

Unit 4

Testing and Evaluation

In this unit we will cover:

- How to assess learners' progress.
- Summative & formative assessment.
- Internal assessment.
- External assessment.
- Subjective Vs. objective assessment.
- The effect of “backwash.”
- The concepts of reliability and validity.

How to assess learners' progress

- Classwork
- Q&A in class
- Homework
- Internal examinations/assessments
- External examinations/assessments

Summative & formative assessment

In simple terms, summative assessment occurs at the end of a course. Formative assessment takes place throughout a course. Here are two examples:

- A-levels are summative. You take an exam at the end of a course of study.
- UK universities these days assess formatively. You are continually assessed.

Strengths & weaknesses:

- For summative assessment, you have to show all you know in a short period of time (e.g. a 3-hour examination). However, the chances of cheating are remote.
- Formative assessment tests the whole person, but is open to abuse.
- Many organisations today like to use both assessment methods (e.g. DELTA).

Internal assessment

A good teacher should be assessing his/her students informally all the time.

- Formal assessment is internally done by means of tests, classwork and homework.
- Many schools will also use internal assessments modelled on external ones.
 - Trial or “Mock” exams is a good example.
- Schools should also assess the teaching quality of staff.
- This is done via observation by management, student questionnaires, etc.
- Schools should also be open to outside inspection (e.g. OFSTED).

External assessment

Examinations are set by an external awarding body (e.g. Cambridge, TOEFL, etc.):

- Strict rules and regulations.
- Random, unannounced inspections.
- Tight security.
- Schools must have the opportunity to challenge results.
- Where different assessment instruments are employed, costs rise.
- Quality assurance systems are mandatory – externally & internally.
- Why do many educational institutions use external assessment?

Subjective Vs. objective assessment

Some authorities distrust subjective assessment. They argue that only objective tests are valid. The latter tests can give a score that is statistically very accurate. However:

- Most objective tests consist of multiple choice questions (MCQs).
- A notable exception is Mathematics.
- If student responses are classified as “right” or “wrong,” objective tests are fine.
- This becomes impossible for most arts and social science subjects.
- Marking guides are required to show markers a range of scores.
- The issue becomes one of how you quantify a quality assessment.

Backwash

What is Backwash?

The backwash effect of testing is when teaching is done solely so that students can pass a test. It doesn't necessarily mean they have learned anything. When this happens, it is usually referred to as negative backwash. Consider this:



Please read this carefully:

**He was crambling with his
tollocks when his autochoyns
simplided. This was because his
oscillating spangers had not been
callibrated or oxymoronogised.**

Now answer these questions

He was crambling with his
tollocks when his autochoyns
simplided. This was because his
oscillating spangers had not been
callibrated or oxymoronogised.

1. What was he crambling with?
2. What happened when he did?
3. Why was this?

Backwash

Avoiding negative backwash

Backwash is not necessarily negative – we just need to be aware of it. Whenever we are setting tests, we need to ask ourselves the question:

“What use is this test to my students?”

If you can't satisfactorily answer the question, but the students all do well in the test, what exactly have they achieved? Answer: they have learned how to pass a useless test!

This leads us on to consider the concepts of reliability and validity.

Reliability and validity

Reliability

This occurs when you can be sure that the test you are administering is fair and that everyone is treated exactly the same. You can be sure the answers will reflect the ability of a group of students. However, it ignores the effect of backwash.

Validity

Is the concept of actually testing what is required to increase student's knowledge. There are several of kinds validity including "construct validity," "content validity" and "face validity." A useful portfolio submission would be for you to write a blog post or an article describing what these terms precisely mean.

Reliability and validity

To conclude:

Sometimes, it is easier to get your head round a difficult concept by giving a concrete example. I have always liked this one which explains the difference between reliability and validity:

Using a tape measure around a person's waist to measure his height is a reliable test. However, it is not remotely valid.

(Think about it!)

Reliability and validity

Assessment in the Age of AI

- Students increasingly use AI tools to support learning and research.
- Teachers should distinguish between appropriate support and academic dishonesty.
- Assessment tasks should encourage critical thinking, analysis and personal reflection.
- Oral presentations, project work and authentic tasks can help demonstrate genuine understanding.
- The aim of assessment remains unchanged: to measure learning in a valid and meaningful way.

Suggestions for further study

What criteria must you use when you want to:

- set test and homework assignments?
- explain why they were set and what you expected from students?
- know effective were they? Reflect upon it/them?

Activities to consider:

- Read as much as you can about assessment methodology.
- What kind of assessment instruments work best with your classes? Why?
- Discuss different assessment strategies with a colleague from another discipline. What are the main differences?

