

14 Assessment

It is important to develop a shared vocabulary in order to plan and evaluate assessment. Without this, positive and well supported initiatives such as Assessment for Learning can become ritualized into a set of habitual practices, without a full enough understanding of the purpose.

Some of the following will be common knowledge, but a quick review of definitions may be helpful.

Summative and formative

When we think of assessment, our first response is often exams. These are a good example of summative assessment designed to evaluate, record and recognize levels of achievement at the end of a period of study.

Formative assessment, on the other hand, is designed to provide ongoing feedback to learners on they are learning, which may result in a change of strategy, revision, recommendations to practice more, and so on.

Although future learning can be planned, and access to or choice of courses granted, in response to summative assessment, formative assessment gives closer and more immediate feedback to the learner. It is generally in the form of a verbal analysis and evaluation, without being reduced to a grade or level.

A particular type of formative assessment may be *diagnostic*, i.e. designed to identify specific difficulties in need of attention, blockages to further progress, etc.

Summative assessment generally concentrates on the product i.e. the result of learning. Formative assessment can involve either product or the learning process.

Formative assessment is more likely to be informal and embedded in day to day classroom events, for example noticing how a pupil tackles a task by observing the process.

Reliability and validity

These terms are often confused. *Reliable* means: can we trust that the same conclusion will be reached by different assessors, that questions aren't biased towards a particular group (e.g. boys), that the question is fairly clear, and so on.

Valid means: the test or performance really does assess what it is meant to. For example, writing an essay about swimming wouldn't prove that you can propel yourself across water. This may seem an extreme example, but the question of validity needs serious thought. Clarity in academic writing is a general requirement, and may distort assessments about physical education, art or mathematics. Many tests which claim to assess reading actually only cover a limited aspect, for example they depend on pronouncing the words aloud rather than understanding them; they might involve pronouncing individual words out of context, rather than making sense of a continuous text.

Both reliability and validity are critical for both summative and formative assessment, and at all ages. When a teacher declares a six year old to be 'more able' than another, it is essential to know that this is based on more than neatness or accent. (Indeed, the word 'able' is particularly dangerous as it can suggest you are saying something about innate intelligence; in reality, it is impossible to make a reliable or valid judgement about raw ability because each pupil has had different experiences in their early lives which impact on school learning.)

Norm referencing and criterion referencing

Traditionally pupils were assessed in terms of rank position or grade. To learn that 'Fiona came third in her class' in French tells us nothing about how good she is in terms of the national population or whether she can buy ice cream. Learning that she got a B at GCSE resolves the first problem but still doesn't tell us much about what he can do. It certainly doesn't reveal strengths and weaknesses.

So norm referencing is useful for summative assessment, at a general level, but not for formative.

In recent decades a greater emphasis has been placed on 'criterion referencing', i.e. judgements about specific knowledge or skills. A good example is the driving test, where you have to be able to carry out particular manoeuvres.

Criterion referencing is essential for formative assessment. It has also been applied to summative assessment which is matched to particular criteria, though (as stated in the previous paragraph) the specific information is often lost when stirred together to produce a composite final grade.

However, there are other limitations to criterion referencing. It is easier to write precise criteria for knowledge or skills than for creativity or appreciation (music, drama) or for ethical or social development.

It can be problematic to use criterion referencing to map progression:

- i) learning is not always linear, and sometimes a person can complete a task in one circumstance but not another;
- ii) when similar actions are used in different grades, too much depends on unclear qualifiers (adjectives and phrases such as appropriate, a variety of, clear)

Functions of assessment

One very clear function of summative assessment is to serve as the gatekeeper for particular careers or further study. It is worth questioning whether this is always valid, but also whether our education system depends too much on sorting and grading. Formative assessment, on the other hand, serves to provide feedback to pupils and teachers and influence the conduct of teaching and learning.

There are various audiences:

teachers, so that they know how well pupils are learning, and perhaps how successful their teaching has been;

parents, for information on their children's progress, and perhaps leading to greater assistance or pressure on their children;

employers, whether to show a general level of achievement or specific skills;

universities, to decide who to admit;

pupils themselves, to give them feedback.

Often there is a disjunction between the assessment tool and the intended purpose. For example, giving a pupil a low grade or mark says little about what to do about it, and may be simply demoralizing.

Authentic assessment

This phrase has been used to show that there are alternatives to assessment as a separate task. Authentic assessment involves activities which are worthwhile in their own right. It can be used summatively or formatively, though it is more common in informal formative assessment e.g. observing the activities of nursery children. (See also Rich Tasks, in section 12)

Accountability

Assessment has increasingly been used to judge other things beyond children's learning, for example to make judgements about a teacher's competence or a school's 'effectiveness'. It is important to consider the distorting effect this may have on teaching. For example, teachers may focus on the test to the detriment of other important learning, or pupils may become very short-term in their learning and even lose interest. This has become known as 'high-stakes testing' to point out that it is doing more than assessing individual learners, in ways which can have unforeseen consequences.

Evaluating assessment

- A. Is the assessment method fit for purpose, i.e. formative / summative, reliable / valid?
- B. Are there any aspects which may stand in the way of fair assessment, such as the impact of untidy writing, an angry teacher or a crisis at home?
- C. Is the teacher using formative assessment in ways which give clear feedback to pupils and influence their learning?
- D. Is formative assessment providing feedback to the teacher, leading to adjustments in teaching?
- E. Is summative assessment (including high-stakes uses for accountability) distorting learning or distracting from formative feedback (e.g. pupils reacting to grades rather than feedback)?
- F. Is sufficient use being made of informal process-based assessment?
- G. Is assessment manageable, and not dominating or distorting teaching and learning?
- H. Are possible negative consequences of assessment avoided, e.g. demoralization?