework contributes to a final grade, for example, it becomes part of the summative assessment. Any formative assessment of that coursework, therefore, does not necessarily provide the student with the opportunity to improve.

Formative and summative assessment can also be used for diagnostic purposes. Diagnostic assessment may be lost in some cases if composite or aggregate grades mean that students are unable to identify their particular strengths or weaknesses.

Product and process

When thinking about assessment, we have a tendency to focus solely on the product outcomes of the students' learning. We think, for example, about the artefacts learners may produce in studios, kitchens or workshops; written work in the form of assignments, projects or exam answers; output from a computer printer or screen; or oral answers to questions given in a classroom or at the workplace.

As we have already noted, there has been a trend towards formally assessing students' transferable, personal skills as well as the content of what they are studying. Sometimes we have educational outcomes which require the assessment of processes – such as motivation, work habits and relationships with other people.

The method which is most appropriate in many cases for this type of assessment is observation by teachers, who recognise many different sorts of characteristics of their students during the times that they are with them. We note the students who are effective in the use of their time and we note those who are persistent and hardworking. In discussion sessions we might observe that some are not only more willing than others to speak but that they seem more capable of framing their thoughts cogently. We may also note learners' reactions to praise and criticism as well as their sensitivity or bigotry.

Product and process can be so closely intertwined sometimes that it may be difficult to identify precisely which is being assessed. This may be the case, for example, when assessing a presentation by students or a group project.

Competence

We mentioned earlier the increasing emphasis on the measurement of competence. Most vocational qualifications have been designed as – or converted into – NVQs (National Vocational Qualifications), following the establishment in 1986 of the National Council for Vocational Qualifications (NCVQ), since incorporated into the Qualifications and Curriculum Authority (QCA).

The analogy of the practical driving test, with which most people are familiar, is often used to illustrate this competence-based approach. As long as you are 17 years old, there are few restrictions in taking the driving test. You can take it at any time of the year. You can be taught by friends or relatives or you can seek, and pay for, specialist training from a range of providers. You may wish to take the test without having had any training if you feel you are competent and ready to be assessed. If you do not pass the test, you can keep trying until you do. You are not competing with anyone else.

Like the NVQ assessor, the driving test examiner may not be concerned about how, when, where and for how long you had prepared for the test. He or she is only interested in assessing competence, although we might hold that there is an additional 'knowledge-based' component in the form of a written test on the Highway Code. Assessment is independent of mode of learning. Again, like the NVQ assessor, the examiner uses national standards to assess your performance.

The competence-based model, therefore, looks at what candidates can *do*. The assessment should be under conditions which are as closely related to working conditions as possible – ideally, in the workplace itself. Candidates are either competent or 'not yet competent' in relation to specific learning outcomes.

The qualifications are written in the form of competences which are related to those skills likely to be found in daily use in industry and commerce. Thus, competences should represent national practices. These competences are endorsed by those responsible for employment in the industry. Having established national levels of competence (see below), NCVQ invited industry to form 'Industry Lead Bodies' (now 'National Training Organisations'), which represented the employers in these occupational areas. Competences are hierarchically structured in levels and relate to higher and higher levels of performance.

- Level 1 – recognises competence in a range of work activities which are primarily routine and predictable or which provide a broad foundation.
- Level 2 – recognises competence in a broader and more demanding range of work activities involving greater responsibility.
- Level 3 – recognises competence in skilled areas that involve performance of a broad range of work activities including many that are complex and non-routine. Supervisory competence may be a requirement at this level.

- Level 4 – recognises competence in the performance of complex, technical and professional work activities, including supervision or management.
- Level 5 – 'higher level' – recognises competence in the pursuit of a senior occupation or profession (as an employee or as a self-employed person) including the ability to apply a significant range of fundamental principles and techniques to diagnosis, planning and problem solving. Extensive knowledge and understanding will be required to underpin competence at this level, together with capability in management and supervision for executive and some professional fields.

While there are considerable advantages to the competence approach, there are also many concerns about this model of assessment. One of the most serious is the debate about *degrees* of competence. Some would argue that simply 'can do' is not enough and that there is no incentive in this model to excel. Think about the people who might undertake work for you, your family or your friends – hairdressers, dentists, builders or car mechanics – and the expectations you have of their competence. If you need to call out a plumber in an emergency, you would probably want a *very good* plumber, not one who once demonstrated competence according to what might be considered as minimum standards. In identifying what he calls 'shades of competence', Race (1993) highlights the difficulties in deciding exactly what we mean by competence (see Figure 15).

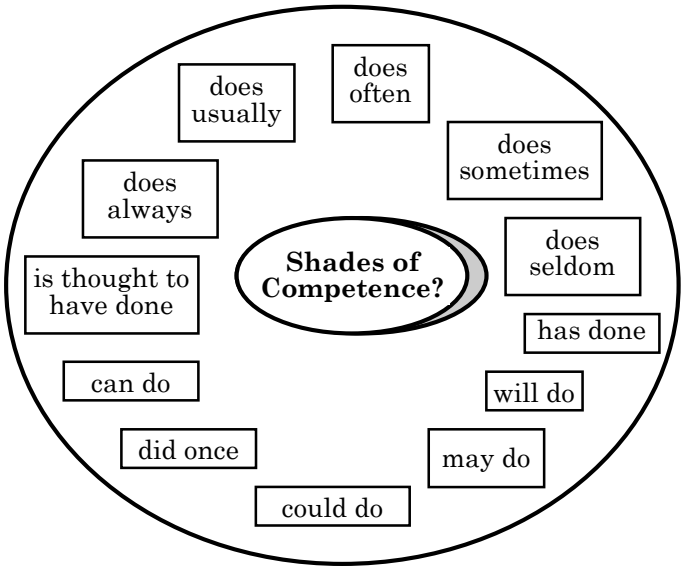

Figure 15 Shades of competence

(Race, 1993: 42)

Race then goes on to suggest some descriptors, which demonstrate the difficulties in assessing competence (see Figure 16).

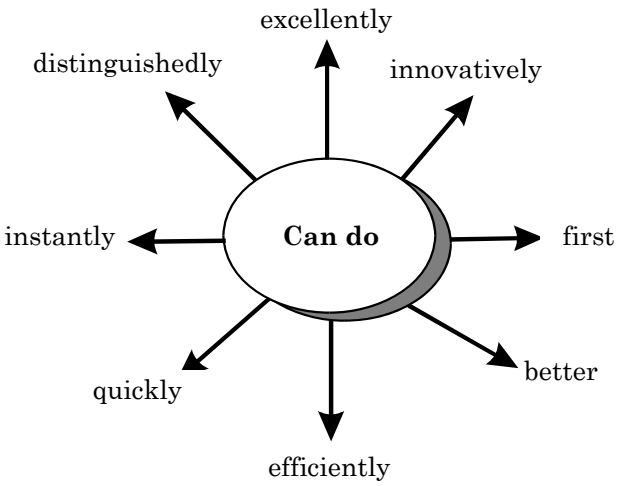

Figure 16 Some competence descriptors

(Race, 1993: 43)

In most competence-based programmes, learners are assessed according to prescribed *performance criteria* as written by the Lead Bodies, or National Training Organisations (NTOs) as they are now known.

We shall be looking at examples of performance criteria in the next section.

8. Types of Assessment

We need now to explore more carefully what we mean by 'criteria' and introduce two important terms in relation to types of assessment – *norm-referenced* and *criterion-referenced*. We will start with a hypothetical example, based on an idea by Brown & Pendlebury (1992), to demonstrate the meaning of these terms.

A caveman says to his daughter 'Now go out and kill your first bear'. This task is an example of criterion-referenced assessment. Criteria provide a list of things that a learner should be able to do in order to complete a course of study successfully. In this example, there is only one criterion – to kill a bear. This criterion has to be met in order to succeed and the learner will either pass or fail. If the task had been changed to 'Go out and kill as many bears as you can' and it had been given to all the children in the locality, then the test would have been a norm-referenced task which put candidates in rank order based on their scores.

Let us look at the purpose of this exercise. If this task had been used to measure the extent of learning at the end of a hunting training course it would have been summative. If, on the other hand, it had been used during the training in order to provide feedback to the children on how well their hunting skills were developing, it could be described as formative.

Norm-referenced assessment

Norm-referenced assessment compares one person's performance with that of everyone else being assessed, either in the present or over a period of time. It places students in rank order or on a *normal distribution curve*. Figure 17 over the page shows a typical normal distribution curve with 90% of students passing, 10% of them with merits and the top 10% with distinctions. The lowest 10% have failed. Although these percentages may remain the same from year to year, the cut-off point for a fail, merit or distinction will change depending on the knowledge and skills of the learners undertaking the particular assessment. So a pass-mark one year may lead to failure the next.

If everyone is coached for the exam, for example, the standard and quality of work may go up considerably for all students, and improved grades may show this. However, the lowest 10% of students will probably continue to fail.

There must be a good spread of marks and the questions must discriminate well between students. If all students get marks of 90%–100% or all get marks below 10% the norm-referencing system breaks down because it cannot then be used effectively to rank-order performance. Norm-referenced assessment determines a level at which someone is deemed to have been successful. Even within a classroom setting a tutor may, for example, describe a student's performance as 'being within the top 5%' of her class. Had that student been in the group in the previous year, she may not have been in the top 5%.

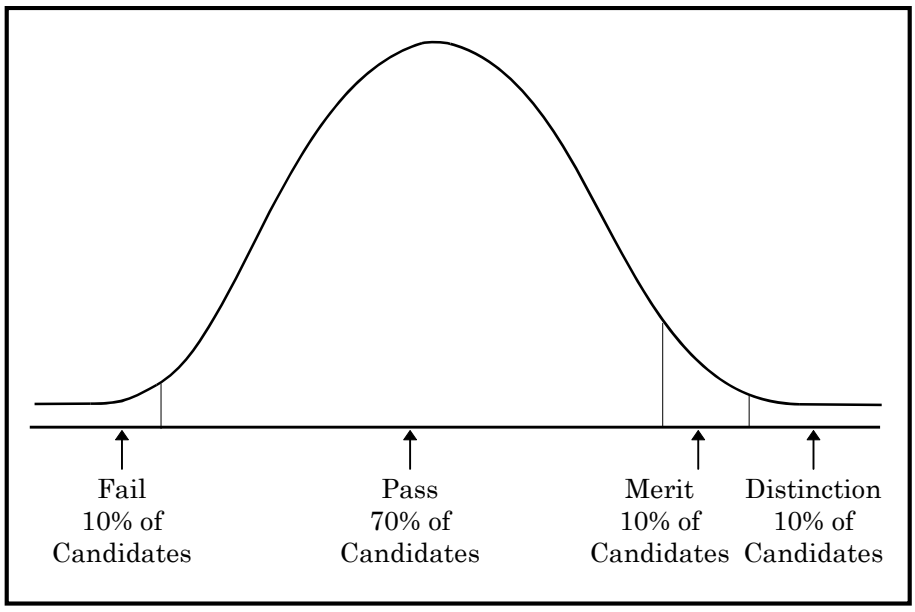

Figure 17 Example of a normal distribution curve

Although norm-referencing is still commonly used in formal – especially nationwide – assessments, it receives much criticism. It may demonstrate people's relative standing but it says little, if anything, about what they can actually do. It also makes many of those involved in the process feel like failures. Some of you – or your friends and relatives – may have experienced this type of assessment in the past. You may, for example, remember the 11+ exam or General Certificate of Education (GCE) 'O' levels, the precursor to GCSEs.

Norm-referenced tests are an essential component of any system which is based on selection, and is likely to be favoured by those who believe that competition with others is an essential component of the educational process.

Criterion-referenced assessment

Criterion-referenced assessment, on the other hand, sets out a list of things that a learner has to be able to do to demonstrate *mastery*. If these things can be done to the required level, then the learner succeeds, regardless of how many other students can or cannot achieve them.

Instead of comparing students with each other, a candidate's performance is assessed against some independent, free-standing definition of what should be achieved. In theory, all learners in a group could succeed or all could fail. To pass, you need simply to meet all the criteria. The practical driving test is criterion-referenced.

Let us look at an example of the use of such criteria. Students on a course leading to a teaching qualification for the PCET sector are likely to be asked to plan a teaching

session, teach it and evaluate it, and design an overall scheme of work. The criteria for assessment may look something like this:

> The work must show:
>
> a) a professional approach to the preparation and planning of teaching
> b) coherence and appropriate organisation in the scheme of work
> c) competence in the delivery of the session
> d) evidence of the ability to reflect on own performance and to respond positively to constructive feedback.

Figure 18 Example of assessment criteria

Assessment based on the public statement of performance criteria has been promoted by policy makers and qualification designers over the last few decades. The Certificate of Secondary Education (CSE), introduced in the early 1980s, was intended to be criterion-based and to replace the norm-referenced GCE 'O' levels. The National Curriculum 5–16, the Scottish National Certificate, National Vocational Qualifications (NVQs) and General National Vocational Qualifications (GNVQs) are all designed to be broadly criterion-referenced.

Wolf (1993) identifies a number of advantages to this approach. She believes this method can improve teaching and learning; make assessment results comprehensible and useful and provide an opportunity for everyone to have their positive achievements recognised.

Learners need to know what is expected of them and what they need to do to succeed. Explicit criteria can provide this. If students do not succeed, of course, they need to know *which* criteria they have not met. Although in recent years it has been common practice to make learners aware of the assessment criteria, many marking schemes in the past remained 'secret' to the learner, especially – though not exclusively – in public examinations.

Figure 19 provides an example of formative feedback given to a student on a teaching course for the post-compulsory education and training (PCET) sector, after an observed lesson.

Note how this feedback, following marking, relates directly to the assessment criteria in Figure 18.

> **a) a professional approach to the preparation and planning of teaching**
> *Your lesson plan is detailed, with appropriate aims and learning outcomes. The assessment strategies are appropriate. The activities are innovative and the timings are generally realistic. Your learning resources are beautifully prepared. However, is it practical to copy them for all your students?*
>
> *Continued...*

b) coherence and appropriate organisation in the scheme of work
The scheme of work relates well to the syllabus. It is realistic, clear and the topics are logically sequenced. You might, however, need more time allocated within the scheme for in-class assessment. It is not clear from the scheme exactly how much time will be allocated for the tasks.

c) competence in the delivery of the session
Your session was successful. Learning outcomes were met. Your style was relaxed and confident. You demonstrated good classroom management skills and you ensured that ALL learners were involved in the activity. Your writing on the whiteboard could be neater (keep practising!) and be careful not to have your back to the group for too long. One possibility would have been to ask one of the students to record the feedback on the whiteboard, leaving you free to manage the discussion.

d) evidence of the ability to reflect on own performance and to respond positively to constructive feedback
In your evaluation you have reflected thoughtfully on your performance. You have tended, however, to focus on the areas for improvement (what you call weaknesses). It is important to think, too, about your strengths. You need to know what it is you did to make things go well. Remember all the positive comments from your peers and tutor!

Overall - a very good lesson. All criteria met. Well done

Figure 19 An example of marking according to given criteria

Criteria are generally written by tutors (as in this example) or by those working for awarding bodies. There are many benefits, however, to encouraging students to devise their own criteria, either independently or in groups. This is discussed further in relation to self- and peer-assessment later in this section.

In *National Vocational Qualifications* (NVQs) the broad areas of competence are broken down into 'sub-areas', entitled *units*. Units are then divided into *elements of competence*. These elements are then broken down still further into *performance criteria*, which represent the fine detail of the activity. This might seem complex, but an example will make it clear.

Figures 20 and 21 are related to a qualification for dental nurses.

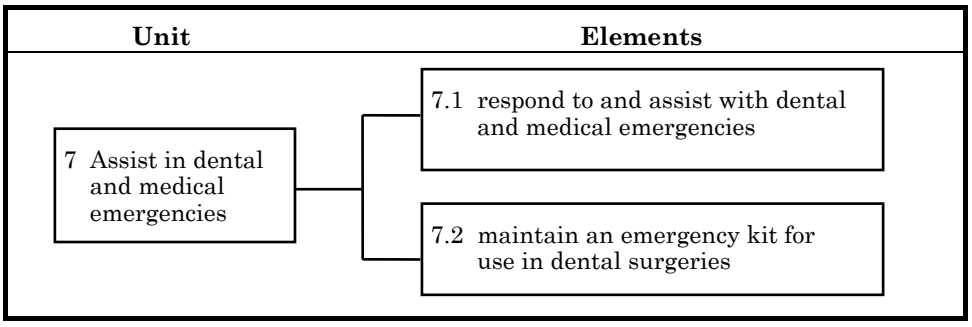

Figure 20 An example of a unit and two related elements

This shows just one of many units within the area of competence. Like other units it includes elements, and in this particular case there are two elements. The performance criteria for first of these elements are shown in Figure 21.

Element	Performance criteria
7.1 respond to and assist with dental and medical emergencies	a) the patient is reassured throughout by verbal and non-verbal means b) immediate/preliminary action is taken appropriate to the medical/dental condition c) the patient's pulse, temperature and breathing are monitored d) medical assistance is called for e) the patient's dignity is respected throughout

Figure 21 An example of an element and the related performance criteria

The candidate, as with all such models, is required to provide evidence to support claims of competence for this and other elements. Ideally, this is achieved in the workplace. If this is not possible, simulations are set up and the candidate is interviewed. The candidate collects evidence to support a claim of competence, usually in a *portfolio*, a method of assessment which is discussed in Section 10.

Criterion-referenced assessment has its drawbacks, too. Criteria may be unclear or ambiguous. They are not easy to write and this is particularly the case where aesthetic judgement by the tutor is required (see Section 10 in relation to creative arts and artefacts). Despite the use of criteria, assessment can remain subjective.

Those who design the assessment tasks may fail to agree on the criteria and, in marking the work or performance, assessors may interpret the criteria in different ways. Using the criteria in Figure 18, for example, assessors may not fully agree on what is meant by 'professional' in the first criterion. These problems, however, are normally resolved by a process of moderation or verification, as we discuss later in Section 9. There may be a wide variation in terms of the quality of work which might meet the specified criteria – excellent, good, sound and barely satisfactory. In some subjects, such as maths or the sciences, there will be answers which can be either right or wrong. In other subjects, there are shades of grey. The differences between levels of work was discussed in the previous section in relation to shades or levels of competence.

One method of recognising 'more than just competent' or 'easily meets the criteria' is to give some form of classification. Sometimes additional criteria are included for a Merit and for a Distinction. They may look something like this:

In order to gain a merit, all the criteria for a pass must be met and, in addition, the work must show …

In order to gain a distinction, all the criteria for a merit must be met and, in addition, the work must show …

Another method of classifying passes, used most frequently in higher education, is to link a numerical grade or percentage to specific criteria, as in Figure 22.

Write a review of a book of your choice for your professional journal.

0-29%	Does not constitute an answer to the question.
30-39%	Lacks many of the characteristics of a book review. Gives inadequate information about the book.
40-49%	Has basic characteristics of a book review. Gives information about book being reviewed.
50-59%	Has characteristics of a book review. Gives information about book being reviewed. Has adequate evaluation.
60-69%	Has characteristics of a book review. Gives information about book being reviewed. Evaluates the book being reviewed and judges its suitability for the readership of the journal.
70+%	Suitable for publication for the journal: is informative, evaluative, creative and interesting.

Figure 22 A set of criteria for writing a book review

(Gibbs, Habeshaw & Habeshaw, 1988a: 115)

With this example it may be hard for a teacher to explain why one student gets 52% and another gets 58% if both learners meet the criterion linked to that range of grades (50%–59%). Some tutors are very fond of this method of marking and will claim that they 'know a 52% when they see it'. In effect, they are probably operating a combination of norm-referenced and criterion-referenced assessment.

One of the problems with criteria is that there may also be *hidden* criteria. Teachers may, for example, dislike errors in grammar, spelling or punctuation. The use – or misuse – of the apostrophe is a particular favourite! While they may admit to themselves and to colleagues that this might influence their view of a student's work, teachers may not give feedback as no reference is made in the criteria. Students may remain unaware of this. This situation can be avoided, of course, by informing students and, if appropriate, making explicit any hidden criteria.

It is important to remember that the terms criterion-referenced and norm-referenced refer to the *type* of assessment, whereas the terms summative, formative and diagnostic, discussed in an earlier section, refer to the *purpose* of assessment.

In practice, the relationships between these terms is not always clear cut. Although criterion-referenced assessments are usually based on a pass or fail system, tutors still compare students' work and they also tend to have in their mind a certain number of students who they expect will pass or fail, based on their experience of previous years and maybe on departmental or organisational norms. It should be noted, too, that norm-referenced tests do have criteria – if only to establish the norm.

Alternatives to assessment by tutors

Assessment is most frequently undertaken by tutors or trainers and by external examiners and assessors. This need not always be the case though. Here are some alternative models.

Self-assessment

Self-assessment is a valuable form of assessment for learners. It can enable students to 'own' their learning. The simplest form of self-assessment is to ask students to mark their own work. They need not share the results with anyone else. There are, however, various ways in which self-assessment can be more formal. Trainees may be required, for example, to keep a reflective diary in which they chart the progress of their own learning and this form of assessment is discussed in Section 10. Self-assessment encourages learners to be reflective.

Self-assessment questions (SAQs), used in many Open University materials, provide a useful way in which learners can check their understanding or progress. SAQs pose questions for the reader to reflect on. If appropriate, answers are then provided. Although designed for distance learning modes of study, they can also be used effectively within a class or workplace environment at all levels. Figure 23 provides an example of an SAQ for you, as a reader of this particular book.

> Read the first pages of this section again (pp79–85). Now write down in your own words what you think the terms **criterion-referenced** and **norm-referenced** mean. Check your descriptions with those we have given. Have you understood these two terms?

Figure 23 An example of a self-assessment question (SAQ)

Peer-assessment

Assessment can also be undertaken by peers. Students may mark each other's work or judge each other's performance. *Peer-assessment* is becoming increasingly popular. It may involve students in marking each other's work using given criteria or devising their own. Peer-assessment does not necessarily involve students in passing or

failing their colleagues. It often requires the group to provide feedback after some type of presentation. This in turn can help students with their own self-assessment and reflection.

Many people find it difficult to assess their peers. Often, we do not want to be too critical and we may be concerned that any criticism will ruin a good relationship. The role of the teacher or facilitator is crucial, in giving guidelines to the group, outlining the process and explaining the benefits that can be drawn from taking part in assessing each other's work.

Race, who is very keen on student-devised criteria and on self- and peer-assessment, comments (1993: 51):

> *Where it is possible to draw assessment criteria from learners themselves, especially in group situations, the sense of ownership which learners develop is very powerful, and leads to them using the criteria with considerable enthusiasm and commitment when self- or peer-assessing.*

He goes on to describe the many tangible benefits for students and focuses on the processes involved. These include identifying criteria, discussing work with colleagues, making judgements and facilitating discussion in relation to fairness and objectivity.

Employer, client or sponsor assessment

In Section 7 we spoke about the many stakeholders who have an interest in the learning outcomes of others. There are many examples of when it is helpful to involve these people in assessment. Where a student is undertaking a project for a client within his or her own workplace, for example, it may be appropriate for the work to be assessed by the client.

Some competence-based programmes, like NVQs, as discussed earlier, are assessed in the workplace by assessors but the underpinning knowledge is 'delivered' and assessed by teachers at educational establishments.

NVQs require competence to be demonstrated in a 'realistic environment', i.e. the workplace, and ideally it is the employer who assesses, usually by observing and questioning the candidate while he or she is at work. Students completing NVQs in a college, however, may have to be assessed by their tutors while they are on work placement or when they are at college, 'working' in a simulated environment, such as a hairdressing salon, training office or training restaurant.

When learners on a business studies course present a business plan for a new venture they are planning, for example, assessment by a local bank manager may be invaluable. This element of 'real life' assessment can contribute significantly to learners' motivation.

However, despite the benefits of self-, peer- or 'outsider'-assessment, it is not unusual for students to feel more comfortable being assessed by the 'expert' teacher. They may believe that teacher assessment is more likely to be fair.

9. Considerations of Fairness

In considering the fairness or unfairness of assessment in this section, we will examine three key concepts – *validity*, *reliability* and *utility*. We will then go on to discuss other factors affecting fairness and conclude by looking at some of the principles of *quality assurance*.

Validity

Validity is the extent to which a test measures what it is supposed to measure. A tape-measure is a valid tool for the measurement of length. It is not valid for measuring weight.

There is little point in conducting any kind of assessment unless tutors can be reasonably sure about what they are trying to measure. If you were assessing how well a student on an office technology programme could answer the telephone, for example, a written essay on telecommunications would not be an appropriate form of assessment. A role-play is likely to be a more valid form of assessment in this particular case.

It is clearly necessary to ensure that what is chosen for testing is representative of what the learner has learnt. Before deciding on the form and techniques of assessment, therefore, it is necessary to define educational aims and objectives for the learning which is to be assessed.

In your own teaching you will be aware of the importance of planning your sessions carefully. During the planning process you will be identifying learning outcomes for your students and also thinking about how you will check whether or not your learners have met these outcomes.

If we look at a section of a lesson plan (Figure 24, overleaf, taken again from the numeracy example we have considered in detail before) we will see that the assessment methods relate directly to the learning outcomes and that the tutor will be able to check whether or not these outcomes have been met.

Assessment techniques work best where learning outcomes have been clearly articulated in advance and, ideally, shared with learners.

The validity of an assessment may be affected by how well the learner has been prepared. It is not just a case of how much they have revised or how much work they have done in class. The feeling conveyed by tutors about the significance and importance of the assessment, as well as all the practical information about rooms and times, is also critical.

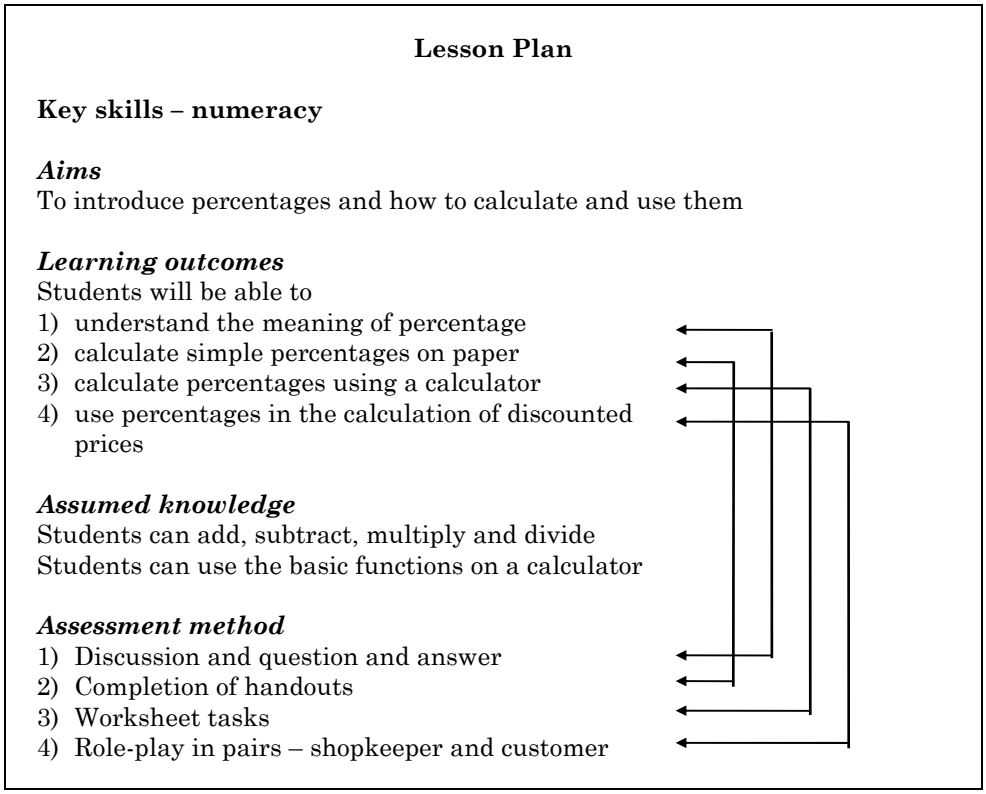

Figure 24 Part of a lesson plan

In terms of validity, there are other issues, such as face, content and criterion-related validity.

Face validity

Sometimes validity can be superficial and this is often referred to as face validity. Questionnaires in magazines, for example, may claim to tell readers about their personalities and presume to conclude that someone is a 'fast living extrovert with a taste for adventure'. Comments like these may tell people something about themselves but they do not really measure personalities.

Content validity

Content validity refers to how well any assessment process samples and measures accurately the knowledge and skills specified in the objectives of the lesson, module or unit. Tests can only sample some of the total possible items which may be included. For true content validity a test must adequately sample the relevant field of knowledge. This is an issue we will return to when we discuss the advantages and disadvantages of examinations. Exams often do not have high content validity. This is because students can pass exams even if they are very selective about what they revise and so they can ignore large proportions of the relevant subject matter.

To return yet again to the analogy of the practical driving test, it would be like allowing people to pass the test even if they had chosen to exclude some parts of the learning, for example, turning right or using indicators.

Criterion-related validity

Criterion-related validity relates to two of the prime purposes for which some tests are constructed – selection and estimation. It may be defined as the extent to which test performance is related to some other valued measure of performance. Some examples may help to clarify this.

Selection – Tests are often constructed with the intention of predicting future performance, but the extent to which they do so is largely a matter of statistical correlation. The 11+ examination, for example, was intended to be an indicator of a pupil's future academic suitability. Where a comparison of test results with some future performance is undertaken (e.g. A level results and final degree classifications) the term 'predictive validity' is often used.

Estimation – This concerns making estimates of present status. A Spanish vocabulary test, for example, might be used to estimate students' current skill in actual conversation in the language. This is often known as 'concurrent validity'.

Reliability

Reliability refers to the consistency of measurement. It refers to the extent to which an assessment will produce the same pattern of scores with the same population of students on two different occasions. However, we cannot expect test results to be perfectly consistent. There are a range of factors which may affect reliability.

If a test is administered to the same group twice in close succession, we might expect some variation in grades on the second occasion due to ill health, stress, inattention or any other reason. Nevertheless, the same general trend in the pattern of test scores should be detectable and, unless the measurement can be shown to be reasonably consistent over different occasions, we will not be able to place much confidence in the results.

With a long time-gap between testing, we may well see marked variation in an individual student's scores. Indeed, we may expect such a 'pay off' if a student has undertaken extra work or received more attention from the tutor.

Teachers normally establish pragmatically the reliability of the assessment measures which they use. They may, for example, look at assignment grades, note a shift in rank order amongst their students and then concern themselves with determining the reasons for any changes.

At a national level, there is a strong argument for seeking reliable end-of-course measures of student achievement. There is much debate, for example, about the reliability of degree classifications. Is a 2:1 (Upper Second) from one university equivalent to a 2:1 from another?

In many organisations, the notion of reliability in assessment is being replaced by the concept of quality assurance, which emphasises the need for procedures within organisations which are designed to ensure that assessments are as fair, reliable and valid as possible. This is discussed more fully below.

Utility

Earlier it was maintained that assessing learning should be a part of the teaching process and not just an afterthought. There are, however, also practical problems that relate to any form of assessment and these must be taken into consideration.

Assessments need to be:

- easily administered
- cost effective to produce and mark
- easily scored against well-understood criteria
- meaningfully interpreted and the results effectively applied in the context of the students' learning needs.

With formative assessment, the principles listed above usually present fewer problems. The devices we use involve little cost, although the greatest cost, of course, is tutors' time. In the case of summative assessment, however, the costs of producing certain assessment tasks may be high. Furthermore, their interpretation and application to students' learning needs may take some time to take effect. Discussion of the results and appropriate remedial work may not be immediately possible.

Factors affecting fairness

Physical or emotional factors

We are probably all familiar with the concept of 'exam nerves' or 'interview nerves'. Some students will perform less well under time constraints or when they know they are being formally assessed. During a workplace assessment, for example, trainees may become less proficient and confident than they would normally be owing to the pressure of the situation. Physical or emotional problems may prevent students from doing their best.

Assessor problems

The tutor may be unfamiliar with a particular programme or method of assessment. She or he may express a bias, either in favour or against a student, based on previous knowledge of the learner or on prejudices. Blind marking is sometimes used to avoid assessor bias. With blind marking students are given a number as a form of identification and therefore when the tutor marks the work, he or she does not know the name of the student. However, this method is only appropriate with written assessments and when large numbers of students are involved.

Differences may exist between assessors in their interpretation of the criteria. This may be because of variations in subject knowledge or degrees of confidence in making a judgement. If differences exist, they should be identified by the *verifier* or *moderator* as part of the organisation's quality assurance and control systems, as discussed below.

Inequalities

Learners may be unfairly penalised if they are assessed in a language which is not their first language. If a certain standard of written or spoken language is required, it should be stated in the criteria and students should be made aware of this and fully prepared.

Materials used for assessment purposes, for example, may be overtly Euro-centric and therefore discriminate against those students who are not European.

The language of assessment itself – whatever the students' first language – can also be a barrier. There is a bewildering amount of jargon surrounding assessment, much of which you may be struggling with yourself! Learners may be presented with not just tasks but also with a language which is alien to them.

Insufficient resources may also be a cause of unfair assessment. Tasks which require the use of the Internet, for example, may favour those learners who have access to the appropriate technology at home or at work.

Can assessment ever be fair? Gipps (1995: 279) suggests that:

> *There is no such thing as a fair test, nor could there be: the situation is too complex and the notion simplistic. However, by paying attention to what we know about factors in assessment, administration and scoring, we can begin to work towards tests that are more fair to all groups likely to be taking them, and this is particularly important for assessment used for summative and accountability purposes.*

Quality assurance

Several mentions have been made of moderators or quality assurance systems. Let us now look at this topic in more detail.

Most organisations involved in education and training have systems in place to avoid the kinds of barriers to fairness discussed above, as part of their quality assurance and quality control policies. This was mentioned above in relation to reliability. There are usually systems of internal moderation or verification, whereby tutors and assessors check each others' marking. In many cases, students' work is marked twice. Internal moderators or verifiers within an organisation may sample a range of work from different tutors. They offer advice and support to assessors and act as a link between the assessors and the awarding bodies.

There are also external moderators, verifiers and examiners. External moderators and verifiers are usually employed by the awarding or examining body. By visiting a range of institutions offering their particular programme or examination, they are able to comment on comparability of standards. They are also involved in ensuring that the assessment process is operated fairly and according to approved procedures and regulations.

External examiners, favoured by the higher education sector, have a similar role but they are usually employed by universities and they normally work on courses similar to the one they are examining.

Those responsible for assessment within NVQ programmes are usually qualified to what were formerly the *Training and Development Lead Body* (TDLB) standards. (The work of this, and the other Lead Bodies, has now been transferred to the new National Training Organisations, NTOs.) This means that the assessors themselves have undergone assessment. They have had to demonstrate that they can make judgements regarding the assessment of competence, which they have collected from a variety of sources.

Assessment policies, and detailed procedures for students who wish to appeal against an assessment decision, provide another safeguard for learners.

10. Methods of Assessment

So far we have briefly mentioned a number of different ways of carrying out assessment, such as oral questioning, portfolios and examinations. There are, of course, many other methods and in this section we look at the main advantages and some of the drawbacks of different types of assessment.

The list of assessment methods below is not exhaustive but it does demonstrate a wide variety of possibilities:

- examinations
- essays
- objective tests
- projects and reports
- observation
- portfolios
- diaries and log books
- creative arts and artefacts
- questioning, interviews, orals and role-play
- problem-solving tasks
- computer-based assessment.

These methods will now be discussed in an attempt to highlight their strengths and weaknesses.

Examinations

It is often argued that written exams are the most efficient form of assessment. All the students sit them at the same time and there is little chance of plagiarism. The scripts can be marked relatively quickly by tutors, who will, in theory, have finished their teaching.

Exams do have advantages. The fact that summative exams play a key role in motivating students should not be underestimated. They can provide a stimulus for students. 'I didn't really understand this course until I started revising for the exams; then it all clicked into place' is not an unusual sentiment. This statement, of course, is more of an indictment of the teaching and learning, than an advantage of examinations as a form of assessment! However, exams can effectively test a learner's factual knowledge and his or her ability to work swiftly and under pressure.

Unfortunately, they can also encourage last minute cramming and a surface approach to learning. Students can regurgitate facts without understanding them. Exams also offer students the opportunity to exclude large sections of the curriculum they have been studying and to concentrate on questions they think will 'come up'.

The fact that the same exam paper is taken by a large number of people does not necessarily make it a valid or reliable form of assessment. In fact, if students have extensive choice within an exam, it can mean, in effect, that no two students sit the same paper. A large number of short questions would lead to greater reliability than asking students to choose just three topics out of fifteen. If students are not required to show an understanding of *all* aspects of a course – which is usually the case in exams – it means that the assessment is invalid, as we discussed in relation to content validity. There are also many issues relating to fairness, as we saw earlier. Some students get very nervous before and during exams. Other students have developed – or have been coached in – good exam techniques.

Not all examinations need follow a traditional format which starts with 'You may now turn over your papers'. There are advantages to experimenting with the format. Exams may have a flexible time constraint. This may, for example, mean giving students a whole day or even week to complete a piece of work. Trainees may be given a case study to read a week before the exam. On the day of the exam they can be given the questions which relate to the case study.

Other formats include giving learners questions they have had a chance to prepare for (seen papers) or allowing them to use their notes and other texts (open-book) during an exam. They may also be given exam papers to take away and complete at home.

Essays

Essays remain a common form of assessment for many 'academic' programmes, particularly in the arts and social sciences and especially on higher education courses. They may be very useful ways of assessing learning in that students are given an opportunity to demonstrate an ability to integrate knowledge, skills and understanding. Essay writing should be about trying out ideas and arguments supported by evidence.

There are, however, many problems with using essays as a form of assessment. Learners need first of all to acquire the skill of *how* to write an essay. In gathering evidence, there is a temptation for students to copy large sections from books (or from other students' work). Essays can be difficult to mark objectively, which is why it is essential to clarify the criteria and make them available to students, as we discussed earlier.

Objective tests

As the name suggests, *objective tests* are a type of assessment in which the marking is objective. Unlike essays, the answers to objective tests are predetermined. Marking the answer to a question like 'What is the capital of Japan?', for example, requires little in the way of subjective judgement. It is important to note, though, that the subjectivity is eliminated only in the marking and that it can remain in the setting. Let us look at the example in Figure 25, from a Level 1 NVQ Hairdressing question paper.

Choosing the correct shampoo usually depends on:

a) what the client asks for

b) the salon stock

c) the client's hair type

d) the condition of the client's hair.

Figure 25 An example of a multiple-choice question

In this example, the marking is objective, in that the person setting the test has identified the 'correct' answer, but there is clearly an element of subjectivity in the setting of the questions.

Multiple-choice questions, as in Figure 25, are just one type of objective test. Others require learners to:

- choose between A and B
- tick TRUE or FALSE for the following statements …
- complete the following sentences …
- identify the missing word in each sentence …
- match the items in one column with the items in the second column (for example, events in one column and dates they took place in the other)
- select the best answer from the following …

Objective tests have many advantages:

- they can be marked easily by tutors;
- they can be marked quickly by tutors;
- they can be marked by optical bar-code readers or scanners;
- they can be marked by someone who is not an expert in the subject;
- they can be very reliable, in that all tutors involved will be marking to the same standard;
- a greater proportion of the syllabus can be covered than in other forms of assessment;
- students are required to answer all questions;
- they can effectively test factual knowledge;
- in some cases, they can test more than just facts, as described below.

Brown & Knight (1994) argue that multiple-choice papers can effectively test higher cognitive skills, such as comprehension, analysis, synthesis, interpretation and reasoning. They describe a 'best answer' paper for medical students where a question comprises a brief description of a patient and the presenting symptoms. The trainee doctor is required to select the most appropriate prescription from a

selection of

several. More than one of the alternatives may be acceptable and experts would have to agree that one of them is the best and that it should get more marks than the others which are possible. Those that are inappropriate would score no marks or possibly negative marks.

While objective tests may be useful as one part of an overall assessment strategy, there are drawbacks to this approach. The main problem is the design of the questions, which need to be absolutely clear and unambiguous. If they are poorly designed then the assessment is not valid. There may also be problems about students guessing the answers, although many papers are designed to penalise guessing by deducting marks for the wrong answer. The over-use of this method can make learning routine and trivialise the answers. Although objective tests can be used to assess higher cognitive skills, they tend to be used primarily for factual recall, and they cannot assess creativity.

Projects and reports

Projects and reports can take many forms and they can be undertaken by individuals or in groups. They enable learners to explore in depth a particular topic. If students have chosen their own title or topic, it is likely to provide them with a form of ownership of the learning.

These tasks can foster independence and creative problem solving. They can also help learners develop project and time management skills as well as those skills relating to research, information-seeking and team-working. Realistic tasks, related to the working environment, combined with continuous assessment can be a more attractive assessment option than essays or exams for many students.

In terms of disadvantages, there is a possibility of time constraints, especially if it is difficult to meet the requirements of an externally set syllabus in the time allocated. There may also be problems with plagiarism and, with groupwork, there are issues relating to the recognition of each individual's contribution. Maintaining reliability in terms of marking may also be a cause for concern. All these problems, though, can be avoided to a large extent by providing appropriate criteria, a clear brief and a system of moderation.

Observation

Observation is about watching students and noting what you see. It can be a very effective way of establishing if learning has taken place and it is particularly useful for assessing processes, as discussed in Section 7. Teachers informally observe their learners, of course, all the time in a classroom, workshop or workplace environment. If processes are being assessed, then it is important to make sure that this is made clear in the criteria. An art tutor, for example, may be assessing not just a completed piece of work (the product) but also the way in which the learner worked through the design stages, met project deadlines or possibly worked with others.

Observation is frequently used for competence-based assessment, to observe learners in a 'real' working environment. It is usual practice, when undertaking such an

observation, for the tutor to have either a checklist or a standard form to complete. This system has advantages. If the checklist is carefully thought out, in advance of the observation, it is likely to cover the relevant aspects to be observed and this can help to ensure the validity of the assessment. It provides a simple means of recording the observation. Using the same checklist for each observation (for the same or different learners) can also increase the reliability of the assessment, particularly if more than one assessor is involved.

Assessment by observation can be motivating for the learner. It can reinforce theory, certify competence and provide an 'I can do' feel-good factor.

Some of the drawbacks of observation as a method of assessment are identified below.

- In the design of a checklist it is difficult to include *all* possible situations and some features will inevitably be omitted. Forms with broad headings allow greater scope for the assessor but may reduce reliability or validity.
- Sufficiency of evidence may prove to be a problem. You may need to observe someone several times (if time permits) to be confident about their abilities and competence.
- There may be bias on the part of the assessor. Some trainees may be given the benefit of the doubt whereas others in a similar situation may not.
- The presence of an observer will change the environment. The candidate may feel nervous and act differently. Other people involved may also behave in an unusual manner.

Portfolios

A portfolio is a documentary record or a collection of evidence. Traditionally used by art students to demonstrate their achievements, the term 'portfolio' is now applied to a whole range of subject areas, at varying levels and in different modes of work or study.

Portfolio compilers may be young people on a vocational course at school or college, collating their coursework or, as noted by Bloor & Butterworth (1996), they may be employees planning their career route through a company, students recording their period of study on a training course or applicants to a college or university putting together a claim for the *accreditation of prior learning* (APL). The advantages of this approach are that it can:

- provide a 'portable record' of someone's achievements or professional experience
- encourage people to take responsibility for their learning and for their assessment
- be a particularly effective method of claiming APL
- help employees to plan their career development
- encourage people to be reflective.

Sometimes, however, learners find the prospect of compiling a portfolio daunting. They may have difficulty in interpreting the language and jargon associated with

portfolios, especially in relation to performance criteria. It can be the case that where students in colleges are required to compile a portfolio, learning is overly assessment-driven. Teachers and students can become so concerned with the mechanics of portfolios and their presentation that this overshadows their purpose and the related learning processes and educational outcomes.

Those claiming APL for the purpose of exemption from part of a programme (for example, one unit) may find that completing the unit in the traditional way is actually easier and less time-consuming than putting together a portfolio.

Diaries and log books

Diaries and log books provide a personal record of experiences. The former tend to be more subjective than the latter but both focus on the process of learning. They are often used alongside other forms of assessment. They may, for example, be used to record experiences whilst on work placement, undertaking a science practical or working in a team on a group project. When we discussed projects earlier, it was noted that marking groupwork assignments can be problematic.

One method of improving this type of assessment is to require each member of the group to keep a diary or log, outlining their own contribution. Diaries are a form of self-assessment; they encourage learners to make reflections on their experience or their behaviour in specific situations. They can provide evidence of process for a tutor to assess. It is possible, though, that in some cases students may not be honest or may not recognise this method as a suitable type of assessment.

Creative arts and artefacts

The competence-based approach is at its most controversial in relation to the assessment of the creative arts. Many would argue that providing learners with explicit criteria is likely to stifle their creativity. However, guidance in the form of criteria does remove the unfairness of a system whereby students are expected to guess what is in the mind of the assessor. Brown & Knight (1994) argue that the mystery surrounding assessment of the creative arts stems from the old master/apprenticeship relationship in which the master had the power to decide whether the apprentice's achievement was aesthetically pleasing or not. Overt and clear criteria can provide the necessary guidelines for technical competence, techniques, time scales and use of materials.

However, even with these criteria, there will remain an element of subjectivity in the assessment of any 'final' product, be it a ceramic bowl, a painting, a hairstyle, a chair, a poem or a meal. It is important for tutors to recognise this element of subjectivity and be open and honest with students in their feedback.

Questioning, interviews, orals and role-play

Oral question and answer is frequently used for formative assessment of students' learning. The main advantage of this type of assessment is its immediacy, which is why it is popular with teachers in the classroom setting. It is normally used in

conjunction with other methods of assessment. Its success depends to a large extent on the skill of the teacher. Teachers have to ensure that questions are appropriate, that they test understanding and not just factual recall and that they are answered by all students and not just a few dominant ones. Teachers often ask 'Are there any questions – anything you don't understand?' but sometimes fall into the trap of assuming that students' silence equates to students' understanding.

Interviews are sometimes used to contribute towards summative assessment. This may be the case, for example, with candidates submitting a piece of work or a portfolio. The interview can be used to confirm that the candidate is the author of the work and to explore certain issues in more depth. This is similar to the use of vivas in colleges and universities for doctorate students or for students who are borderline in terms of their grades.

In preparing students for finding employment, mock interviews can assess learners' personal qualities and attitudes and interpersonal skills, particularly if used with a check-list of characteristics.

Oral exams and role-play exercises tend to be associated with language learning. In this context, they can very effectively assess fluency and comprehension. Role-plays can also be used at all levels on a wide range of programmes to assess students' ability to think quickly and react in an appropriate way to varying situations. Students can be given a scenario and then asked to act it out, either with each other, or with an examiner or teacher. A student on a travel and tourism course, for example, may be required to demonstrate how he or she would react to a very awkward and unpleasant customer or holidaymaker. Law students may be required to act out scenes in a court to assess their negotiating skills and knowledge of the law.

Clearly, these forms of assessment have many advantages. One-to-one interviews and oral exams enable the questioner to explore understanding and pick up issues not covered in a student's written work. Oral exams and role-plays are very practical, 'real life' forms of assessments and, as such, students will appreciate their relevance.

However, these types of 'on the spot' situations can also be very stressful for some students. Although there are likely to be criteria for passing or failing, the assessment can appear to be very subjective, which raises questions about reliability. There may also be practical problems related to arranging orals, interviews or role-plays for large numbers of students within a limited time frame.

Problem-solving tasks

This is a broad heading for a wide range of activities, which can take many forms. Problem-solving tasks are often related to practical assignments which take into account commercial pressures and practices. A day-long in-tray assessment, for example, may require students on an office technology course to read all the papers in their 'tray'. They then have to scan through all the documents, prioritise the tasks and complete those that need to be finished by the end of that day. Clearly this type

of assessment can assess some working competences more effectively than a written exam.

Mini-enterprise tasks are of a similar nature. Students on business studies programmes may be required to set up their own business to raise money. These assessments can continue throughout a term, semester or even a whole academic year. Students may, for example, offer a car cleaning service to colleagues and staff. In doing so they need to write a business plan, design, print and distribute leaflets, organise 'staffing' schedules and training as well as keep the accounts. Such activities provide the opportunity to assess not just business knowledge but also interpersonal and team working skills. They can very effectively demonstrate how students have related theory to practice. This type of assessment can be very time-consuming to arrange and to monitor for the tutors involved and, particularly if students take on different roles within their team, criteria for assessment have to be made very clear to ensure reliability and validity.

Some problem-solving tasks are computer-based. Simulations, for example, enable students to experiment with different variables to see the effect this has on certain situations.

Computer-based assessment

Technology is used in all sorts of way to assist in assessment. Optical mark scanners or readers are used in the marking of objective tests, as noted earlier in this section. Computers are, of course, also used to create assignments as well as to record, analyse and communicate results. Feedback tutorials are increasingly undertaken, as we saw earlier, by email, electronic or video conferencing.

There are software packages available which enable learners to respond to questions on the computer screen. Tutors can write their own banks of questions at different levels. Banks of questions for some syllabi – notably GCSE and A levels – are sold commercially. On most systems, multimedia resources such as sound or video can be brought in to present a test or give feedback on answers. The responses of the students are stored and collated automatically. The system can provide students with instant feedback and it is generally easy to use. Tutors may, of course, have to spend time initially devising the questions. Appropriate computing facilities are necessary for each student, although not all students have to take the assessments at the same time.

It is important not to be overawed by the technology and to remember that these objective tests can have the same drawbacks and limitations as the manual versions and can lead to a surface approach to learning.

Computers are also used to provide simulations of events which could not be otherwise simulated. Assessments can be designed around these simulations. Gibbs, Habeshaw & Habeshaw (1988a) describe a computer simulation of the economy. A desired economic state for five years from now is specified and today's economic indicators are input at the beginning of the session. Students are then set the task of achieving this desired state by manipulating those variables within the control of, say, the Chancellor of the Exchequer. With this, as with other simulations, students explore a variety of 'what if...' scenarios.

Advances in technology will no doubt continue to facilitate more 'real life' assessment. Virtual reality is already used in teaching and training in some institutions. It places a person into a simulated environment that looks and feels to some degree like the real world. People in this synthetic environment can move their head and eyes to explore it, feel the space that surrounds them and can interact with the objects in it. Simulated objects appear solid: they can be picked up, examined from all sides, navigated around, heard, touched and explored in many sensory ways.

Virtual reality has enormous potential in relation to teaching, learning and assessment. Using virtual reality systems it is possible for the learner to create images and objects which would take a long time to construct, or not be feasible to construct at all. Trainee architects – or their assessors – can walk around the buildings they design. Trainees in the printing industry can demonstrate their ability to operate expensive equipment without the risk of losing vast sums of money for their employers if they make a mistake. Those involved in the performing arts can use virtual reality for stage lighting and stage design courses.

Conclusion

We have discussed a diverse range of assessment strategies. All types have their strengths and weaknesses and it is likely that each method disadvantages some students in some way. It is preferable, therefore, for teachers and trainers to include in their repertoire a range of assessment methods. They will need to support educational aims, though the choices may be restricted because of resource or time constraints. Variety is as important for effective assessment as it is for all other aspects of teaching and learning.