5

Why Assess Innovatively?

Phil Race

Introduction

If traditional assessment practices and instruments were working perfectly, there would be no need for innovative assessment. With something as crucial as assessment, it is essential that innovation is not approached lightly, or engaged in for its own sake. The future careers and lives of students are at stake, and the scope for experimentation must be carefully delineated and planned. Before justifying the need to assess innovatively, it is important to be convinced that traditional methods are not serving the purposes for which they were designed. In this chapter, we take a fresh look at the role of assessment in higher education, including discussion of the various client groups that assessment should be serving, and the needs of these groups. We then explore how some traditional approaches to assessment are failing, in that they are not measuring the intended learning outcomes that we claim are important for students to develop, and are often promoting surface approaches to learning. The ways that students perceive their learning, and the ways that academics perceive their teaching, are all significantly dependent on the nature and formats of assessment. Therefore, changing assessment can best be justified if the quality of both teaching and learning are improved as a result, and if the assessment itself can be shown to be demonstrably fairer, and better related to the intended learning outcomes. It is argued that assessment must play a better part in the ways that teachers teach and learners learn.

I will conclude with checklist questions about some of the principal factors that can be adjusted either separately or in combination, to form the basis of innovations in assessment. The questions aim to be an aid to fuel productive innovation, to counter the dangers of overassessment, and help to ensure that assessment may be used to promote deeper approaches to learning.
Assessment for whom?

Higher education is often talked about using industrial and commercial metaphors. The sector has been described as being driven by market forces. Management systems which may work well in commerce and industry are increasingly transposed into higher education, however inappropriately. Students themselves are much better informed about their rights as consumers in this market, and their awareness and sensitivity to the performance criteria of educational institutions has never been more acute. Whether we like it or not, the most significant product of higher education is now the qualifications that students gain, rather than the quality of their learning experience. Part of the aim of this chapter is to point to ways of balancing assessment so that it enhances and enriches students’ learning experience, as well as leading to qualifications that are valid and appropriate for their future lives and careers. The client groups served by assessment are diverse, and it is relevant to look at the respective expectations of students themselves, the employers who will take them on when qualified, and other significant stakeholders in higher education such as the teachers who work in the system, subject reviewers who review the system, parents who increasingly pay for the system, and society in general for whom the system should be geared.

What should assessment do for students?

Ultimately, assessment should be for students. It should not only serve to ensure that their qualifications are valid and relevant to their career development, but also should be a formative part of their learning experience. Other chapters in this book explore in detail how assessment can deliver much more significant learning payoff to students than is often the case with familiar, traditional approaches. Various contributions to Knight’s (1995) book show the extent to which assessment practices can let students down. At present, students often feel that they are excluded from the assessment culture, and that they have to use trial and error to make successive approximations towards the performances that are being sought in assessed work. Students developing their technique the best tend to succeed in assessment, and it is not always the students who could have most to offer to the future development of the fields in which they are studying.

What do employers want from assessment?

Employers have a vested interest in assessment. Students’ achievements are the principal basis of selection for employment. Employers wish to be able to make well-informed decisions when short-listing candidates for interview. They also wish to be able to tell, from assessment data, whether students are going to turn out to be good employees, and not just subject experts. Employees need to be able to work together, to communicate effectively, and to think creatively, and it is not enough for them just to be knowledgeable about the subjects that they have studied. Employers, therefore, have a strong interest in what is assessed, as well as in how fairly the assessment practices and instruments function. Employers also want it to be easy for them to use students’ assessment data in short-listing and recruitment. This often means that they find it more acceptable to use degree results as a quick indicator of candidates’ suitability for posts, even when such results may be poor indicators of actual employability potential. Employers also tend to remember assessment as it was when they themselves studied, and this can make them more reluctant to think about the benefits that may be linked to innovations in assessment. It is therefore necessary to be able both to spell out these benefits to them, as well as to explain why the innovations are necessary in terms of what was wrong with some of the assessment formats they themselves encountered.

What do parents expect of assessment?

Parents remain a very significant factor in students’ learning. This is not least because the monetary aspect of their vested interest has increased significantly, as their financial contribution to students’ education has become more important. Parents increasingly subsidize the time their offspring spend in higher education, and are naturally concerned about all factors contributing to quality in teaching, learning and assessment, even though they are often distanced from the day-to-day experience of students. They need to be assured that assessment processes and practices are fair, equitable and valid, and that students are well informed about the rules of the assessment game. Parents too, however, may distrust innovations in assessment, especially if they believe that such innovations are made merely to save money.

What do teaching staff need of assessment?

For most teaching staff, designing and implementing assessment instruments and processes, and marking students’ work, are crucial parts of their day-to-day work. However, lecturers and tutors now have less face-to-face time with individual students, and often carry out most of their dealings with students in large-group situations. Their ability to base assessment results on really knowing students is reduced, and therefore they need to take steps to ensure that assessment practices and instruments are well designed and valid. In an educational system where the prominence of institutional and departmental performance indicators has increased dramatically, they need to ensure that assessment (which is one of the most public performance indicators) is measuring what it is intended to be measuring. With syllabus content expressed in terms of intended learning outcomes, and with standards associated with graduateness becoming increasingly public, the onus is
What does society require from assessment?

Taxpayers and electors are naturally concerned that funding spent on all aspects of education is well managed. With a much higher proportion of the population aspiring to degree qualifications, the level of interest in assessment has increased. The extent to which people regard themselves as stakeholders in education has increased accordingly. This leads to a greater demand that the outcomes of higher education are relevant to society in general, and that successful students are well qualified to contribute to society.

Why innovate? – there’s no need if it’s working!

One of the most significant dangers facing higher education is that of trying to replace things which aren’t broken. Any risks associated with innovations in assessment are far more serious than any dangers linked to innovations in teaching practices, learning environments or syllabus content. Assessment directly affects students’ immediate futures and their whole future careers. Even decades after students complete their last examinations, their assessment results continue to be scrutinised when they apply for promotion, or change jobs or career directions.

Before investing the time and energy necessary to introduce innovations in assessment, it is important to be convinced about the areas where traditional methods are not fulfilling their purposes. Those opposing innovations in assessment often refer to traditional methods as being tried and tested. This can be a strong argument against innovation. Anything new brings with it its own risks. However, it is equally valid to propose that when something is indeed tried and tested, there is a tendency to become blind to the shortfalls and weaknesses that are involved. McDowell and Sambell in Chapter 6 present interesting evidence of students’ reactions to innovations in assessment, showing that students themselves are all too aware of some of the shortcomings of traditional assessment methods.

There is widely shared concern that assessment does not measure the knowledge, skills and attributes that are intended to be assessed. In other words, assessment is often simply not valid. For example, traditional unseen examinations tend to measure students’ skills to perform in this particular environment, which is in most respects one far removed from both the working environment, and the lifestyle, for which higher education should be aiming to equip them. Similarly, assessment tends to be dominated by measurements of things which are relatively straightforward to assess, such as written answers, essays, reports and dissertations. The students who fare best in such assessments are those who develop those skills which relate directly to succeeding to extract marks from assessors marking such work. Again, what is actually measured may only be a pale reflection of students’ real learning.
I therefore continue this chapter with a short analysis of some of the principal weaknesses of two of the most familiar kinds of assessment, traditional unseen written exams and continuous assessment. In each case, I present first some of the tensions that can occur between these kinds of assessment and the quality of students’ learning, then move on to some of the operational weaknesses that can easily jeopardize the validity and reliability of these assessment formats, however safe, familiar and comfortable they may seem to be to assessors.

**Some failings of traditional unseen written exams**

Much has been written about the weaknesses of unseen traditional examinations. I do not wish to suggest that such examinations should be abandoned altogether, and elsewhere in this book are suggestions about ways of using many other forms of examination. Alternatives include open-book exams, time-unconstrained exams, take-away exams, multiple-choice exams, computer-delivered assessments, oral exams and assessed presentations. However, the unseen written exam continues to dominate many educational programmes. In particular, this assessment format seems to be at odds with the most important factors underpinning successful learning. For example, Brown et al. (1998) in an edited collection of articles on student motivation, demonstrate fundamentally how assessment can affect (and often damage) students’ enthusiasm to learn. Of particular significance are the chapters by Mortimer (Brown et al. 1998: 173-88), advocating the benefits of portfolios over traditional assessment methods, Newstead (ibid.: 189-200), presenting startling evidence about the frequency of cheating, and Leach, Neutez and Zepke (ibid.: 201-11) giving insight into the links between assessment and motivation from studies in New Zealand. Moreover, there is abundant evidence that assessors are not particularly good at making exams valid, reliable, or transparent to students. Some of the principal concerns that can be expressed are summarized below.

Firstly, there are tensions between examinations and the quality and depth of students’ learning experience. Exams do not do much to increase students’ motivation in terms of their want to learn. They may, however, cause students to need to learn, which is a significant driving force for learning, if not the happiest one. Students often make choices in modular schemes so that they avoid this kind of assessment if they can. This can lead them to choose subjects in which they are less interested than those which they fear to select because they will be subjected to such exams. Most exams are not ideal learning experiences in their own right. Though students may do a lot of learning before formal unseen written examinations, their actual experiences of learning in such situations is very limited.

Secondly, traditional exams are cause for concern in the context of a further important aspect of deep learning: feedback. The amount of feedback that students receive after most kinds of exam is not optimal. Feedback plays a vital role in learning, yet most systems require exam scripts to be regarded as secret documents, not to be shown to students on any account! Traditional unseen written exams can, in this situation, be said to be lost learning experiences. Exams do not do much to help students make sense of what they have learned. While during the time leading up to exams, there may be a significant amount of making sense of what has been learned, the assessment experience itself does little to help students to gain any further deepening of their knowledge or skills.

Thirdly, exams usually force students into surface learning, and into rapidly clearing their minds of previous knowledge when preparing for the next exam. Students are encouraged by the assessment regime, to clear their brains of the knowledge they have stored for each exam in turn. One of the defences of those supporting the continuation of unseen written exams, is that these concerns about the tensions between assessment and learning quality are not important as the intention of the assessment is summative rather than formative. They argue that feedback to students should be handled elsewhere in an educational programme, rather than in the context of examinations. However, I argue that the extent of the tensions exceeds that associated with the contribution of feedback to learning quality.

Fourthly, there are concerns about validity and reliability of traditional unseen written exams. Examinations scripts are usually marked far too quickly. Most staff who mark exams agree that the task usually has to be completed in haste, in preparation for timetabled exam boards. It can then be said that one of the most important things that we do for students is performed as a bolt-on addition, rather than a well-planned central part of the design of their education. Lecturers and examiners are often tired and bored when marking students’ scripts. Because of the speed with which exam scripts need to be marked, and the pressure to do the task well, staff are not functioning at their best while undertaking the task. This leads to increased danger that the assessment is not reliable. Even under good conditions, there is abundant data on the problems both of inter-assessor reliability and intra-assessor reliability. This suggests that even the best groups of assessors find it a difficult and complex process.

Finally, there is cause for concern that traditional unseen written exams do not really measure the learning outcomes which are the intended purposes of higher education. Brown (Chapter 8) expands on this concern in the context of assessing practice. Exams tend to favour candidates who happen to be skilled at doing exams, rather than at anything more important. If we look at exactly what skills are measured by unseen written exams, the most important of these from the students’ point of view turns out unsurprisingly to be the techniques needed to do unseen written exams! This can be interpreted as a serious threat to the validity of such exams; in other words, what is being measured may be much less important than what should have been measured. There are many important qualities which are not tested by traditional exams. For example, unseen written exams are limited or useless for measuring teamwork, leadership, and are rarely a suitable vehicle for measuring creativity and lateral thinking.
Some failings of continuous assessment systems

The umbrella term 'continuous assessment' tends to be applied to ongoing measurement of the work that students do throughout a course rather than at fixed end-points. The term 'formative assessment' is often mentioned in such contexts, particularly when continuous assessment is coupled with the provision of feedback to students on their work. However, there is still the need for much of continuous assessment to play a summative role, in that it is counted towards overall assessment, including sometimes final-degree classifications.

Firstly, there are tensions between continuous assessment and the learning payoff which is associated with the time students spend doing assessed work. When students are under too much coursework pressure, their 'want to learn' is damaged. When almost everything that students do is measured, they naturally adopt strategic approaches to their learning, and only concentrate on those things that are going to be assessed. Students' learning becomes driven by assessment, and students may only do those things that are assessed. Such strategic approaches can be made beneficial if the nature and range of the assessed tasks are adjusted to make all the learning that students do in their assessed work relevant as possible to the intended learning outcomes. Furthermore, the range of learning experiences associated with continuous assessment can be too narrow. For example, essays and reports make up the majority of students' continuous assessment formats in many disciplines, and the skills tested are primarily those associated with preparing essays and reports, rather than the deeper knowledge or understanding that may be intended. The value of feedback to students may be eclipsed by marks or grades. Students pay most attention to their scores or grades when they get back marked work, and often are quite blind to valuable feedback which may accompany their returned work. Particularly when there is a long delay in getting feedback to students, they may already have moved on to learning other topics, and they do not then find learning from the feedback a priority. Students are often quite unaware of the criteria used to assess their work. When students are practised in interpreting and making use of assessment criteria, the standard of their assessed work rises dramatically. Alerting students to the detail of the assessment agenda is regarded by some staff as a move towards 'spoonfeeding'. However, enabling students to demonstrate their full potential is a desirable goal.

Secondly, there are concerns about validity and reliability of continuous assessment. For example, it can be difficult to detect unwanted collaboration. Particularly with assignments submitted in word-processed formats, it is hard, if not impossible, to detect every instance of plagiarism or copying. For teaching staff, continuous assessment often results in too much time spent in marking. In many courses, lecturers continue to try to use the same continuous assessment processes that worked quite well when student numbers were much smaller. This leads to reduced reliability of assessment. Students have their own problems with continuous assessment, and may get the balance wrong between ongoing assessment and exams. For example, students feeling under pressure to submit coursework by stated deadlines, may still be attempting such work at a late stage in their studies on a particular module, when they would be better advised to cut their losses regarding coursework and prepare for important exams.

The failings both of traditional unseen written exams in particular, and continuous assessment regimes in general, should not be taken as sufficient causes to abandon all traditional approaches to assessment. Despite these failings, traditional assessment systems have worked for a long time, and much of the experience that has been developed in making them work continues to represent valuable foundations for further development of assessment processes and practices. I suggest that innovations in assessment need to build on this experience, and that innovations need to be made incrementally, and tested and refined on an ongoing basis. It should not surprise anyone if an assessment innovation proves not to work well at first. As long as each innovation is carefully regulated to ensure that any teething troubles do not translate into casualties in students' learning experiences or their subsequent qualifications, there remains every reason to experiment with and research into the validity and reliability of innovative approaches to assessment.

Dimensions of innovation

There are many variables which can be the basis for experimentation, research and analysis in redesigning student assessment. Any of these variables can be adjusted independently, or two or more can be the basis for innovations. I suggest that the most productive way of deciding upon directions for innovation is, however, to interrogate each of the variables separately at first, and to have good reasons for wishing to adjust them before planning the implementation of any innovation. The dimensions I will consider are timing, content, choice of assessment methods, the balance between ongoing and assessing evidence. The anticipated effects of each and every adjustment should, I suggest, be interrogated against the following five questions:

- How can the innovation increase the learning payoff associated with the assessment experience?
- How can the innovation claim to make the assessment more valid? (In other words, how does the new assessment instrument or process get closer to measuring students' genuine and direct achievement of the intended learning outcomes?)
- How can the innovation be defended in terms of the assessment being more reliable than that which it is replacing?
• How can the innovation be seen to make the assessment workload of the staff involved more realistic, productive and relevant to students' qualifications?
• How can the innovation reduce the extent to which students are forced to jump repeatedly through the same hoops?

In the foregoing, I suggest that each of the questions above should be used together with the separate questions I will propose about each of the dimensions of innovation addressed in turn.

**Timing of assessment**

In traditional educational programmes, there has always been a tendency to have most of the important assessment episodes towards the end of a course or module. While this can be justified in terms of this being the time when students should have learned enough to be ready for assessment, a consequence is that too much weight is placed on students' performance during a period of a few hours spread over a week or two. Moreover, the assessment of all the different subjects being studied tends to converge at these times. This is bad news for the significant numbers of students who do not happen to be at their best during this narrow bandwidth of time. Some questions about timing which may fuel innovations are presented below.

- Is that which is measured in three hours much more reliable than that which could have been measured in 30 minutes? A short exam, for example, can measure some things just as well as a much longer one (and take much less time to assess, with more chance of the assessment being done in a reliable way).
- Is speed so important? What is the desired balance between students doing something well and doing the same thing quickly? How significant is the danger that we force all students to do things quickly, then end up measuring how well they do them under such conditions, missing out measuring how well they could have done them? When speed is deemed to be important, are we in danger of measuring speed too often?
- Why can a three-hour exam be justified? What is it that can be measured in three hours that could not be measured in much shorter assessment episodes? There are indeed good answers to these questions, but then we might ask two further questions: How often do we need to measure these things? Are we in danger of using a lot of time to measure these same things repeatedly?
- Do all summative assessments have to occur during the same weeks? An alternative would be to have modules spanning different lengths of time, and starting and finishing at different times of the year. This would at least allow students to concentrate their energies on assessments (coursework or exams) in one subject at a time. It can be argued that this will be fairer in terms of students demonstrating their optimum performance in each subject.
- With coursework, what is the learning payoff per unit time involved? For example, how much greater is the learning payoff of writing a full essay, than of producing a clear essay plan? How many times do we need to get students to produce full essays (or reports) to show that they can do so? How often could we use short-form assessments to generate significant learning payoff over much shorter timescales?

**Content being assessed**

The validity of assessment needs to be the prime directive in deciding how the content agenda is addressed. What are we actually measuring? Some content-related questions which may fuel innovation in assessment are listed below.

- What are we trying to assess? Is it merely how much students can remember about what they have read? Is it merely how well they can write about what they remember about what they have learned? Is it how well students can handle a given body of information? Is it how well students can find their own way through all the available information, and pick out what is the most important information?
- Is assessment measuring what has been learned, and not just what was taught? The Quality Assurance Agency in England emphasizes the importance of this in the briefing of subject reviewers.
- Are we measuring learning, or are we paying too much attention to the ways that students communicate their learning in writing? Is the latter, are we giving appropriately weighted credit to students' communication skills, and making sure that we are not measuring the same skills more frequently than we need to?
- What content agenda are we trying to measure? How much of it is what has been covered in lectures? How much of it should students have found out for themselves? How do we communicate to students the content agenda that we intend to set for them? How can we show them what we will be looking for in their assessed work, without stunting their own initiative and creativity?
- Where the content has already been delineated in terms of learning outcomes, how appropriate is the assessment method to the measurement of students' actual achievement of the outcomes?

**Selection of assessment methods**

Throughout this book are mentioned different assessment methods. A wide range of assessment methods are discussed by Brown and Knight (1994).
These range from traditional unseen written exams, through many kinds of alternative exam formats, to a wide variety of other 'measurables' that can be products of students' individual work or of their collaborative work. Each assessment method advantages some students. Each assessment method disadvantages other students. The following general questions about assessment methods can be the starting point for productive innovations in assessment.

- Is there a sufficient diversity in the mixture of assessment methods being used on each course? Does this allow students who happen not to show themselves at their best in particular assessment situations adequate opportunities to demonstrate their full potential through other assessment situations?
- Does the amount of time devoted by students to their work related to each assessment method reflect the relative importance of what is being measured in each instance?
- Are we using diverse assessment methods to ensure that we are not measuring the same things again and again?

The balance between collaborative and individual work

As mentioned earlier in this chapter, and in many other places in this book, employers value collaborative skills. Indeed, most avenues of working life or professional development involve collaboration at least as much as individual effort or personal achievement. Traditionally, assessment has tended to measure individual achievement, not least because this is easier to measure 'fairly' or 'objectively', but I have placed these words in quotes because it is often the individuality that is measured thus, rather than that which it should have been intended to measure fairly and objectively. When trying to measure collaborative work objectively, there is the danger that 'if we can assess it, it probably isn't it!' A couple of questions about the balance between assessment of individual work and collaborative work, which may provide food for thought for assessment innovations, are given below.

- Is the difficulty in attributing individual achievement standing in the way of using assessed collaborative tasks to measure valuable skills and attributes? It is likely that in the overall assessment profile of the course, individual achievement will already be addressed substantially. Therefore, it may be worth striking a balance between the value of the collaborative work and the objectivity of some aspects of its assessment.
- Are there ways in which triangulation of individual achievement within collaborative work can be made? For example, when it is only possible to measure the overall product of collaborative work, can a viva or presentation be used to make a realistic estimate of the individual contribution to the work?

The balance between assessing performance or assessing evidence

To achieve a realistic, rounded assessment, any assessment programme needs to balance the measurement of students' performance alongside that of the evidence which they produce. With some kinds of evidence, there may be concerns about the authenticity of the ownership of the work being assessed. Yet some aspects of performance are too difficult or overly time-consuming to attempt to measure directly, and assessment may need to concentrate on evidence arising from the performance. The balance between evidence and performance needs to be a consideration in all forms of assessment. Even with as familiar an assessment format as traditional unseen written exams, the evidence which is assessed relates both to students' learning achievements and to their performance under the conditions of the assessment. The following questions are intended to provide some starting points for assessment innovations addressing the need to optimize the balance between assessing evidence and performance.

- Are the conditions under which the assessment takes place appropriate? For example, if the real intention is to assess performance of some kind, does the assessment method lend itself to measuring this objectively, and are the assessment criteria relating to the performance clearly understood by the students themselves?
- Are the same elements of performance being measured too often? Some aspects of performance are easier to measure than others, and there is the danger of measuring the same ones again and again, while failing to try to measure other (perhaps more important) elements. For example, the assessment of essays in written exams at least partially reflects the performance of writing an essay in handwriting (rather than using a word-processor) under exam conditions, and good performance at this task may be repeatedly rewarded, and cumulatively may play too significant a part in distinguishing a first-class degree from a lower one.
- Where it has been decided that a particular performance aspect is important, is the measurement being done in a way that promotes learning and provides feedback to students? Using self-assessment of performance as a contributory factor can cause deepening of the associated learning experience. Using peer assessment of the performance can lead to greater reliability of the assessment, and can provide increased richness of feedback to students about their performances.

Conclusions

The dimensions of innovation I chose to discuss in this chapter are by no means exhaustive. They are intended to be indicative, and to illustrate just some of the possibilities regarding assessment innovations. They should
enable readers to find their own rationale in looking for ways in which innovative assessment can be introduced, not least because we are increasingly aware that the familiar, tried-and-tested methods are not matching up to the requirements of a widening range of stakeholders in higher education. Most importantly, however, assessment innovations are the best way forward in researching and evaluating the effects of assessment on the two most vital factors involved: the quality of students' learning and the validity of their qualifications.

References

UCoSDA/Loughborough University (1996) *Making the Grade*. Sheffield: UCoSDA.