

Basic assessment issues and terminology

Chris Rust

Summary

This introduction to basic assessment issues and terminology is aimed at those who are relatively new to teaching and who have little or no experience in assessing students' work. Before designing a course or undertaking any student assessment, it is important to clarify some of the fundamental principles and issues which need to be applied. For this purpose, let us define assessment as evaluation or appraisal; it is about making a judgement, identifying the strengths and weaknesses, the good and the bad, and the right and the wrong in some cases. It is more than simply giving marks or grades, although that may well be a part of it. And because it involves making a judgement it will almost always inevitably include an element of subjectivity by the assessor. However, we should strive to make assessment as objective, fair and transparent as possible.

Assessment plays a crucial role in the education process: it determines much of the work students undertake (possibly all in the case of the most strategic student), affects their approach to learning and, it can be argued, is an indication of which aspects of the course are valued most highly.

Chris Rust, 2000

Much of this was first published as "Guide to assessment", in *Assessment strategies for Pop Music Performance*, University of Salford Dept of Music, FDTL Project POP - CD-Rom.

Biography

Chris Rust is Principal Lecturer and Senior Development Officer in the Oxford Centre for Staff and Learning Development at Oxford Brookes University, and a Fellow of the Staff and Educational Development Association. He has researched and published on a range of issues including: the experiences of new teachers in HE, the positive effects of supplemental instruction, ways of diversifying assessment, and the effectiveness of workshops as a method of staff development.

Keywords

Purposes of assessment, formative, summative, reliability, validity, criterion v norm referenced assessment, assessment criteria, types of assessment, CAA

Purposes of assessment

It is easy to become so immersed in the job of teaching that we lose sight of the exact purpose of a particular element of assessment. There is then the possibility that we are not achieving that purpose, or that we overlook another form of assessment which might be more appropriate. We actually assess students for quite a range of different reasons - motivation, creating learning opportunities, to give feedback (to both students and staff), to grade, and as a quality assurance mechanism (both for internal and external systems). Because all too often we do not disentangle these functions of assessment, without having really thought it through assessments are frequently trying to do all these things,

to varying degrees. In fact it is arguable that while it is desirable for assessments meeting the first three of these functions to be conducted as often as possible, the final two do not need to be done anywhere near so frequently; it is simply important that they are done somewhere. The implications of this are that while an essay question, where all the answers are double marked and the marks count towards the students' final grades, may fulfil all these functions, for all assessments to be this rigorous would be prohibitively expensive in staff time, while a peer-assessed seminar presentation, which does not count towards the students' final grades but is simply a course requirement, could fulfil the first three functions and may not even require a tutor to be present.

Formative v Summative assessment

This is the distinction between assessment which is mainly intended to help the student learn and assessment intended to identify how much has been learnt. Formative assessment is most useful part way through a course or module, and will involve giving the student feedback which they can use to improve their future performance. In practice, to varying degrees, most forms of assessment probably try to do both although the end of course exam where the only feedback received is a mark is almost totally summative. It is arguable that assessment in British higher education is too often focused on the summative, and the accumulation of marks, coming at the end of courses, while students would benefit from more opportunities to build on their strengths and learn from their mistakes through the feedback from formative assessment activities staged throughout their course or module.

Assessment and Course Design

Assessment should be seen as an intrinsic part of the learning process rather than something which is just 'tacked on' at the end in order to get some marks. It should therefore be seen as a vital part of the initial design of the course or module. A model of course design can be described in the following three stages:

Stage 1: Decide on the intended learning outcomes. What should the students be able to do on completion of the course, and what underpinning knowledge and understanding will they need in order to do it, that they could not do when they started? (This obviously begs the questions what have they done before and what prior ability and knowledge can you expect?). These learning outcomes should each be described in terms of what the student will be able to do, using behavioural verbs, and described as specifically as possible. (Verbs like 'know' and 'understand' are not helpful because they are so general. Ask yourself, "What could the student do to show me that they know or understand?"). You may find it useful to group your outcomes under the following four headings: skills (disciplinary), skills (general), values and attitudes, underpinning knowledge and understanding.

Stage 2: Devise the assessment task/s. If you have written precise learning outcomes this should be easy because the assessment should be whether or not they can satisfactorily demonstrate achievement of the outcomes.

Stage 3: Devise the learning activities necessary (including formative assessment tasks) to enable the students to satisfactorily undertake the assessment task/s.

These stages should be conducted iteratively, thereby informing each stage by the others and ensuring coherence.

Principles of assessment

Reliability

If a particular assessment were totally reliable, assessors acting independently using the same criteria and mark scheme would come to exactly the same judgement about a given piece of work. In the interests of quality assurance, standards and fairness, whilst recognising that complete objectivity is impossible to achieve, when it comes to summative assessment it is a goal worth aiming for. To this end, what has been described as the 'connoisseur' approach to assessment (like a wine-taster or tea-blender of many years' experience, not able to describe exactly what they are looking for but 'knowing it when they find it') is no longer acceptable. Explicitness in terms of learning outcomes and assessment criteria is vitally important in attempting to achieve reliability. They should be explicit to the students when the task is set, and where there are multiple markers they should be discussed, and preferably used on some sample cases prior to being used 'for real'.

An additional advantage of having explicitly stated criteria is that they can be used to create assignment attachment sheets. This is a sheet attached to the student's work which tabulates the criteria and enables feedback to be given in a mechanistic way. Assignment attachment sheets have the dual benefit of reducing the marker's time which otherwise would have been spent writing repetitive comments, while focusing the attention of both the tutor and student on how well specific criteria have been met.

Validity

Just as important as reliability is the question of validity. Does the assessed task actually assess what you want it to? Just because an exam question includes the instruction 'analyse and evaluate' does not actually mean that the skills of analysis and evaluation are going to be assessed. They may be, if the student is presented with a case study scenario and data they have never seen before. But if they can answer perfectly adequately by regurgitating the notes they took from the lecture you gave on the subject then little more may be being assessed than the ability to memorise. There is an argument that all too often in British higher education we assess the things which are easy to assess, which tend to be basic factual knowledge and comprehension rather than the higher order objectives of analysis, synthesis and evaluation.

Relevance and transferability

There is much evidence that human beings do not find it easy to transfer skills from one context to another, and there is in fact a debate as to whether transferability is in itself a separate skill which needs to be taught and learnt. Whatever the outcome of that, the transfer of skills is certainly more likely to be successful when the contexts in which they are developed and used are similar. It is also true to say that academic assessment has traditionally been based on a fairly narrow range of tasks with arguably an emphasis on knowing rather than doing; it has therefore tended to develop a fairly narrow range of skills. For these two reasons, when devising an assessment task it is important that it both addresses the skills you want the student to develop and that as much as possible it puts them into a recognisable context with a sense of 'real purpose' behind why the task would be undertaken and a sense of a 'real audience', beyond the tutor, for whom the task would be done.

Criterion v Norm referenced assessment

In criterion-referenced assessment particular abilities, skills or behaviours are each specified as a criterion which must be reached. The driving test is the classic example of a criterion-referenced test. The examiner has a list of criteria each of which must be satisfactorily demonstrated in order to pass - completing a three-point turn without hitting either kerb for example. The important thing is that failure in one criterion cannot be compensated for by above average performance in others; neither can you fail despite meeting every criterion simply because everybody else that day surpassed the criteria and was better than you.

Norm-referenced assessment makes judgements on how well the individual did in relation to others who took the test. Often used in conjunction with this is the curve of 'normal distribution' which assumes that a few will do exceptionally well and a few will do badly and the majority will peak in the middle as average. Despite the fact that a cohort may not fit this assumption for any number of reasons (it may have been a poor intake, or a very good intake, they have been taught well, or badly, or in introductory courses in particular you may have half who have done it all before and half who are just starting the subject giving a bimodal distribution) there are even some assessment systems which require results to be manipulated to fit.

The logic of a model of course design built on learning outcomes is that the assessment should be criterion-referenced at least to the extent that sufficiently meeting each outcome becomes a 'threshold' minimum to passing the course. If grades and marks have to be generated, a more complex system than pass/fail can be devised by defining the criteria for each grade either holistically grade by grade, or grade by grade for each criterion (This is elaborated on further below; but for an example of a criterion grid, and even more on designing and using criteria, visit <http://www.brookes.ac.uk/services/ocsd/fw24.html>).

Writing and using assessment criteria

Assessment criteria describe how well a student has to be able to achieve the learning outcome, either in order to pass (in a simple pass/fail system) or in order to be awarded a particular grade; essentially they describe standards. Most importantly they need to be more than a set of headings. Use of theory, for example, is not on its own a criterion. Criteria about theory must describe what aspects of the use of theory are being looked for. You may value any one of the following: the students' ability to make an appropriate choice of theory to address a particular problem, or to give an accurate summary of that theory as it applies to the problem, or to apply it correctly, or imaginatively, or with originality, or to critique the theory, or to compare and contrast it with other theories. And remember, as soon as you have more than one assessment criterion you will also have to make decisions about their relative importance (or weighting).

Graded criteria are criteria related to a particular band of marks or honours classification or grade framework such as Pass, Merit, Distinction. If you write these, be very careful about the statement at the 'pass' level. Preferably start writing at this level and work upwards. The danger in starting from, eg first class honours, is that as you move downwards, the criteria become more and more negative. When drafted, ask yourself whether you would be happy for someone meeting the standard expressed for pass, or third class, to receive an award from your institution. Where possible, discuss draft assessment activities, and particularly criteria, with colleagues before issuing them.

Once decided, the criteria and weightings should be given to the students at the time the task is set, and preferably some time should be spent discussing and clarifying what they mean. Apart from the argument of fairness, this hopefully then gives the student a clear idea of the standard they should aim for and increases the chances they will produce a better piece of work (and hence have learnt what you wanted them to). And feedback to the student on the work produced should be explicitly in terms of the extent to which each criterion has been met.

Types of assessment

As has been argued already, the type of assessment chosen should be related to learning outcomes and governed by decisions about its purpose, validity and relevance. In addition, as it is probably true to say that every assessment method will place some students at a disadvantage to some extent, a range of types of assessment is desirable to hopefully reduce the element of disadvantage suffered by any particular student. Types of assessment to choose from include:

Essay

An answer to a question in the form of continuous, connected prose. The object of the essay should be to test the ability to discuss, evaluate, analyse, summarise and criticise. Two dangers with essays are that they are easy to plagiarise, and that undue weight is often given to factors such as style, handwriting and grammar.

Assignment

A learning task undertaken by the student allowing them to cover a fixed section of the curriculum predominantly through independent study. Different methods of presenting the results can be used dependent on the nature of the task - a report (oral or written), a newspaper or magazine article, a taped radio programme, a video, a poster, a research bid, a book review, a contribution to a debate, etc. It is vital to be clear in the assessment criteria how important the medium is compared with the message, so if it is a video how important is the quality of the lighting, the style of the editing, etc, compared with the content that is covered. If aspects of the medium are important then time must be given in the course for these to be taught.

Individual project

An extended investigation carried out by an individual student into a topic agreed on by student and assessor. In many ways similar to an assignment, the main difference is the onus on the student to choose the particular focus and/or medium of presentation. As with any assessment where the product will vary significantly from student to student it is vital that the criteria are sufficiently well written to be fair when applied to different undertakings and results.

Group project or assignment

Where either an assignment or project is undertaken collectively by groups of students working collaboratively. This has the pragmatic advantage of potentially reducing the tutor's assessment workload and the educational advantage of helping to develop the students' teamworking skills. There are also some forms of product such as collaborative performance that can by definition only be achieved in a group. The major assessment problem is how to identify each individual's role and contribution and to reward it fairly. Solutions (none of which is problem free) tend to include combinations of: an individual component which can be individually assessed, tutor observation, and involving the students in some self and/or peer assessment as the ones in the best position to judge.

Dissertation

Written presentation of results of an investigation or piece of research, normally taking the form of an extended essay being less rigorous in its style and layout requirements than a thesis. The content reflects the findings of the investigation. This has similar assessment problems to an individual project.

Examination

This can take a variety of different forms. The most common factors are that it is done under comparatively short, timed conditions and usually under observed conditions which ensures it is the student's own work (although there are examples of exams where students take the questions away). Major criticisms are that because of the comparatively short time allowed answers may inevitably be superficial and/or not all the learning outcomes may be assessed. They may also encourage the rote learning of potential model answers. This can be avoided if the focus of the tasks set is on the application of what has been learnt, presenting the student with a previously unseen context or scenario or set of data which they have to 'do' something with. Some of the most common variations of exams are:

- 'seen' where the questions to be answered are given at a pre-specified date beforehand. The intention is to reduce the need for 'question-spotting', to reduce the anxiety, and to increase the emphasis on learning.
- 'open-book' during the exam students have access to specified texts and/or their notes. the intention is to reduce the emphasis on memorising facts, to reduce anxiety and allows more demanding questions to be set.
- 'unseen' arguably makes the student revise the whole syllabus because anything may appear on the paper (although in practice may do the opposite as the student may 'question-spot' and gamble on certain topics coming up.
- 'MCQ' objective tests asking multiple choice questions (MCQ) where the student simply selects from a bank of potential answers. Easy to mark (can be done by a machine or even administered on a computer) and can ensure students revise the complete syllabus but arguably difficult, if not impossible, to assess higher order skills. Writing good questions is however very difficult. If you can find an appropriate US textbook there will probably be a bank of questions which come with it on disk.

Viva

Possibly used in conjunction with any of the above methods, this involves the student having to answer questions orally. In a comparatively short space of time it is possible to ascertain both what the student knows and the depth of this understanding (and possibly the amount they contributed to a group project and the nature of that contribution).

Performance

In many cases, when it comes to practical outcomes, the only sensible way of really assessing whether an outcome has been learnt is through watching the student actually perform it - whether 'it' is literally a performance, as in the performing arts, or a nursing student taking a patient's blood pressure. Because in such cases the assessed "product" is transient, for purposes of moderation and external validation you may need to find ways of recording the event (audio or video). Such recordings can also play a vital role in giving the student feedback.

Self and peer assessment

There is strong evidence that involving students in the assessment process can have very definite educational benefits. Not so much a type of assessment like those already listed, this is something which can be done in conjunction with any type of assessment. The important aspect is that it involves the student in trying to apply the assessment criteria for themselves. This might include: a marking exercise on 'fictitious' or previous years' student work; the completion of a self-assessment sheet to be handed in with their work; 'marking' a peer's work and giving them feedback (which they can then possibly redraft before submission to the tutor); or really marking other students' work (i.e. allocating marks which actually count in some way) - a seminar presentation, for example, or a written product using a model answer. The evidence is that through trying to apply criteria, or mark using a model answer, the student gains much greater insight into what is actually being required and subsequently their own work improves in the light of this. An additional benefit is that it may enable the students to be set more learning activities on which they will receive feedback which otherwise would not be set because of lack of staff time.

Computer-aided assessment (CAA)

Some people may resist the notion of computer-aided assessment because they do not believe it can validly assess the range of outcomes they are looking to test. This may be true (although very sophisticated multiple-choice questions can be quite challenging) but CAA does have the very definite advantage that it may significantly save staff time spent on assessment. If this enables you to introduce more assessed tasks, and therefore increases the amount of feedback students (and you) get on their learning, and possibly also helps them to pace their learning, it may well be worth considering. For more information, the best place to start is the CAA Centre website at <http://www.caacentre.ac.uk>.

Other Websites

In the UK, a useful website to start at would be:

<http://www.lgu.ac.uk/deliberations/assessment/index.cgi>

In the USA, a number of organisations have developed statements that include references to fair assessment practices and include:

Leadership Statement of Nine Principles on Equity in Educational Testing and Assessment by the first National Symposium on Equity and Educational Testing, North Central Regional Educational Laboratory
www.ncrel.org/sdrs/areas/issues/content/contareas/math/ma1newst.htm

Nine Principles of Good Practice for Assessing Student Learning by the American Association for Higher Education
www.aahe.org/principles.htm

Writing Assessment: A Position Statement by the Conference on College Composition and Communication
<http://www.ncte.org/cccl12/sub/state6.html>

Further reading

John Biggs, Chapter 8 Assessing for learning quality in *Teaching for quality learning at university* 1999 Buckingham: SRHE/OUP

George Brown et al, *Assessing student learning in higher education* 1997 London: Routledge

Sally Brown, Phil Race and Brenda Smith, *500 Tips on assessment* 1996 London: Kogan Page

Sally Brown, Chris Rust and Graham Gibbs, *Strategies for diversifying assessment* 1994 Oxford: Oxford Centre for Staff and Learning Development

Sue Habeshaw, Graham Gibbs and Trevor Habeshaw; *53 Interesting ways to assess your students* 1993 Bristol: TES

David Newble and Robert Cannon, Chapter 6 Assessing the students, in *A Handbook for teachers in Universities and Colleges* 1995 London: Kogan Page

Paul Ramsden, Chapter 10 Assessing for understanding, in *Learning to teach in higher education* 1992 London: Routledge

Derek Rowntree, *Assessing students - How shall we know them?* 1977 London: Harper Row