Creating valid assessments for Curriculum for Excellence

Dr Alastair Pollitt, Director, Cambridge Exam Research

Creating valid assessments for new National Courses and Units

What's important?

An assessment is a tool, and like any tool, it is meant to serve a purpose, such as to support learning, to inform parents, or to summarise learning. What matters most is that each assessment should satisfy the purpose, or purposes, for which it is needed. This is what we mean by **validity**.

Validity

Validity cannot be created just by looking at test results — the potential for validity must be designed in from the beginning. Assessment is a process, with several steps, and at each one of them, some of that potential will be lost through the compromises that must be made. Everyone who takes part, whether curriculum specialist, qualification developer, statistician, parent, teacher, lecturer or potential employer, shares the responsibility to keep the validity as high as possible.

The 'bucket brigade'

Imagine potential validity as the water in a fire-fighter's bucket. There should be as much as possible in it at the start, and as the bucket is passed from hand to hand — under pressure to be as quick as possible — everyone tries to keep as much water in it as possible. If even one person spills too much, the bucket will not very well serve its purpose in helping to put out the fire.

The steps in the process

There are at least seven different steps in the making and using of an assessment tool. For the present purpose, we will concentrate mostly on the first four of them.

Assessment consists of gathering, evaluating and interpreting evidence of what is important in learning. We begin with the concept of **importance**.

1 Agreeing what is important

Key participants: The subject community, especially the teachers, lecturers and colleges and local authorities

Getting started properly is critical. Everyone involved needs to know, and to accept, a clear description of what really matters in the teaching, learning and assessment of a particular subject. Introductions to National Courses and Units

give a good, concise statement of what is important, of the value of learning a subject, or of what young people can achieve through studying it. Let's call this an **importance statement** (IS).

The importance statement concisely sums up the purpose of learning, in a form that every learner and assessor should be able to remember. In particular, assessors should constantly — almost instinctively — judge every decision they have to make against it. Here are a few key issues:

- For assessment, the importance statement needs to define importance in the subject being assessed; thus, for example, a team designing an examination for Dance may need to write their own specific IS, based on the statement for Expressive Arts in general.
- Such an importance statement must be made public, for teachers, students, and others with an interest in the assessment of Dance.
- However, higher level importance statements — for subjects, a curriculum area, perhaps, or for all curriculum areas — should never be forgotten.
- An importance statement should apply to the whole subject; to ensure continuity, it should not vary much, if at all, for different phases or levels.

2 Specifying the test

Key participants: Curriculum and assessment specialists

The ideal assessment is never practicable: inevitably, compromises must be made. Whenever you face an awkward decision, try to resolve it as: Given the constraints of time, cost and practicality, how best can I measure what's important?

Some important elements, such as 'enhance their enjoyment' cannot be assessed, but the three key expectations — breadth, level of challenge, application — all can, and must, be.

- The job of the specification is to define the evidence the assessors must seek.
- Assessing breadth means a range of content must be compulsory.
- The assessment must match the level of demands, or challenge, required by the SCQF level.
- There may be many more constraints, of content or style, or caused by the availability of equipment and materials, or to avoid bias or unfairness, that influence what evidence can be elicited: the criterion must always be the importance statement.

3 Writing the items

Key participants: Question writers and scrutineers

A 'question' is the words used to communicate the task to the learner; in this document we use 'item' to refer to the question and the marking instructions to the Marker; the 'task' is the actions the candidate is required to do. 'Task'

includes written responses, performance tasks, or practical, project or any other assessable work that can provide evidence of achievement.

Writers face many challenges to create a test that keeps the validity as high as possible. Here are some examples, to show how the general principle of concentrating on the evidence of what's important can be applied:

- It is very important that writers keep the importance statement in mind at all times, to ensure that each item measures some of what is important, and that the whole assessment measures as much as possible of it.
- This implies a broad level of consistency in both the coverage of content and demands from test to test; it does not imply two versions should be the same in detail.
- The job of an assessment (the text, diagrams etc that appear on a test paper) is to convey the required task to the learner. The language should be as natural and clear as possible. This does not mean only using 'simple' sentences.
- Use emphasis, such as bold, in a question to help candidates understand the task; do not follow 'conventions' in using it.
- Make sure the question tests what's important, not just reading or irrelevant skills.
- In writing a question, pretend the candidates are two years younger than they actually are — this helps deal with the anxiety many will suffer during examinations.
- Be careful not to trick candidates with 'clever' questions, or with humour or irony: stress will make their minds much more susceptible to error or distraction.
- Use a limited range of command words — the words like 'sketch', 'describe' or 'explain'.
- When testing 'application', be very careful with contexts, which often introduce biases: if some candidates are familiar with the context, it may not test application at all.
- The job of marking instructions is to evaluate the evidence for what's important.
- Make sure the marking requires and credits what the command word asked for.
- Make sure the marking gives only credit for what is considered important in the subject.
- Avoid 'points'-based marking schemes for questions that ask for explanations, or any other response that should be judged qualitatively.
- The marking instructions should help Markers with the difficult decisions, not just tell them the right answer. Don't just give good answers, but poor and average ones too.
- Try to define the boundaries between 0 and 1, 1 and 2, etc. Be explicit about what is required to gain extra marks; use principles rather than (or as well as) examples to show this.

4 Communicating with learners

Key participants: Assessment managers

Assessment involves a triangle of Setter (question writer), learner, and Marker. Many problems in exams are caused by a failure of the communication from Writer to learner or from learner to Marker. Some of these can be avoided in advance:

- Make sure teachers and learners know what the assessment will be like; but don't make it too predictable, since the skills you are assessing may then not be the ones that are important.
- In particular, if quality is more important than correctness, make sure that the scoring criteria are known to the teachers, and that they are consistent with the importance statement. It's only fair for learners to know how they will be judged.
- Make sure the command words will be understood by the candidates; consider providing a glossary of them, with examples, for teachers.

5 Scoring

Key participants: Markers and judges

More validity will be lost if the communication from the Writer to the Marker is not perfect. In normal life, the person who asks a question knows what sort of answer they are looking for: in exams, the Marker needs to be told what range of evidence the Writer was expecting to elicit:

- Make sure the marking instructions describe the variety of responses the question elicits — good and poor — and do not just show a perfect answer. This will help ensure consistent marking.
- Ensure that Markers pay more attention to what is important than to mere correctness or to trivial errors.
- Make sure a single error, repeated several times by a candidate, is only punished once. This applies equally to all curriculum areas.

6 Grading

Key participants: Assessment Managers and authorities

7 Interpreting

Key participants: Teachers, learners, parents, 'next phase', employers

These last two stages are not really relevant here, except to note that they cannot deliver valid assessment if the scores are not valid — if the bucket is already empty.

Conclusion

Maximising the validity of assessment is everyone's responsibility. From the initial idea for an assessment, through specifying what it should contain, choosing tasks, writing questions (including the marking instructions), and making sure the

candidates are suitably prepared, through scoring, and grading, to the final stage of interpreting the results — no one can duck their responsibility for validity.

When it is your turn to play your part in the assessment process: Don’t pass the buck — pass the bucket.