

Assessment: general points

Meaning:

Assessment is a process that involves

- collecting evidence in a planned and systematic way
- interpreting the evidence to produce a judgement
- communicating and using the judgement

Notes

- Assessment is a broad term and goes far beyond formal and informal testing; it includes judgements made in the course of teaching, marking work, in fact any situation where evidence is gathered and turned into a mark, remark, grade, number, etc that indicates that some judgements has been made of it.
- Assessment always involves three aspects: the collection of evidence, the interpretation of that evidence to give a judgement and some way of communicating the judgement.

Does this accord with your view of assessment?

Purposes:

- To help students' learning (formative assessment)
- To summarise achievement at certain times (summative assessment)
- To evaluate the effectiveness of teaching (where the focus of interest is the class not the individual)
- To monitor the performance of students across a state, region or nation (where only a sample of students is assessed)
- To assist in research or evaluation of curriculum materials

Notes

- The wide range of purposes makes it evident that what is assessed and the way in which the assessment is done has to fit the purpose.
- In many cases the purpose is best achieved when other information that student performance is also collected and used.

Can you think of other purposes of assessment?

Methods

Information can be collected by

- observing students carrying out their regular class work (observing includes listening, questioning and discussing)
- studying the products of their regular work (including drawings, artefacts, writing)
- introducing special activities into students' class work (eg concept mapping, diagnostic tasks)
- giving tests (teacher-made and external; performance and written)

electricity would not give a valid assessment of understanding of a simple electric circuit, although it would be an assessment of quite high reliability.

- Reliability and validity are not independent qualities of assessment. Attempts to increase reliability tend to restrict the range of evidence that can be used and hence reduce validity. Similarly, attempting to increase validity often means extending the range of the assessment to learning outcomes such as attitudes and values, which do not lend themselves to reliable assessment.

The characteristics of formative and summative assessment

Using the terms just discussed we can describe the characteristics of the assessment for the two purposes that are our concern in this module.

Characteristics of formative assessment

- It is an integral part of teaching for understanding. It is not optional; it cannot be taken away without changing the whole nature of the teaching and learning
- It relates to progression in learning but takes individual progress and effort into account; that is, it is both criterion-referenced and child-referenced
- It leads to action that supports further learning
- It can be used in all learning contexts
- It provides information about all learning outcomes
- It involves students in assessing their performance and deciding their next steps
- It uses methods that protect validity rather than reliability.

Characteristics of summative assessment

- It takes place at certain intervals when achievement is to be reported
- It relates to progression in learning judged against publicly available criteria
- It give results for different students that are based on the same criteria and so can be compared and combined if required
- It requires methods that are as reliable as possible without endangering validity
- It involves some quality assurance procedures.

Act 4. Looking closely at formative assessment

Notes

- These are methods of collecting evidence, they should *not* be described as methods of *assessment*, for the reason that assessment involves two other steps.
- Tests have many disadvantages for certain purposes, but they are useful for others. The pros and cons often relate to the relative importance, for a particular purpose, of reliability and validity. (these are terms that should relate to the whole process of an assessment and not just to the methods of gathering information – see later).

Can you add other ways in which evidence can be gathered for assessment?

Basis for making a judgement

In making a judgement the evidence is compared with some standard or criterion. This can be

- the standard of performance of other students of the same age and/or experience (norm-referenced)
- the description of a certain level of performance (criterion-referenced)
- the previous performance of the student (child-referenced or ipsative)

Notes

- Since an assessment always involves a judgement it is important to be clear about how it was made.
- It is particularly important in classroom assessment by teachers to be clear as to whether a student's performance is being compared with his or her previous performance (ipsative) or with a standard or criterion against which all students are being judged (criterion-referenced). Both are important but they serve different purposes.

What basis are you using for judging work when you write or give a verbal comment on a student's work and when you give a grade or number?

Reliability and validity of different methods of assessment

When different ways of collecting information and judging it are combined, we get the different *methods of assessment*, eg

- standardised tests are special tasks, where the judgement is norm-referenced, and the result communicated as a score compared with the average performance;
- class tests usually indicate what someone can and cannot do and so are criterion-referenced;
- marks on regular work can be either criterion-referenced or ipsative;
- on-going classroom observations are usually a mixture of ipsative and criterion-referenced and lead to judgements about action to take, as in the formative assessment cycle.

These vary in relation to *reliability* and *validity*.

- Reliability is the term used to express the degree of accuracy of the result of assessment, i.e. if the assessment were to be repeated, the extent to which the second result would agree with the first. The reliability of assessment of a student's skill in writing is likely to be less than that of a test of addition and subtraction sums, mainly because it is far easier to mark the latter consistently.
- Validity refers to the extent to which what is assessed really reflects what it was intended to assess. For example, a multiple-choice test of knowledge about materials that conduct