Learning and Teaching in Higher Education
Issue 1, 2004-05

Editors' Introduction

Assessment in higher education has come under increasingly close scrutiny both as a focus of scholarship and polemical contestation. This, the first issue of LATHE, addresses the scholarship of assessment. For many years assessment was an unquestioned 'given' in higher education, the preserve of the few who judged the many. Its foundation on tacit knowledge added to its power and mystique. Elites reproduced themselves by promoting the very practices through which they had succeeded. Newly appointed academic staff would be introduced to opaque and arcane assessment practices — not through programmes of professional development but through the gradual (or in some cases instantaneous) grasp of the 'feel' for an appropriate grade. Students would passively receive the grade — which might be accompanied by words of praise or 'could do better'.

The articles in this issue demonstrate clearly how assessment, at the level of strategy and practice, is now being addressed through the scholarship of a wide range of staff in higher education — academics, researchers, learning technologists and educational developers.

Professor Graham Gibbs and Dr Claire Simpson discuss how assessment arrangements can promote student learning. They propose a framework of ten 'conditions under which assessment supports learning', to enable staff to assess and evaluate their own practice. Dr Scott Fleming and Dr David James investigate agreement in student assessment performance within and between modules from similar disciplines, and explore the widely-held view that grades awarded for examinations tend to be lower than those for coursework. Professor Lorraine Stefani provides a challenging perspective in discussing how to raise awareness among academic staff of the scholarship of teaching, learning and assessment, and how to support staff in developing a scholarly approach to assessment practices. Martin Jenkins considers approaches through which computer-aided assessment can provide formative feedback to enhance student learning. In the final paper, by provocatively inviting the reader to consider assessment techniques which actually prohibit student learning, Professor Sally Brown emphasises the crucial role of assessment.

The seven case studies address assessment through presentations; the use of grade descriptors for marking; aligning assessment with learning activities; self-evaluation of coursework; peer review in

scenario-based assessments; learning conversations as a framework for formative assessment; and the support of higher-order cognitive skills through self-assessment.

As the forgoing articles and case studies amply illustrate, the significance of assessment in relation to student learning undoubtedly justifies its selection as the focus of the first issue of this new journal.

Phil Gravestock and Kristine Mason O'Connor
University of Gloucestershire, UK

Conditions Under Which Assessment Supports Students' Learning

Graham Gibbs[1] & Claire Simpson[2]

[1] Oxford University, [2] Open University, UK

Abstract

Much evaluation of teaching focuses on what teachers do in class. This article focuses on the evaluation of assessment arrangements and the way they affect student learning out of class. It is assumed that assessment has an overwhelming influence on what, how and how much students study. The article proposes a set of 'conditions under which assessment supports learning' and justifies these with reference to theory, empirical evidence and practical experience. These conditions are offered as a framework for teachers to review the effectiveness of their own assessment practice.

Introduction

When teaching in higher education hits the headlines it is nearly always about assessment: about examples of supposedly falling standards, about plagiarism, about unreliable marking or rogue external examiners, about errors in exam papers, and so on. The recent approach of the Quality Assurance Agency (QAA) to improve quality in higher education has been to focus on learning outcomes and their assessment, on the specification of standards and on the role of external examiners to assure these standards. Where institutional learning and teaching strategies focus on assessment they are nearly always about aligning learning outcomes with assessment and about specifying assessment criteria. All of this focus, of the media, of quality assurance and of institutions, is on assessment as measurement. This article is not about measurement at all — it is about learning. The most reliable, rigorous and cheat-proof assessment systems are often accompanied by dull and lifeless learning that has short lasting outcomes — indeed they often directly lead to such learning. We are not arguing for unreliable assessment but we are arguing that we should design assessment, first, to support worthwhile learning, and worry about reliability later. Standards will be raised by improving student learning rather than by better

measurement of limited learning. This article is about how to design assessment that supports worthwhile learning. The case studies elsewhere in this issue are about particular assessment methods — tactics if you like. Guidance on how to implement a wide range of assessment tactics can be found elsewhere (e.g. Gibbs, 1995). This article is about strategy — about the functions that assessment performs (Gibbs, 1999) that enable a teacher to select appropriate assessment tactics. We will argue that assessment works best to support learning when a series of conditions are met. The article will examine the nature of these conditions.

The dominant influence of assessment

In the early 1970s researchers on both sides of the Atlantic (Snyder, 1971; Miller & Parlett, 1974) were engaged in studies of student learning at prestigious universities. What they found was that, unexpectedly, what influenced students most was not the teaching but the assessment. Students described all aspects of their study — what they attended to, how much work they did and how they went about their studying — as being completely dominated by the way they perceived the demands of the assessment system. Derek Rowntree stated that 'if we wish to discover the truth about an educational system, we must first look to its assessment procedures' (Rowntree, 1987, p.1). The Snyder and Miller & Parlett studies went further and highlighted the way students respond to these assessment procedures. More recently, qualitative studies have emphasized the importance of understanding the way students respond to innovations in assessment (Sambell & McDowell, 1998).

Snyder's work gave birth to the notion of the 'hidden curriculum' — different from the formal curriculum written down in course documentation, but the one students had to discover and pay attention to if they wanted to succeed:

> 'From the beginning I found the whole thing to be a kind of exercise in time budgeting … . You had to filter out what was really important in each course … you couldn't physically do it all. I found out that if you did a good job of filtering out what was important you could do well enough to do well in every course.' (Snyder, 1971, pp.62-63)

Once students had worked out what this hidden curriculum consisted of they could allocate their effort with great efficiency:

> 'I just don't bother doing the homework now. I approach the courses so I can get an 'A' in the easiest manner, and it's amazing how little work you have to do if you really don't like the course.'
> (Snyder, *ibid.*, p.50)

Miller & Parlett focused on the extent to which students were oriented to cues about what was rewarded in the assessment system. They described different kinds of students: the 'cue seekers', who went out of their way to get out of the lecturer what was going to come up in the exam and what their personal preferences were; the 'cue conscious', who heard and paid attention to tips given out by their lecturers about what was important, and the 'cue deaf', for whom any such guidance passed straight over their heads. This 'cue seeking' student describes exam question-spotting:

> 'I am positive there is an examination game. You don't learn certain facts, for instance, you don't take the whole course, you go and look at the examination papers and you say 'looks as though there have been four questions on a certain theme this year, last year the professor said that the examination would be much the same as before', so you excise a good bit of the course immediately …'
> (Miller & Parlett, 1974, p.60)

In contrast, these students were described as 'cue-deaf':

> 'I don't choose questions for revision — I don't feel confident if I only restrict myself to certain topics'
>
> 'I will try to revise everything …'
> (Miller & Parlett, 1974, p.63)

Miller & Parlett were able to predict with great accuracy which students would get good degree results:

> '… people who were cue conscious tended to get upper seconds and those who were cue deaf got lower seconds.'
> (Miller & Parlett, 1974, p.55)

Many students are perfectly capable of distinguishing between what assessment requires them to pay attention to and what results in worthwhile learning, as this postgraduate Oceanography student explained:

> 'If you are under a lot of pressure then you will just concentrate on passing the course. I know that from bitter experience. One subject I wasn't very good at I tried to understand the subject and I failed the exam. When I re-took the exam I just concentrated on passing the exam. I got 96% and the guy couldn't understand why I failed the first time. I told him this time I just concentrated on passing the exam rather than understanding the subject. I still don't understand the subject so it defeated the object, in a way.' (Gibbs, 1992, p.101)

Whether or not what it is that assessment is trying to assess is clearly specified in documentation, students work out for themselves what counts — or at least what they think counts, and orient their effort accordingly. They are strategic in their use of time and 'selectively negligent' in avoiding content that they believe is not likely to be assessed. It has been claimed that students have become more strategic with their use of time and energies since the 1970s and more, rather than less, influenced by the perceived demands of the assessment system in the way they negotiate their way through their studies (MacFarlane, 1992).

The role of coursework assignments

Students tend to gain higher marks from coursework assignments than they do from examinations (*Eds: see James & Fleming, this issue, for a discussion on this topic*). Chansarkar & Raut-Roy (1987) studied the effects of combinations of various forms of coursework with examinations. They found that all combinations of coursework of varying types with examinations produced better average mark rates than did examinations alone — up to 12% higher average marks. Gibbs & Lucas (1997) reported an analysis of marks on 1,712 modules at Oxford Polytechnic. Modules with 100% coursework had an average mark 3.5% higher than modules with 100% examinations, and there were three times as many failed students on modules where there were only examinations. There was a significant positive correlation between the proportion of coursework on a module and average marks ($r = +0.36$, $p<.0001$). Bridges *et al*. (2002) studied the differences in coursework and exam marks in six subjects at four universities. They found coursework marks to be higher by one third of a degree classification in English and History (similar to the Oxford Polytechnic finding) and higher by two thirds of a degree classification in Biology, Business Studies, Computer Studies and Law.

Students also prefer coursework. Starr (1970) reported that 90% of students from four departments preferred half or more of their marks to come from coursework and 56% preferred all their marks to come from coursework. Students consider coursework to be fairer than exams, to measure a greater range of abilities than exams and to allow students to organize their own work patterns to a greater extent (Kniveton, 1996).

Higher average marks and student preference would not count for much if coursework were inherently less valid as an assessment — but it is not. First, examinations are very poor predictors of any subsequent performance, such as success at work. A review of 150 studies of the relationship between exam results and a wide range of adult achievement found the relationship to be, at best, slight (Baird, 1985). For example, first degree results explain less than 10% of the variance in postgraduate performance (Warren, 1971).

Second, coursework marks are a better predictor of long term learning of course content than are exams. Conway *et al*. (1992) reported a study of the performance of psychology students on a range of tests of their understanding and recall of content of a cognitive psychology course taken many years before. They found that student marks on coursework assignments undertaken up to 13 years before correlated with these test scores while students' original exam marks did not. Presumably the kind of learning that coursework involves has long term consequences while the kind of learning involved in revision for exams does not. Studies of surface and deep approaches to learning have shown similar results: that any positive impact on test results of students taking a surface approach in preparation for the test are very short-lasting (Marton & Wenestam, 1978).

Third, in experimental studies in which students have either studied exam-based or assignment-based courses, the quality of their learning has been shown to be higher in the assignment-based courses. For example Tynjala (1998) compared two student groups: the first group studied via conventional lectures, a text-book and an exam; the second group studied via assignments based on the text-book, discussion with other students about these assignments, and a course-work essay marked by the teacher. This second group then also took the exam so as to enable a comparison with the first group, even though they had not studied for the exam. The second group were found to place more emphasis on thinking and had developed more sophisticated conceptions of learning (see Säljö, 1982). In their exam answers they revealed more comparisons, more evaluations and more sophisticated structures to their answers

in terms of the SOLO taxonomy of learning outcomes (Biggs & Collis, 1982). These results (achieved with less teaching) were interpreted in terms of the assessment requirements for the second group being more constructivist.

It is a common observation of higher education teachers that if coursework is taken away from a module due to resource constraints, students simply do not do the associated studying; for example students will rarely write unassessed essays. It is argued that you have to assess everything that moves in order to capture students' time and energy. However, coursework does not have to be marked to generate the necessary learning. Forbes & Spence (1991) reported a study of assessment on an engineering course at Strathclyde University. When lecturers stopped marking weekly problem sheets because they were simply too busy, students did indeed stop tackling the problems, and their exam marks went down as a consequence. But when lecturers introduced periodic peer-assessment of the problem sheets — as a course requirement but without the marks contributing — students' exam marks increased dramatically to a level well above that achieved previously when lecturers did the marking. What achieved the learning was the quality of student engagement in learning tasks, not teachers doing lots of marking. The trick when designing assessment regimes is to generate engagement with learning tasks without generating piles of marking.

The decline in formative assessment

A traditional characteristic of teaching in higher education in the UK has been the frequent provision of detailed personalized feedback on assignments. The archetype has been that of Oxford or Cambridge University where students wrote an essay a week and read it out to their tutor in a one-to-one tutorial, gaining immediate and detailed oral feedback on their understanding as revealed in the essay. This was almost the only teaching many Oxbridge students experienced: teaching meant giving feedback on essays. This formative assessment was quite separate from marking and at Oxford and Cambridge the only summative assessment often consisted of final examinations at the end of three years of study that had involved weekly formative assessment.

Few institutions have been able to match the quantity or quality of feedback provided by Oxford or Cambridge but the assumption for most has been that frequent assignments and detailed (written) feedback are central to student learning. Until quite recently, for example, many

science courses involved weekly problem sheets and laboratory reports, all of which were marked by teachers and returned to students within the week. In most forms of distance education, feedback on frequent assignments is the main interactive component of teaching and the Open University has placed great emphasis on frequent assignments, training and paying tutors to provide comprehensive feedback, and monitoring the quality of this feedback. For some Open University students this is their only contact with their tutor. They can cope without much, or even any, face-to-face teaching, but they cannot cope without regular feedback on assignments.

Resource constraints in conventional universities have led to a reduction in the frequency of assignments, in the quantity and quality of feedback and in the timeliness of this feedback. Modularisation has tended to shorten courses and has reduced the timescale within which it is possible to set assignments and provide feedback, while increasing the number of examinations. Some courses have abandoned formative assignments altogether. Others may involve just one assignment but with feedback not being provided until very late in the course, or even after the exam. At the same time the diversity of students has increased enormously, so that previous assumptions of the level of sophistication of knowledge background, study skills, conception of learning (Säljö, 1982), or conception of knowledge (Perry, 1970) of students are now likely to be very wide of the mark. Far more guidance is likely to be required by these students who need more practice at tackling assignments and more feedback on their learning, not less. Because regular assignments and comprehensive feedback is understood to be central to distance education, it has in contrast largely been retained; as a result today's Open University students may receive fifty times as much feedback on assignments over the course of an entire degree programme as do students at conventional universities.

The effectiveness of feedback

In a comprehensive review of 87 meta-analyses of studies of what makes a difference to student achievement, Hattie (1987) reports that the most powerful single influence is feedback. Similarly, Black & Wiliam's (1998) comprehensive review of formative assessment emphasizes the extraordinarily large and consistent positive effects that feedback has on learning compared with other aspects of teaching. There have been many attempts both to understand the nature of this impact and to harness its power through innovation, at least in schools, as a consequence of this incontrovertible evidence.

In higher education, feedback to individual students in class must have declined significantly as class sizes have increased, though we have no evidence about this. Writing comments on assignments, however, remains a major component of teachers' workload in higher education. As class sizes have increased there have been some economies of scale in teaching (simply by packing more students into classrooms), but economies of scale are difficult to achieve for assessment: most assessment costs go up in direct proportion to the number of students. As a result, assessment costs can overtake teaching costs and teachers can find themselves spending much of their time marking. Is all this effort worthwhile?

In the Course Experience Questionnaire (Ramsden, 1991), used extensively in Australia and elsewhere to evaluate the quality of courses, the questionnaire item that most clearly distinguishes the best and worst courses is 'Teaching staff here normally give helpful feedback on how you are going' (Ramsden, 1992, p.107). This does not mean that higher education teachers in fact give helpful feedback — it means that whether or not they give helpful feedback makes more difference than anything else they do. How well does feedback actually work?

Maclellen (2001) surveyed 130 students and 80 lecturers at the University of Strathclyde about their perceptions concerning assessment. Amongst the 40 questions asked, four were about feedback and these revealed wide discrepancies between students and lecturers. While most teachers responded that feedback is frequently helpful in detail, frequently helps students to understand and frequently helps learning, most students responded that feedback was only sometimes helpful in these ways. 30% of students reported that feedback never helps them to understand. While 63% of lecturers responded that feedback frequently prompts discussion with a tutor, only 2% of students responded the same way and 50% of students responded that feedback never prompted discussion.

There may be a problem here with the quantity and quality of feedback such that it is not actually helpful to students — after all, teachers are under enormous time pressure and it is difficult to provide comprehensive and useful feedback under such circumstances. But there are other problems. Studies of what students do with feedback makes for depressing reading. Feedback is often not read at all (Hounsell, 1987) or not understood (Lea & Street, 1998). Wotjas (1998) reported:

> 'Some students threw away the feedback if they disliked the grade, while others seemed concerned only with the final result and did not collect the marked work.'

There is also a problem associated with both marks and feedback being provided. A grade is likely to be perceived by the student as indicating their personal ability or worth as a person as it is usually 'norm-referenced' and tells you, primarily, where you stand in relation to others. A poor grade may damage a student's 'self-efficacy', or sense of ability to be effective. Yorke (2001) elaborates on the positive or negative ways in which formative assessment can affect student retention and emphasizes its role in 'academic integration' (Tinto, 1993). In contrast, feedback on its own is more likely to be perceived as a comment on what has been learnt. In the absence of marks it has been reported that students read feedback much more carefully (Black & Wiliam, 1998) and use it to guide their learning. In the light of this (school-based) research evidence, some schools have adopted policies that all assignments should only have feedback and that no marks should be provided. The Alverno College 'assessment as learning' system is probably the best known higher education example of 'grade-less' assessment.

This is not a pretty picture. Assessment sometimes appears to be, at one and the same time, enormously expensive, disliked by both students and teachers, and largely ineffective in supporting learning. In the light of these problems the remainder of this article sets out and attempts to justify a set of 'conditions under which assessment can support learning'. The evidence is rarely conclusive enough to argue that if your assessment fulfils these conditions then learning will inevitably be more effective. They are offered as a plausible set of guidelines.

This is not the first attempt to identify such 'conditions' but is the first attempt in the context of higher education. School-based research has identified lists of effects of formative assessment such as the one below, based on Gagne (1977):

1. Reactivating or consolidating prerequisite skills or knowledge prior to introducing the new material
2. Focusing attention on important aspects of the subject
3. Encouraging active learning strategies
4. Giving students opportunities to practise skills and consolidate learning

5. Providing knowledge of results and corrective feedback
6. Helping students to monitor their own progress and develop skills of self-evaluation
7. Guiding the choice of further instructional or learning activities to increase mastery
8. Helping students to feel a sense of accomplishment.

(Crooks, 1988)

The conditions outlined here refer to two relatively distinct categories of influence:

- the influence of the design of assessment systems and assignments on how much students study, what they study and on the quality of their engagement
- the influence of feedback on learning.

Influences of assessment on the volume, focus and quality of studying

Condition 1

Sufficient assessed tasks are provided for students to capture sufficient study time

This issue concerns how much time and effort students allocate — the 'time on task' principle (Chickering & Gamson, 1987) that if students don't spend enough time on something they won't learn it. Berliner (1984), summarising research in the 'time on task' principle, concluded that there was strong empirical evidence of a direct relationship between time allocation by courses, student time management and actual student time on task, on the one hand, and student achievement on the other.

The relationship between effort and marks is not always straightforward. Kember *et al*. (1996) found that students' perceptions of their effort depended on their motivation more than on the number of hours they actually allocated, and that it was possible for students to put in many hours unproductively, especially if they adopted a surface approach to their studies. Some kinds of assessment can generate long hours of ineffective memorization.

Courses in UK higher education are designed to involve a specified number of learning hours relating to the number of credits for the course. Students are normally expected to spend between about one and four hours out of class for each hour in class (depending largely on the discipline involved). Innis (1996) found students at Leeds Metropolitan University spend between 1.4 and 3.0 hours out of class for each hour in class. How much of this 'out of class' time is actually allocated to studying may be determined largely by assessment demands. In the USA, higher education students on average spend less than half as many hours out of class for each hour in class as teachers expect: between 0.3 and 1.0 hours out of class when teachers, on average, expect 2.1 hours out of class for each hour in class (Moffat, 1989; Hutchings *et al*., 1991; Gardiner, 1997; Brittingham, 1998). The emphasis in the USA on attempts to improve student performance through assessment is on 'classroom assessment' — activities undertaken in class to test students and use this assessment information to guide both students and teaching (Angelo & Cross, 1993). This focus on the classroom could be interpreted as a recognition of the failure to generate much out of class learning through the type of assessment they use. Diary studies (e.g. Innis, 1996) show how students in the UK allocate their time largely to assessed tasks and that this becomes a more narrow focus over time as they become more experienced students, allocating as little as 5% of their time to unassessed study tasks by year three.

Subject areas with less frequent assessed tasks (e.g. text-based subjects) have students who study fewer hours (Vos, 1991). Science and technology subjects that generate greater total study effort tend to have more frequent (though smaller) assessed tasks, such as problem sheets and laboratory reports.

Studies of the impact of students undertaking paid employment in parallel to full time study show that such students study fewer hours (Curtis & Shami, 2002) and perform significantly less well (Paton-Salzberg & Lindsay, 1993). Studies show that up to three quarters of full time students work during term time and they are likely to allocate their reduced study hours especially strategically in relation to assessment requirements. They report reduced reading and other out of class study activity.

Assignments are not the only way to capture student time and effort through assessment. The conventional way to do this is by having unpredictable sampling of course content in unseen examinations so that for a student to ignore anything is a high risk activity. The quality, quantity and distribution of the study effort captured in this way is somewhat unpredictable and probably varies with student perceptions of the likely exam demands and the risks involved.

Time and effort can also be captured through social pressure, for example:

- the potential embarrassment of the poor quality of your work being seen by colleagues, as when a seminar presentation is assessed, or when a laboratory report is written and displayed publicly in the form of a poster
- the potential censure from colleagues if a student were to fail to complete their component of an assessed group assignment.

Condition 2

These tasks are engaged with by students, orienting them to allocate appropriate amounts of time and effort to the most important aspects of the course.

This condition concerns what the effort is oriented towards and what quality of effort is involved. Students usually distribute their time unevenly across courses, often focusing on topics associated with assessment and nothing else. If they drew a graph of weekly study effort for all the weeks of an individual course involving a sequence of assignments, it might look more like the Alps than like Holland. Exams can have the effect of concentrating study into a short intense period at the end of the course with little study of, for example, lecture notes, until many weeks after the lecture. Frequent assignments (such as short problem sheets) or tests (such as computer-based assessment) can distribute student effort across the course, often on a weekly basis, while infrequent assignments (such as extended essays) may result in intensive studying for a week or two immediately prior to the assignment deadline, while topics not covered by the assignment can be largely ignored.

We know very little about the distribution of student effort and higher education teachers also tend to know little about what their students do with their time and when.

Condition 3

Tackling the assessed task engages students in productive learning activity of an appropriate kind

This issue concerns the kinds of study and learning activity involved in tackling the assignment or in preparing for tests. Some assessment generates unhelpful and inappropriate learning activity, even if it produces reliable marks. Studying for multiple choice question

(MCQ) tests can orient students to a surface approach (Scouler & Prosser, 1994; Tang, 1994; Scouler, 1998), as can exams, though the approach to learning of students may have as much impact as the form of test. Students may take a deep approach to preparing for MCQ tests and adopting effective study strategies even when the test only makes low level demands, and Macdonald (2002) has reported that at least some students adopted a deep approach to examination revision and learning effectively as a result of the integration of material that their revision involved.

Much assessment simply fails to engage students with appropriate types of learning. Submitting a laboratory report of a teacher-designed procedure is unlikely to help students to learn how to design experiments. Probably the only way to learn how to solve problems is to solve lots of problems. Probably the only way to gain facility with the discourse of a discipline is to undertake plenty of practice in using that discourse, for example through writing. Assignments are the main way in which such practice is generated. Students are unlikely to engage seriously with such demanding practice unless it is assessed, or at least required, by the assessment regulations. It seems unlikely that this student would write essays, and acquire the learning that resulted, without being required to:

> 'It's just work, in a way. Just all these essays, and reading's the worst part, it's just labouring really.' (History student) (Hounsell, 1987)

Some assessment can mis-orient student effort. Snyder (1971) described how students encouraged to be creative at Massachusetts Institute of Technology abandoned any such aspiration on discovering that most of the marks were derived from rote memorization of material for multiple choice tests. Some assignments create appropriate learning activity as a by-product. For example, setting essays can generate 'reading around' and can support the working up of coherent arguments in a way that simply asking students to read what is on the reading list does not. If you were to take the essay away, the appropriate form of studying would not occur even in the unlikely event of a similar volume of reading of similar material taking place. The product, the essay, and the marks associated with it, may be less important to the learning than the framework the assignment provides for the learning activities of 'reading around' and of 'constructing arguments'. Similarly, with laboratory reports or design briefs, the product may be less important than details of the studying required to fulfil the assignment requirements. Group projects can engage students in much discussion and confront individuals with

alternative views and different standards of work. The quality of the group product (such as a report) that is marked may be less important than the qualities of the learning process that created it.

Students can tackle assignments that are intended as learning activities so as to maximize the marks they obtain rather than maximising the learning achieved from engaging with the assignment. This may involve 'faking good' and pretending to be competent or knowledgeable, deliberately covering up misunderstanding and ignorance, telling teachers what they want to hear rather than what they as students believe, and so on. To some extent this is a consequence of the student's orientation, but assessment tasks, marking regimes and the way feedback functions can override such individual orientations and even encourage student behaviour that reduces learning. In the example below an intrinsically oriented student describes, in a learning log, the means he used to tackle assignments in Engineering in a way designed to obtain marks at the expense of learning:

> 'The average lecturer likes to see the right result squared in red at the bottom of the test sheet, if possible with as few lines of calculation as possible — above all else don't put any comments. He hates that. He thinks that you are trying to fill the page with words to make the work look bigger. Don't leave your mistakes, either, even corrected. If you've done it wrong, bin the lot. He likes to believe that you've found the right solution at the first time. If you're still making mistakes, that means you didn't study enough. There's no way you can re-do an exercise a few months after because you've only got the plain results without comments. If you have a go, you may well make the same mistakes you've done before because you've got no record of your previous errors.' (Gibbs, 1992)

The influence of feedback on learning

> 'Knowing what you know and don't know focuses learning. Students need appropriate feedback on performance to benefit from courses. In getting started, students need help in assessing existing knowledge and competence. In classes, students need frequent opportunities to perform and receive suggestions for improvement. At various points during college, and at the end, students need chances to reflect on what they have learnt, what they still have to learn, and how to assess themselves.' (Chickering & Gamson, 1987)

Conventionally, feedback is conceptualized as an issue of 'correction of errors' (Bruner, 1974) or 'knowledge of results' in relation to learning itself; if a student is informed that she is accurate then she will learn. In this article we are concerned with how the provision of feedback affects student learning behaviour — with how feedback results in students taking action that involves, or does not involve, further learning.

Condition 4

Sufficient feedback is provided, both often enough and in enough detail

This issue concerns what is conventionally defined as formative assessment: the impact on learning of feedback on progress, usually provided after a 'performance' on an assignment. The volume and thoroughness of feedback varies enormously between courses — we suspect far more than the variation in quantity or quality of teaching.

This feedback may need to be quite regular, and on relatively small chunks of course content, to be useful. One piece of detailed feedback on an extended essay or design task after ten weeks of study is unlikely to support learning across a whole course very well. There has been very widespread adoption of computer-based testing to provide at least some feedback on progress, and in some assessment software it is possible to provide 'remedial feedback' when incorrect answers are selected. Cook (2001) has reported that students' final exam marks were closely related to the number (and therefore frequency) of computer marked assignments students had tackled. The frequency and speed of response of such feedback, which is possible to provide reasonably economically, may compensate for its relatively poor quality and lack of individualization.

Feedback has to be quite specific to be useful. The Open University trains its 7,500 part time tutors to give quite detailed and extensive feedback. Cole *et al*. (1986) list the characteristics of effective feedback in distance learning and Roberts (1996) found that students' preferences for feedback closely match this list. The specific forms of feedback that are effective vary from discipline to discipline. Evidence about the most effective forms of feedback in language learning, for example, is summarized in Hyland (2001). In both Psychology (Stephenson *et al*., 1996) and Mathematics (Rice *et al*., 1994) students have been reported as wanting specific, detailed facilitative feedback. Greer (2001) reports a study that illuminates exactly what kind of impact feedback was achieving on the learning of Accountancy.

Much of the feedback to students provided in the rest of higher education would be picked up by the Open University's Staff Tutors (who monitor tutors' marking) as being totally inadequate and would lead to quality assurance and staff development interventions.

Condition 5

The feedback focuses on students' performance, on their learning and on actions under the students' control, rather than on the students themselves and on their characteristics

Literature on formative assessment distinguishes between feedback which tells students they are hopeless, or amongst the bottom 10% of students (a grade D, for example), and feedback which tells students exactly where they have gone wrong and what they can do about it. Grades without feedback may be particularly damaging. A focus of critical feedback on personal characteristics can be demotivating and can negatively affect students' 'self-efficacy' or sense of competence. This is important because self-efficacy is strongly related to effort and persistence with tasks (Schunk, 1984; 1985), predicts academic achievement well and is associated with adopting a deep approach to learning (Thomas *et al*., 1987). In contrast, feedback concerning content provides the student with options for action and is less closely associated with their ego — it is about their action rather than about themselves. Wootton (2002) has written passionately about the negative impact of assessment on 'at risk' students and asks whether the system exists 'to encourage learning or to measure failure'.

Condition 6

The feedback is timely in that it is received by students while it still matters to them and in time for them to pay attention to further learning or receive further assistance

This issue was highlighted in the 'seven principles of good practice in undergraduate education' (Chickering & Gamson, 1987; 1991). It is based on a range of studies of the timing of feedback (for summaries, see Dunkin, 1986; McKeachie *et al*., 1986). A teaching method which places great emphasis on immediate feedback at each stage of a student's progress through course units, the Personalised System of Instruction (PSI), has been demonstrated in many studies to improve student performance (Kulik *et al*., 1980).

If students do not receive feedback fast enough then they will have moved on to new content and the feedback is irrelevant to their ongoing studies and is extremely unlikely to result in additional appropriate learning activity, directed by the feedback. Due to resource pressures feedback is being provided more slowly and as courses in the UK are now shorter, this may mean that feedback on coursework is not provided until after the course has finished. Much such expensively provided feedback is likely to be wasted. There may be a trade off between the rapidity and quality of feedback so that, for example, imperfect feedback from a fellow student provided almost immediately may have much more impact than more perfect feedback from a tutor four weeks later.

Carroll (1995) described 'formative assessment workshops' for classes of 300 medical students which consisted of multiple choice question test items followed immediately by a short remedial tutorial on the question. There was no individualized feedback in this system but the feedback was very immediate and the workshop sessions were scheduled to allow students time to study more material before moving on to the next section of the course. 85% of students reported wanting more such sessions. Sly (1999) reported the impact of 'practice tests' on subsequent exam performance. Students had the option of taking a practice test, with computer-based feedback, sufficiently in advance of an exam to enable them to use the feedback to undertake some more studying to address their weaknesses. 197 weaker students chose to take these practice tests and these students improved their exam scores so much that they outperformed 417 stronger students. The benefits were still evident in a subsequent exam.

Condition 7

Feedback is appropriate to the purpose of the assignment and to its criteria for success

This issue concerns the relationship of feedback to what an assignment has been set for and what counts as a successful attempt at the assignment. Feedback can perform several functions. For example it can be used primarily to:

- correct errors
- develop understanding through explanations
- generate more learning by suggesting further specific study tasks

- promote the development of generic skills by focusing on evidence of the use of skills rather than on the content
- promote meta-cognition by encouraging students' reflection and awareness of learning processes involved in the assignment
- encourage students to continue studying.

Which of these is appropriate depends on why the particular assignment was set in the first place. For example, was the intention to provide a single opportunity to practise the use of a procedure or algorithm in an accurate way, to provide one of many opportunities to practise in the use of a transferable skill, to offer a rich opportunity to reflect on learning, or to provide an easy first hurdle in a course that it would be motivating for a student to complete?

A recent study at the Open University suggested that maintaining motivation was the most important and influential issue for new students for their first assignment in a course (Gibbs & Simpson, 2002). If a student is looking for encouragement and only receives corrections of errors this may not support their learning in the most effective way.

Students need to understand why they have got the grade or mark they have and why they have not got a higher (or lower) grade. Criteria need to be explicit and understood by students, and demonstrably used in forming grades. Often criteria are not accompanied by standards and it is difficult for a student to tell what standard is expected or would be considered inadequate. Much of the literature on the use of self- and peer-assessment is about the reliability of such marking, and assumes that self- and peer-assessment is primarily a labour-saving device. But the real value may lie in students internalising the standards expected so that they can supervise themselves and improve the quality of their own assignments prior to submitting them.

Students need to understand criteria in order to orient themselves appropriately to the assignment ask. Penny & Grover (1996) have reported the extent to which students misunderstood the criteria to be used to assess their final year research project. The students expected criteria to be concerned with low-level goals such as style and presentation while their teachers emphasized high level goals such as theoretical and conceptual understanding. Opportunities to provide feedback at multiple stages during an ongoing project can re-orient student effort in appropriate ways (Carless, 2002).

Assessment also performs a role in conveying the standard that students have to aspire to. Conveying high expectations is one of the 'seven principles of good practice in undergraduate education' (Chickering & Gamson, 1987). Feedback, model answers and especially exemplars (Orsmond *et al*., 2002) help to establish these expectations and self-assessment helps students to internalize them.

Condition 8

Feedback is appropriate, in relation to students' understanding of what they are supposed to be doing

Students' conceptions of the task

Students have to make sense of what kind of a task they have been set when they tackle an assignment and what would count as a 'good' attempt at it. They can misunderstand and be confused by whatever briefing and feedback they have been given in the past, as in this example:

> 'What do you think the tutor was looking for in this essay?
>
> Ah … well!, this is confusing me. I know the tutor likes concise work, but doesn't like generalisations, and doesn't like too much detail, although on the whole I think he'd like more detail than generalisations. And because it was such a general question, I though 'oh help!', I don't know what he's looking for.'
> (Hounsell, 1987)

Whatever feedback this student's tutor gives will be interpreted in the light of this student's conceptions of what the tutor really wants or what the task really consists of. Students can have a great deal of difficulty understanding what form of communication an essay is (when the only audience knows more than they do about the topic), or what a laboratory report is for (when it has already been written hundreds of times before in exactly the same format), or what a design task has been set for (when only the product is assessed and not the learning that was involved in creating it). Many academic tasks make little sense to students. This inevitably causes problems when they come to read feedback about whether they have tackled this incomprehensible task appropriately.

Students' conceptions of learning

Underlying the above students' confusion about what the tutor really wants could be an unsophisticated conception of learning. Säljö (1982) describes students as having one of five conceptions of learning:

1. Learning as passive receipt of information
2. Learning as active memorization of information
3. Learning as active memorization of information or procedures, to be used at some time in the future
4. Learning as understanding
5. Learning as a change in personal reality: seeing the world differently.

A student with conceptions of learning 1, 2 or 3 might have trouble interpreting feedback that stated: 'Not enough discussion' if they had accurately provided the tutor with information they had diligently collected. Feedback needs to be sensitive to the unsophisticated conceptions of learning that may be revealed in students' work.

Students' conception of knowledge

Perry's 'scheme of intellectual and ethical development' describes how students develop over time, and through academic experience, their understanding of what knowledge itself is (Perry, 1970). He describes students as starting off thinking that there are an enormous number of right answers and that their job is to learn these and give them back to the teacher correctly. Perry describes this learning process with the memorable phrase 'quantitative accretion of discrete rightness'. He describes students as moving through a number of stages of increased understanding of the nature of knowledge involving, for example, extreme relativism, in which all answers are seen as equally right. A student who does not draw a conclusion to an essay may be leaving it up to the reader to decide, given that all conclusions are seen as equally valid. Feedback that simply read 'No conclusion' might not help such a student to progress! Teachers' feedback is often (though not always) generated from a more sophisticated epistemological stance than that of the student and this offers plenty of scope for misunderstanding of feedback or blank incomprehension.

Students' conception of the discourse of the discipline

Lea & Street (1998) describe a student who, after submitting an essay on a History course, received the feedback 'I like your conclusions to what is a carefully argued and relevant essay.' At the same time the student received feedback on an essay submitted on a parallel Anthropology course which was so critical of the student's ability to write a clear argument or produce a justified conclusion that they were advised to seek study skills counselling. Lea & Street interpret this as a consequence of Anthropology involving a very different form of discourse involving different forms of argumentation and use of evidence, as it was clearly not a case of generalized essay writing inadequacies. If the student did not understand the discourse of Anthropology and was unpractised in using it, then generalized essay writing advice was unlikely to be helpful, whether from the lecturer or from a study skills counsellor. Feedback needs to be sensitive to what kind of writing is expected and what students are likely to understand about it. In modular course structures it is common for students to cross disciplinary boundaries and have to cope with such differences in discourse. Science and Technology students often have particular difficulties with social science-type essays even if they can write in an articulate way in their own discipline, but there are also profound differences in discourse within the social sciences, for example between Sociology and Psychology, and within the Humanities, for example between History and Literature.

Similarly, Higgins *et al*. (2001) discuss the failures of communication that take place in feedback. They describe a case in which the tutor's entire feedback consisted of: 'A satisfactory effort. More critical analysis of key issues would have helped.' The student, who wanted to be better than 'satisfactory', was left frustrated by the poor quality of critical analysis by the tutor.

Condition 9

Feedback is received and attended to

A number of studies have described students receiving their assignment back, glancing at the mark at the bottom, and then simply throwing it in the bin, including all the feedback.

> 'Sometimes I do read the comments but I find that I'll never write the same essay again anyway … . I tend to ignore them in some ways, unless there is something very startling.'
> (Hounsell, 1987)

Crooks (1988) has summarized a range of research on this issue; where marks on intermediate tests or coursework assignments count significantly towards final marks, students pay less attention to accompanying feedback. Jackson (1995) found that third year students were particularly likely only to look at the grade rather than at feedback on essays. He reported that students like to see the feedback, but more to assure them that their essay had been read carefully and marked fairly.

It is not inevitable that students will read and pay attention to feedback even when that feedback is lovingly crafted and provided promptly. Special steps may need to be taken to engage students with feedback, such as:

- asking students to specify, on their assignment, what they would like feedback on, and giving feedback on nothing else
- providing feedback but no marks, so that students have to read the feedback to get any idea how they are progressing
- requiring assignments to be self-assessed (without any marks being involved) so that students pay attention to whether teachers' views correspond to their own. In a review of literature on self- and peer-assessment, Dochy *et al*. have reported that overt self-assessment has been shown to increase student performance (compared with a control group, in controlled studies) and increase students' control over their learning strategies (Dochy *et al*., 1999)
- using two-stage assignments with feedback on the first stage, intended to enable the student to improve the quality of work for a second stage submission, which is only graded. Cooper (2000) has reported how such a system can improve almost all students' performance, particularly the performance of some of the weaker students
- providing a grade only after self-assessment and tutor feedback has been completed. Taras (2001) reports the successful use of such a sequence as a component of summative assessments.

Condition 10

Feedback is acted upon by the student

This issue concerns the impact of feedback on future learning. Feedback may accurately correct errors but still lead to no change in the way a student goes about the next assignment or tackles any future learning task. This may occur for a variety of reasons:

- feedback may come too late to be acted on by students
- feedback may be backward looking — addressing issues associated with material that will not be studied again, rather than forward-looking and addressing the next study activities or assignments the student will engage with
- feedback may be unrealistic or unspecific in its aspirations for student effort (e.g. 'read the literature' rather than 'for the opposite view, see Smith Chapter 2 pages 24-29')
- feedback may ask the student to do something they do not know how to do (e.g. 'be more Sociological' or 'express yourself more clearly')
- feedback may be context-specific and only apply to the particular assignment rather than concerning generic issues such as study skills or approaches that generalize across assignments
- feedback may be discouraging and lead to less study effort rather than more
- there may be no follow-up to check if students have taken any action, so students can ignore feedback with impunity.

Ding (1998) suggests that even if students read feedback comments, they do little with them. In contrast Brookhart (2001) found that successful students use both marks and feedback and actively self-assess, both to learn and to direct their future studying. The most important variables here may be, as so often, to do with the student rather than with the teacher. Teaching students to monitor their own performance is, in Sadler's theoretical analysis of the role of feedback, the ultimate goal of feedback (Sadler, 1989). Research on the impact of the use of 'classroom assessment' in college in the USA again and again stresses the impact not on the learning of specific content but on the development in students of 'meta-cognition' and the ability to gain control over their own learning (see Steadman, 1998, for a summary). Students are likely to need to be taught how to use feedback to develop meta-cognitive control (Sadler, 1998). Improved ability to learn may not have the effects hoped for, however. Ramsden *et al*. (1987), studying the impact of a 'study skills' programme designed to increase the extent to which students adopted a deep approach, found it actually achieved the opposite. Students' increased awareness enabled them to adopt a surface approach to a greater extent in order to meet the perceived low level demands of their courses' assessment! Again this illustrates the way students' perceptions of assessment influence their learning.

Conclusion

These 'conditions under which assessment supports learning' are in the process of being tested out in practice in the context of a large scale project starting with a study of assessment in science courses at two universities. Teachers of courses with a wide range of assessment practices are collecting evidence from their students about, for example, how they distribute their effort in relation to assessment demands, and how they respond to feedback. They are using this evidence to diagnose potential problems with their courses, making changes to the assessment to address these problems, and then evaluating whether the changes have had positive impacts on the ways their students go about their learning. This is much like any action research process involving the improvement of teaching, but with one major difference: the focus is not on teaching but on assessment. The starting assumption is that there is more leverage to improve teaching through changing aspects of assessment than there is in changing anything else and, at the same time, the teachers know less about how students respond to assessment than about anything else. As this project progresses, teachers' insights and evidence of effective changes to courses will lead to these 'conditions' being developed further. It is a large scale collaborative venture in the 'scholarship of assessment' that will lead both to case studies of changes that turned out to be effective but also to an elaborated conceptual framework that helps to explain why they were effective. The intention is that these conditions can be used as a checklist by any teacher wishing to review and make sense of the effectiveness of their own course's assessment system to support student learning.

Acknowledgements

The literature review that forms the basis of this article was undertaken by the Student Support Research Group at the Open University. The research project that the review was developed for is entitled 'Improving the effectiveness of formative assessment in Science' and is in receipt of £250,000 from the Higher Education Funding Council for England under the Fund for the Development of Teaching and Learning. Their support is gratefully acknowledged.

References

Angelo, T.A. & Cross, K.P. (1993) *Classroom Assessment Techniques: a handbook for college teachers*, San Francisco, Ca: Jossey-Bass.

Baird, L.L. (1985) Do grades and tests predict adult accomplishment? *Research in Higher Education*, vol. 23, no. 1, pp. 3-85.

Berliner, D.C. (1984) The half-full glass: a review of research on teaching, in P.L. Hosford (ed) *Using What We Know About Teaching*, Alexandria, Va.: Association for Supervision and Curriculum Development.

Biggs, J.B. & Collis K.F. (1982) *Evaluating the Quality of Learning: the SOLO taxonomy*, New York: Academic Press.

Black, P. & Wiliam, D. (1998) Assessment and classroom learning, *Assessment in Education*, vol. 5, no. 1, pp. 7-74.

Bridges, P., Cooper, A., Evanson, P., Haines, C., Jenkins, D., Scurry, D., Woolf, H. & Yorke, M. (2002) Coursework marks high, examination marks low: discuss, *Assessment and Evaluation in Higher Education*, vol. 27, no. 1, pp. 36-48.

Brittingham, B.E. (1988) Undergraduate students' use of time: a classroom investigation, *To Improve the Academy*, vol. 7, pp. 45-52.

Brookheart, S. M. (2001) Successful students' formative and summative uses of assessment information, *Assessment and Evaluation in Higher Education*, vol. 8, no. 2, pp. 154-169.

Bruner, J.S. (1974) *Toward a Theory of Instruction*, Cambridge, Mass: Harvard University Press.

Carless, D.M. (2002) The 'mini-viva' as a tool to enhance assessment for learning, *Assessment and Evaluation in Higher Education*, vol. 27, no. 4, pp. 353-363.

Carroll, M. (1995) Formative assessment workshops: feedback sessions for large classes, *Biomedical Education*, vol. 23, no. 2, pp. 65-67.

Chansarkar, B.A. & Raut-Roy, U. (1987) Student performance under different assessment situations, *Assessment and Evaluation in Higher Education*, vol. 12, no. 2, pp. 115-122.

Chickering, A.W. & Gamson, Z.F. (1987) *Seven Principles to Good Practice in Undergraduate Education*, Racine, Wi.: The Johnson Foundation Inc.

Chickering, A.W. & Gamson, Z.F. (1991) *Applying the Seven Principles to Good Practice in Undergraduate Education*, San Francisco: Jossey-Bass.

Cole, S., Coates, M. & Lentell, H. (1986) Towards good teaching by correspondence, *Open Learning*, vol. 1, no. 1, pp. 16-22.

Conway, M.A., Cohen, G. & Stanhope, N. (1992) Why is it that university grades do not predict very long-term retention? *Journal of Experimental Psychology: General*, vol. 121, no. 3, pp. 49-57.

Cook, A. (2001) Assessing the use of flexible assessment, *Assessment and Evaluation in Higher Education*, vol. 26, no. 6, pp. 539-549.

Cooper, N.J. (2000) Facilitating learning from formative feedback in level 3 assessment, *Assessment and Evaluation in Higher Education*, vol. 25, no. 3, pp. 279-291.

Crooks, T.J. (1988) The impact of classroom evaluation practices on students, *Review of Educational Research*, vol. 58, no. 4, pp. 438-481.

Curtis, S. & Shami, N. (2002) The effect of taking paid employment during term-time on students' academic studies, *Journal of Further and Higher Education*, vol. 26, no. 2, pp. 129-138.

Ding, L. (1998) Revisiting assessment and learning: implications of students' perspectives on assessment feedback. Paper presented to *Scottish Educational Research Association Annual Conference, University of Dundee, September 25-26*.

Dochy, F., Segers, M. & Sluijsmans, D. (1999) The use of self-, peer- and co-assessment: a review, *Studies in Higher Education*, vol. 24, no. 3, pp. 331-350.

Dunkin, M.J. (1986) Research on Teaching in Higher Education, in M.C. Wittrock (ed) *Handbook of Research on Teaching*, 3rd ed, New York: Macmillan.

Forbes, D. & Spence, J. (1991) An experiment in assessment for a large class, in R. Smith (ed) *Innovations in Engineering Education*, London: Ellis Horwood.

Gagne, R.M. (1977) *The Conditions of Learning*, 3rd ed, New York: Holt, Rinehart and Winston.

Gardiner, L.F. (1997) Redesigning higher education: producing dramatic gains in student learning, *ASHE-ERIC Higher Education Reports*, vol. 23, no. 7, Washington, DC: Association for the Study of Higher Education.

Gibbs, G. (1992) *Assessing More Students*, Oxford: Oxford Centre for Staff Development.

Gibbs, G. (1995) *Assessing Student Centred Courses*, Oxford: Oxford Centre for Staff Development.

Gibbs, G. (1999) Using assessment strategically to change the way students learn, in S. Brown & A. Glasner (eds) *Assessment Matters in Higher Education*, Buckingham: Society for Research into Higher Education and Open University Press.

Gibbs, G. & Lucas, L. (1987) Coursework assessment, class size and student performance: 1984-94, *Journal of Further and Higher Education*, vol. 21, no. 2, pp. 183-192.

Gibbs, G. & Simpson, C. (2002) *Evaluation of Regional Retention Activity: interim report*, Open University: Student Support Research Group Report 40/2002.

Greer, L. (2001) Does changing the method of assessment of a module improve the performance of a student? *Assessment and Evaluation in Higher Education*, vol. 26, no. 2, pp. 128-138.

Hattie, J.A. (1987) Identifying the salient facets of a model of student learning: a synthesis of meta-analyses, *International Journal of Educational Research*, vol. 11, pp. 187-212.

Higgins, R., Hartley, P. & Skelton, A. (2001) Getting the message across: the problem of communicating assessment feedback, *Teaching in Higher Education*, vol. 6, no. 2, pp. 269-274.

Hounsell, D. (1987) Essay writing and the quality of feedback, in J.T.E. Richardson, M.W. Eysenck & D. Warren-Piper (eds) *Student Learning: research in education and cognitive psychology*, Milton Keynes: Open University Press and Society for Research into Higher Education.

Hutchings, P., Marchese, T., & Wright, B. (1991) *Using Assessment to Strengthen General Education*, Washington, DC: American Association for Higher Education.

Hyland, F. (2001) Providing effective support: investigating feedback to distance language learners, *Open Learning*, vol. 16, no. 3, pp. 231-247.

Innis, K. (1996) *Diary Survey: how undergraduate full-time students spend their time*, Leeds: Leeds Metropolitan University.

Jackson, M. (1995) Making the grade: the formative evaluation of essays [online]. UtiliBASE: available from http://ultibase.rmit.edu.au/Articles/jacks1.html.

Kember, D., Ng, S., Tse, H., Wong, E.T.T. & Pomfret, M. (1996) An examination of the interrelationships between workload, study time, learning approaches and academic outcomes, *Studies in Higher Education*, vol. 21, no. 3, pp. 347-358.

Kniveton, B.H. (1996) Student Perceptions of Assessment Methods, *Assessment & Evaluation in Higher Education*, vol. 21, no. 3, pp. 229-238.

Kulik, C.C., Kulik, J.A. & Cohen, P.A. (1980) Instructional technology and college teaching, *Teaching Psychology*, vol. 7, pp. 199-205.

Lea, M. & Street, B. (1998) Student writing in higher education: an academic literacies approach, *Studies in Higher Education*, vol. 23, no. 2, pp. 157-172.

Macdonald, J. (2002) 'Getting it together and being put on the spot': synopsis, motivation and examination, *Studies in Higher Education*, vol. 27, no. 3, pp. 329-338.

Maclellen, E. (2001) Assessment for learning: the different perceptions of tutors and students, *Assessment and Evaluation in Higher Education*, vol. 26, no. 4, pp. 307-318.

MacFarlane, B. (1992) The 'Thatcherite' generation of university degree results, *Journal of Further and Higher Education*, vol. 16, pp. 60-70.

Marton, F. & Wenestam, C.G. (1978) Qualitative differences in the understanding and retention of the main points in some texts based on the principle-example structure, in M.M. Gruneberg, P.E. Moris & R.N. Sykes (eds) *Practical Aspects of Memory*, London: Academic Press.

McKeachie, W.J., Pintrich, P.R., Lin, Y. & Smith, D. (1986) *Teaching and Learning in the College Classroom: a review of the research literature*, Ann Arbor: National Centre for Research to Improve Postsecondary Teaching and Learning, University of Michigan.

Melton, R. (2000) *Assessment and Assignments — a review of related issues based on data collected from the 1998 courses survey*, Report SRC 200, Open University.

Mentkowski, M., Astin, A., Ewell, P. & Moran, E. T. (1991) *Catching Theory up with Practice: conceptual frameworks for assessment*, Washington DC: AAHE Assessment Forum.

Miller, C.M.I. & Parlett, M. (1974) *Up to the Mark: a study of the examination game*, Guildford: Society for Research into Higher Education.

Moffat, M. (1989) *Coming of Age in New Jersey: college and American culture*, New Brunswick, NJ: Rutgers University Press.

Orsmond, P., Merry, S. & Reiling, K. (2002) The use of exemplars and formative feedback when using student derived marking criteria in peer and self assessment, *Assessment and Evaluation in Higher Education*, vol. 27, no. 4, pp. 309-323.

Paton-Saltzberg, R. & Lindsay, R. (1993) *The Effects of Paid Employment on the Academic Performance of Full-time Students in Higher Education*, Oxford: Oxford Polytechnic.

Penny. A.J. & Grover, C. (1996) An analysis of student grade expectations and marker consistency, *Assessment and Evaluation in Higher Education*, vol. 21, no. 2, pp. 173-184.

Perry, W.G. (1970) *Forms of Intellectual and Ethical Development in the College Years*, New York: Holy, Rinehart and Winston.

Ramsden, P. (1991) A performance indicator of teaching quality in higher education: the Course Experience Questionnaire, *Studies in Higher Education*, vol. 16, pp. 129-150.

Ramsden, P. (1992) *Learning to teach in Higher Education*, London: Routledge.

Ramsden, P., Beswick, D. & Bowden, J. (1987) Learning processes and learning skills, in J.T.E. Richardson, M.W. Eysenck & D.W. Warren-Piper (eds) *Student Learning: research in education and cognitive psychology*, Milton Keynes: Open University Press and Society for Research into Higher Education.

Rice, M., Mousley, J. & Davis, R. (1994) Improving student feedback in distance education: a research report, in T. Evans & D. Murphy (eds) *Research in Distance Education 3: revised papers from the third research in distance education conference*, Geelong, Victoria: Deakin University Press.

Roberts, D. (1996) Feedback on assignments, *Distance Education*, vol. 17, no. 1, pp. 95-116.

Rowntree, D. (1987) *Assessing Students — how shall we know them?* London: Kogan Page

Sadler, D.R. (1989) Formative assessment and the design of instructional systems, *Instructional Science*, vol. 18, pp. 119-144.

Sadler, D.R. (1998) Formative assessment: revisiting the territory, *Assessment in Education*, vol. 5, no. 1, pp. 77-84.

Säljö, R. (1982) *Learning and Understanding*, Goteborg: Acta Universitatis Gothoburgensis.

Sambell, K. & McDowell, L. (1998) The construction of the hidden curriculum: messages and meanings in the assessment of student learning, *Assessment and Evaluation in Higher Education*, vol. 23, no. 4, pp. 391-402.

Schunk, D. (1984) Self-efficacy perspective on achievement behaviour, *Educational Psychologist*, vol. 19, pp. 48-58.

Schunk, D. (1985) Self-efficacy and classroom learning, *Psychology in the Schools*, vol. 22, pp. 208-223.

Scouler, K. (1998) The influence of assessment method on students' learning approaches: multiple choice question examinations vs. essay assignment, *Higher Education*, vol. 35, pp. 453-472.

Scouler, K. & Prosser, M. (1994) Students' experiences of studying for multiple choice question examinations, *Studies in Higher Education*, vol. 19, no. 3, pp. 267-280.

Snyder, B.R. (1971) *The Hidden Curriculum*, Cambridge, MA: MIT Press.

Sly, L. (1999) Practice tests as formative assessment improve student performance on computer managed learning assessments, *Assessment and Evaluation in Higher Education*, vol. 24, no. 3, pp. 339-344.

Starr, J.W. (1970) Student opinion on methods of assessment, *Educational Review*, vol. 22, pp. 243-253.

Steadman, M. (1998) Using classroom assessment to change both teaching and learning, *New Directions for Teaching and Learning*, vol. 75, pp. 23-35.

Stephenson, K., Sander, P. & Naylor, P. (1996) Student perceptions of the tutor's role in distance learning, *Open Learning*, vol. 11, no. 1, pp. 22-30.

Tang, C. (1994) Effects of modes of assessment on students' preparation strategies, in G. Gibbs (ed) *Improving Student Learning: theory and practice*, Oxford: Oxford Centre for Staff Development, pp. 151-170.

Taras, M. (2001) The use of tutor feedback and student self-assessment in summative assessment: towards transparency for students and for tutors, *Assessment and Evaluation in Higher Education*, vol. 26, no. 6, pp. 605-614.

Thomas, J.W., Iventosh, L. & Rohwer, W.D. (1987) Relationships among student characteristics, study activities and achievement as a function of course characteristics, *Contemporary Educational Psychology*, vol. 12, pp. 344-364.

Tinto, V. (1993) *Leaving College: rethinking the causes and cures of student attrition,* 2nd ed, Chicago: University of Chicago Press.

Torrance, H. & Pryor, J. (1998) *Investigative Formative Assessment: teaching, learning and assessment in the classroom*, Buckingham: Open University Press.

Tynjala, P. (1998) Traditional studying for examination vs constructivist learning tasks: do learning outcomes differ? *Studies in Higher Education*, vol. 23, no. 2, pp. 173-191.

Vos, P. (1991) *Curriculum Control of Learning Processes in Higher Education*, 13th International Forum on Higher Education of the European Association for Institutional Research, Edinburgh.

Warren, J.R. (1971) *College Grading Practices: an overview, Report No. 9*, Washington D.C., ERIC Clearinghouse on Higher Education.

Wootton, S. (2002) Encouraging learning or measuring failure? *Teaching in Higher Education*, vol. 7, no. 3, pp. 353-357.

Wotjas, O. (1998) Feedback? No, just give us the answers, *Times Higher Education Supplement*, September 25 1998.

Yorke, M. (2001) Formative assessment and its relevance to retention, *Higher Education Research and Development*, vol. 20, no. 2, pp. 115-126.

Agreement in Student Performance in Assessment

David James & Scott Fleming
University of Gloucestershire, UK

Introduction

Since 1990, there has been increasing attention to assessment in higher education, and there have been various attempts to inform the professional discourse of assessment in the sector. Some have been concerned with the philosophy of assessment and of assessment practice (e.g. Miller *et al*., 1998; Swann & Ecclestone, 1999); others have focused on promoting good practice (e.g. Baume & Baume, 1992; Brown, 2001) and the application of specific examples (e.g. Habeshaw *et al*., 1993; Race, 1995; 1996). It has also been claimed that student learning is assessment-driven (Habeshaw *et al*., 1993), and even that assessment is of *singular* importance to the student experience (Rust, 2002).

The rationale underpinning effective assessment in higher education, as well as its importance, have both been widely explored (e.g. Race, 1995). Broadly, the key features include: diagnosis (of different kinds); evaluation of progress, providing feedback (to learners, tutors and external agencies); motivation; demonstration of the acquisition of skills and/or competencies and measuring achievement. It is now a widespread view that multiple methods of assessment should be used for multiple assessment expectations (Brown & Knight, 1994), and that students should experience a wide and varied 'assessment diet' within a programme of study[1]. Brown *et al*. (1996, p.14) explain: 'Assessment that is 'fit for purpose' uses the best method of assessment appropriate to the context, the students, the level, the subject and the institution.'

Innovation in assessment practice has been endorsed by different agencies (e.g. Institute for Learning and Teaching in Higher Education, Learning and Teaching Support Network), but there is still a culture of traditionalism in many universities. As recently as 1996, Brown *et al*. reported that over 80% of assessment in universities is based on essays, reports and traditional, timed, unseen examinations and,

as Buswell (2002) has noted, traditional unseen examinations (as well as coursework essays) may stifle some principles of innovative assessment. Some of the impetus for innovation in assessment has also been developed through concerns about the prevalence of plagiarism and other forms of 'dishonesty in assessment' (Yorke *et al*., 2000; Larkham & Manns, 2002)[2]. Other examples of more educationally progressive forms of innovation in assessment (e.g. Fullerton, 1995) are often connected to good practice in feedback to students (e.g. Cantwell, 2002). Assessment, as Race (1995, p.82) observes, 'at best, is a very inexact science'. Inevitably, and quite properly, validity and reliability in assessment continue to be emphasized, although evaluation of the degree of validity and reliability is rarely undertaken. In those institutions where a wide range of assessment methods is practised, suggestions of differential levels of performance often raise questions of comparability.

At the module level, evidence for different performance across assessment points is superficial. Yorke *et al*. note the general perception that 'coursework marks tend to be higher than marks awarded in examinations' (2000, p.14; *Eds: see also Gibbs & Simpson, this issue*), and they point to some preliminary evidence to that effect. These matters need to be considered with some care and rather more attention to detail than has often been the case hitherto. Leaving aside some of the technical debates about whether students' assessment data are actually interval or ordinal data (cf. Yorke, 2001), there are some important implications for modular programmes in particular. For instance, the diversity of assessment practice across different disciplines and subject areas raises profound questions about equity — especially, Yorke *et al*. (2000) claim, in modular schemes.

While distinctions are often made between the natural and social science subjects, it is perhaps useful to first consider the range of assessment tasks employed within a discipline, and their effect on performance. For example, within a discipline such as exercise physiology, which forms part of most sport and exercise science programmes, a range of assessment tasks is often employed, drawing on essay type, mathematically-based and practical skill assessments. Therefore, an initial question when exploring the broad area of performance in assessment might usefully be: 'how does student performance vary across assessment tasks within the same module?' A second question would be: 'how does student performance vary in the same assessment task across modules?' Such variation in student performance is perhaps best thought of as the level of 'agreement' in performance.

Previous traditional attempts to investigate 'agreement' in performance (of any type) have involved significance difference tests and intraclass correlation coefficients (Bland & Altman, 1990), but neither of these approaches is suitable, and both limit the extent to which the findings are meaningful (see Technical Note). There have, however, been recent advances in statistical techniques suitable for examining 'agreement' in student performance (Bland & Altman, 1986). Specifically, a 'limits of agreement' approach, widely used in medicine and sport science (Webber *et al*., 1994; Atkinson & Nevill, 1998), is suggested as a 'user-friendly' and robust way to undertake this analysis.

Assessment of a student's performance on a particular module may often be thought of as a single evaluation of the extent to which the student has met some or all of the module's learning outcomes. More helpfully, however, when there is more than one assessment task in a particular module, it may be thought of as the combination of different assessment tasks (whatever the weighting attached to each of them). In this sense, the level of agreement between performances on the different tasks may elucidate the nature of overall student performance further still. Typically, assessment tasks within a module tend not to be of the same kind; often they are complementary, sometimes through the use of different media. The primary aim of the present study was to investigate agreement in student performance between assessment tasks *within* two modules.

Additionally, however, many modules adopt conventional combinations of assessment tasks. Previously, in the social sciences and the humanities for example, this might have been a written essay and unseen examination (containing essay questions) or, more recently, perhaps a group poster presentation and individual seminar (accompanied by written paper). There has been little substantive exploration of the level of agreement in student performance in similar tasks across different modules. A secondary aim, therefore, was to investigate agreement in student performance in the same assessment task *between* modules from similar disciplines (i.e. Anatomy and Physiology).

As a final but important contextualising note, the nomenclature adopted for the statistical techniques that underpin this study is, of course, value-laden. 'Agreement' should not necessarily be interpreted as a virtue in this regard, anymore than 'failure to establish agreement' (or the even more pejorative term 'disagreement') should be regarded as a deficiency or shortcoming in assessment protocols. There are important reasons why, for

example, 'within module' assessments tasks might not demonstrate agreement — they might be examining different skills, competencies and knowledge through different media. There are also reasons why similar tasks from different modules might evidence differential patterns of student performance — they may involve conceptually different material requiring different kinds of cognitive competencies. Examination of the extent to which agreement exists within a module's assessment protocol, or between similar tasks in different modules, however, may signal some important characteristics about the diet of student assessment experiences, and of their performance. The levels of agreement may, therefore, provide a basis for more nuanced and context-sensitive examination of student assessment. This is a theme to which the discussion will return in the conclusion.

Method

Study design

The sample for this study was drawn from two modules, both of which form part of the introductory level curriculum for students undertaking one of the three 'science' programmes of study in sport and exercise[3]. The two modules were 'Anatomy and Assessment of Structure' (hereafter referred to as Anatomy) and 'Introduction to Physiology of Sport, Exercise & Health' (hereafter referred to as Physiology). The basis for selection of this sample reflected the need to assess agreement across assessment points within a module, and across modules. When looking at agreement across modules, it was possible to assess student performance in the same type of assessment. To ensure potential confounding variables were minimized, the modules were taken from the same level of study, and ran in the same academic year (2000-2001). There were 267 students registered for the Anatomy module, and 196 for the Physiology module. A total of 180 were registered for both.

Student performance was assessed on each module through three assessment points. Agreement of performance within each module was assessed by comparing performance in each assessment point against each other assessment point in turn. This resulted in three comparisons within each module. Additionally, two of these assessment points were similar when comparing the two modules, which also allowed cross module comparisons of student performance. Specifically, the common assessment points were a Skills Test and an Examination. In the case of the Physiology module, the other assessment point was a Laboratory Report, and in the case of the

Anatomy module, the other assessment point was a Practical File. In both modules, the Examination was multiple-choice, and of one-hour duration, the only difference being that the Anatomy Examination was computer-based, whereas the Physiology Examination was a traditional paper-based examination.

The Skills Test was a practical test, designed to assess a student's ability to undertake skills developed through the module. There were four different skills testing stations, and the test required that each student spend a maximum of ten minutes at one of them. Students had prior knowledge of the skills upon which they would be assessed, but were randomly assigned to one of the stations on arrival for the test. The Anatomy Skills test required the students to identify an anatomical landmark and to measure a length, breadth, girth or skin-fold. The Physiology Skills Test required the students to undertake assessment of lung function, blood pressure, minute ventilation or a progressive exercise protocol whilst complying with health and safety guidelines.

Data analysis

Student performance data (i.e. percentage marks) for each assessment point were acquired from central student records of electronic module results. The data were initially cleaned by removing student marks when no attempt was made at an assessment point, and were then sorted by student identification number in order to match students across modules. This process allowed for the deletion of marks if a student was not registered on both modules. Clearly this was only necessary when student performance was compared across modules. Once paired data were available after the initial cleaning, it was no longer necessary to store students' identification numbers.

The cleaned data were then used to assess agreement between assessment points, following the procedure described by Bland & Altman (1986; see Technical Note). The first part of this process involved calculating the arithmetic mean mark for each student, and the difference between the two marks for each student. The arithmetic mean of the differences was then calculated, and used to represent the accuracy or 'bias'. The standard deviation (SD) of the differences was also calculated, and used to represent the precision or 'agreement'. Normally the extent of agreement is represented as 95% confidence intervals (i.e. 1.96 × SD), and the findings are presented through a 'limits of agreement plot' for each comparison[4].

However, in the case of many comparisons in the present study, the limits of agreement plot showed a clear trend in the data, such that the differences (plotted on the y-axis) increased or decreased as the arithmetic mean performance (plotted on the x-axis) increased. This is a common finding when examining agreement data (Bland & Altman, 1999), so an approach was adopted to account for the trend. Accounting for the trend is necessary, since failure to do so results in a meaningless value for bias and an exaggerated value for agreement. The approach for accounting for the trend involved fitting a least squares regression line to the limits of agreement plot. The equation of the regression was used to remove the trend from the data, allowing revised differences to be calculated. These differences were then used to determine agreement (i.e. 1.96 × SD) around the regression line, and plotted on the original limits of agreement plot, the bias then being reflected by the regression line.

Ethics statement

The University's principles and procedures on research ethics were adhered to throughout the study. In particular, data on student performance were presented such that identification of individual student performance was impossible, thereby complying with the requirements of the Data Protection Act. Restricted access to the data is permitted only to those who have administrative (e.g. data collation and processing) and academic functions (e.g. management roles with teaching, learning and assessment responsibilities, roles overseeing pastoral responsibility and course leaders). In this instance one of the authors had joint responsibility for teaching, learning and assessment within the School of Sport & Leisure.

Results

The findings are considered by first examining agreement of assessment within a module (Anatomy followed by Physiology module), followed by agreement of assessment between modules (Skills Test followed by Examination). In all cases, the findings are presented as figures (limits of agreement plots) and in the form of summary tables. Throughout, the application of legends to figures, and headings to tables, shows which assessment point is subtracted from another to give the bias. For example, 'Practical File – Exam' identifies that the Examination score is subtracted from the Practical File score to give the bias.

The second summary table for each module considers students' performance across the assessment points being compared. For example, a 'high' level of performance is indicated by an arithmetic mean score of greater than 70% in the two assessment points being compared. A 'low' level of performance represents a score of less than 40%, and a medium level of performance represent a score of ~55%.

Anatomy module

Agreement between the student performance in the Practical File and Examination is shown in Figure 1 (top panel) and summarized in Tables 1 and 2 respectively. With the exception of the students with the high level of performance, performance in the Examination was stronger than performance in the Practical File. The general trend is that as the students' overall performance deteriorates (moving from high to low levels of performance), the performance in the Practical File gets relatively weaker, and the performance in the Examination gets relatively stronger. The limits of agreement plot shows an agreement of ±32.9% between the Practical File and the Examination.

Table 1: Bias when comparing assessment performance within the Anatomy module

Performance level	**40%**	**50%**	**60%**	**70%**
Practical File – Examination	-10.0%	-6.1%	-1.6%	2.9%
Practical File – Skills Test	-9.4%	-12.0%	-14.7%	-17.4%
Skills Test – Examination	-4.0%	2.7%	9.4%	16.1%

Agreement between the student performance in the Practical File and Skills Test is shown in Figure 1 (middle panel) and summarized in Tables 1 and 2 respectively. In general, performance in the Skills Test was stronger than performance in the Practical File. The general slight trend is that as the students' overall performance deteriorates (moving from high to low levels of performance), the performance in the Practical File ceases to be so relatively weak, and the performance in the Skills Test ceases to be so relatively strong. The limits of agreement plot shows an agreement of ±43.1% between the Practical File and the Skills Test. In this module, the agreement between the two non-examination assessment points demonstrated greater bias at the good performance extreme but, perhaps more importantly, considerably greater lack of agreement across the entire performance range.

Figure 1: Agreement between student performance in three assessment points in Anatomy module

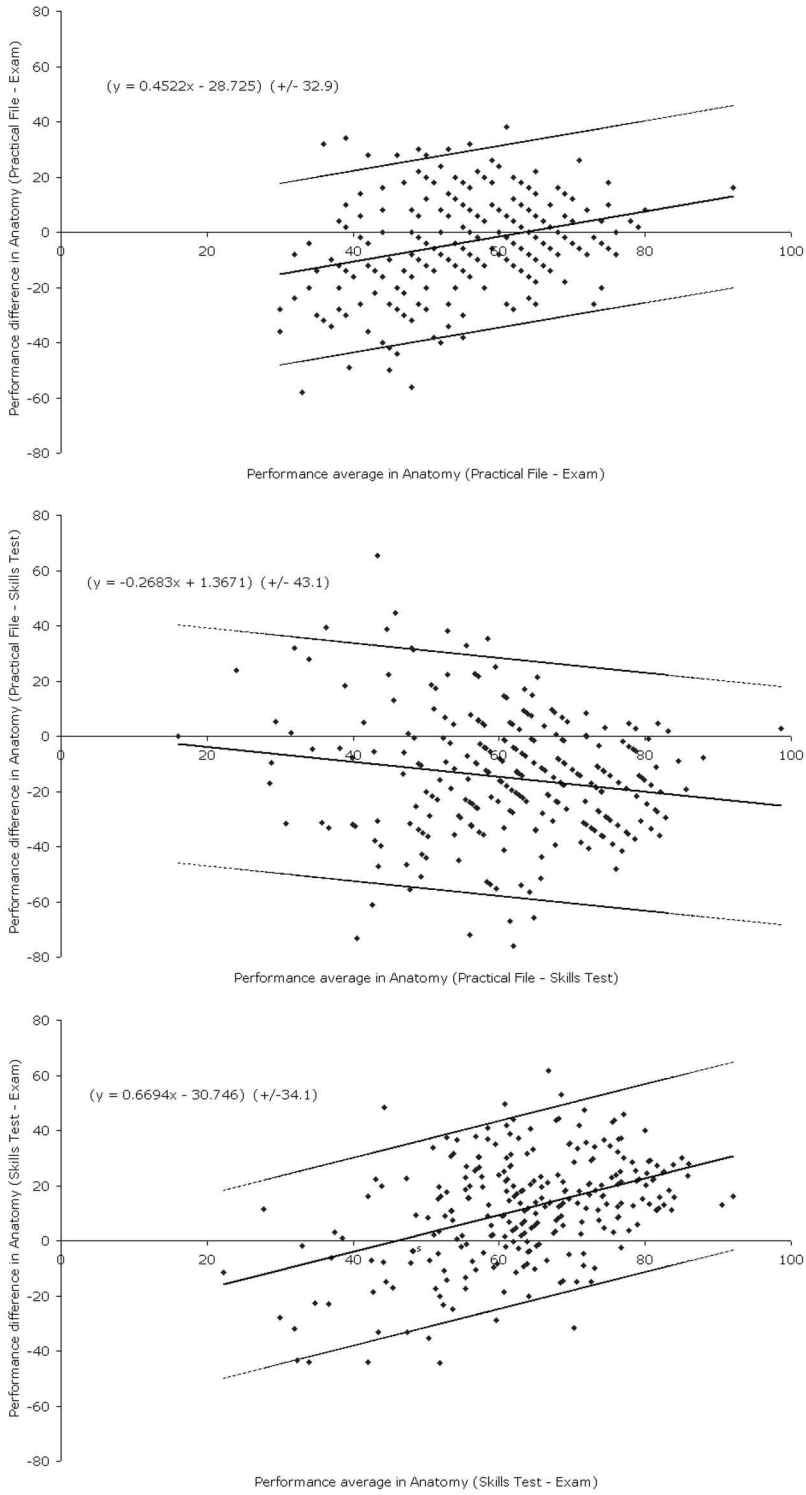


Table 2: Summary of assessment performance in the Anatomy module according to performance category

Low performance	Medium performance	High performance
Examination > Practical File	Examination > Practical File	Practical File > Examination
Skills Test > Practical File	Skills Test > Practical File	Skills Test > Practical File
Examination > Skills Test	Skills Test > Examination	Skills Test > Examination

Agreement between the student performance in the Skills Test and Examination is shown in Figure 1 (bottom panel) and summarized in Tables 1 and 2 respectively. The general trend is that as the students' overall performance deteriorates (moving from high to low levels of performance), the performance in the Skills Test gets relatively weaker and the performance in the Examination gets relatively stronger. The limits of agreement plot shows an agreement of ±34.1% between the Skills Test and the Examination.

An overall rank order of relative performance in assessment tasks therefore indicates that, in general, students performed better in the Skills Test than in the Examination, and better in the Examination than in the Practical File. However, it is interesting to note that students with a low level of performance tend to do better in the Examination, relative to other points of assessment (see Table 2).

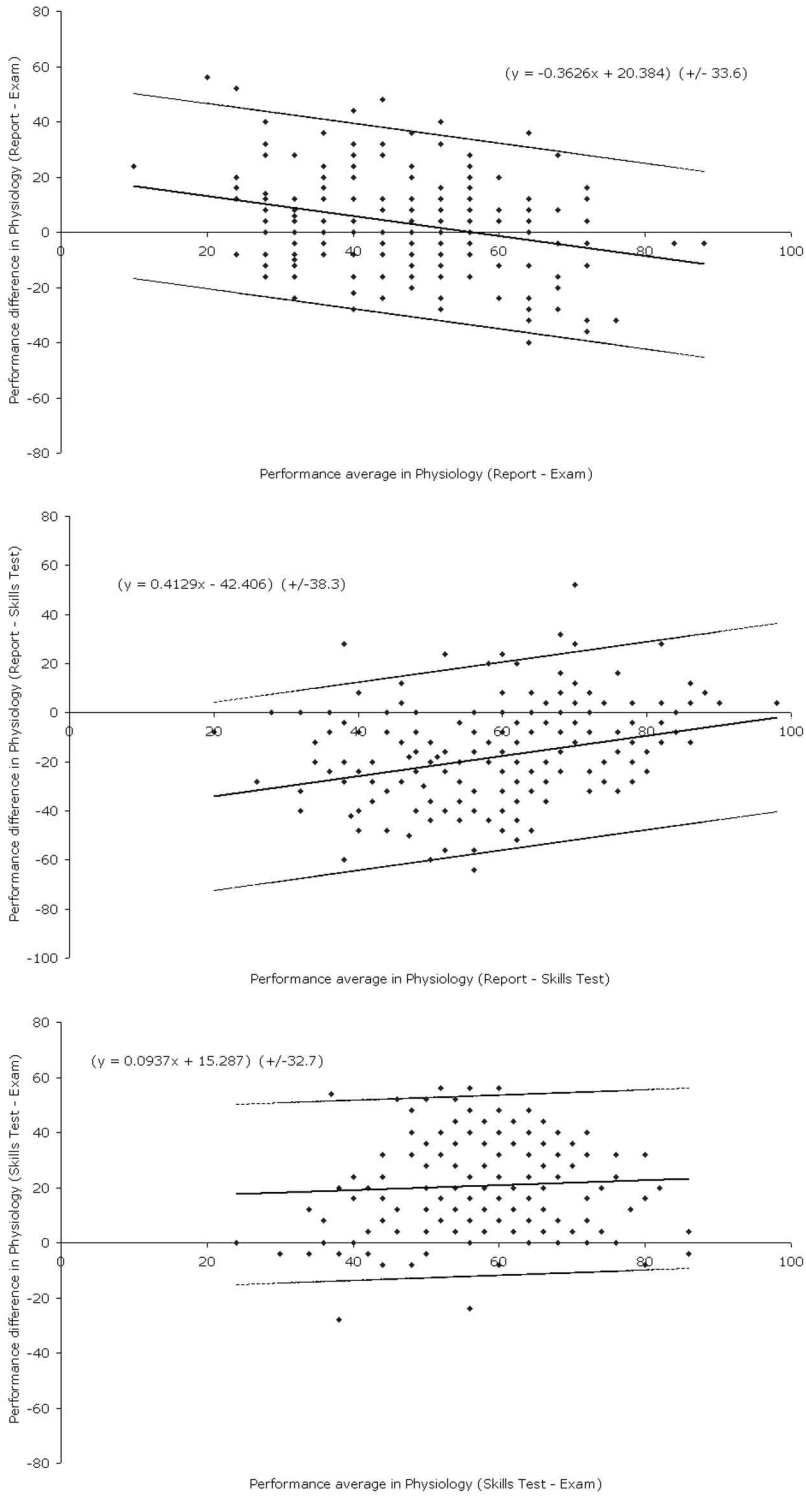
Physiology module

Agreement between the student performance in the Report and Examination is shown in Figure 2 (top panel) and Tables 3 and 4 respectively. The students with a high level of performance tended to perform relatively better in the Examination, whereas the students with a low level of performance tended to perform relatively better in the Report. The limits of agreement plot shows an agreement of ±33.6% between the Report and the Examination.

Table 3: Bias when comparing assessment performance within the Physiology module

Performance level	40%	50%	60%	70%
Report – Examination	5.9%	2.3%	-1.4%	-5.0%
Report – Skills Test	-25.9%	-21.8%	-17.6%	-13.5%
Skills Test – Examination	19.0%	20.0%	20.9%	21.8%

Figure 2: Agreement between student performance in three assessment points in Physiology module



Agreement between the student performance in the Report and Skills Test is shown in Figure 2 (middle panel) and in Tables 3 and 4 respectively. Throughout the range of student performance (i.e. low to high level of performance), the bias suggests that students perform poorly in the Report relative to the Skills Test. Also, the general trend was that as students' overall performance deteriorated (moving from high to low levels of performance), students tended to perform relatively worse in the Report. The limits of agreement plot shows an agreement of ±38.3% between the Report and the Skills Test. In this module, the agreement between the two non-examination assessment points demonstrated greater bias at the poor performance extreme and, interestingly, a greater lack of agreement across the performance range. The greater lack of agreement and considerable bias is a feature shared with similar assessment points in the Anatomy module.

Table 4: Summary of assessment performance in the Physiology module according to performance category

Low performance	Medium performance	High performance
Report > Examination	Report > Examination	Examination > Report
Skills Test > Report	Skills Test > Report	Skills Test > Report
Skills Test > Examination	Skills Test > Examination	Skills Test > Examination

Agreement between the student performance in the Skills Test and Examination is shown in Figure 2 (bottom panel). Throughout the range of student performance (i.e. low to high level of performance), the bias suggests that students perform poorly in the Examination relative to the Skills Test. Also, the general slight trend was that as students' overall performance deteriorated (moving from high to low levels of performance), students tended to perform relatively worse in the Skills Test. The limits of agreement plot shows an agreement of ±32.7% between the Skills Test and the Examination. The positive bias, whereby students perform better in the Skills Test rather than the Examination, is a striking feature of this comparison.

An overall rank order of relative performance in assessment tasks therefore indicates that, in general, students performed better in the Skills Test than in the Report, and better in the Report than in the Examination. However, a distinction is evident between high performing students and others, in that the Examination performance is better than the Report performance (see Table 4).

Skills test

Agreement between the student performance in the Anatomy and Physiology Module is shown in Figure 3 and summarized in Table 5. The general slight trend was that as students' overall performance deteriorated, students tended to perform relatively worse in Anatomy. It is worth mentioning at this point that Anatomy and Physiology took place in different semesters, and any comparison might usefully note this potential confounding variable. The limits of agreement plot shows an agreement of ±41.5% between the Anatomy and Physiology modules.

Figure 3: Agreement in student performance in a skills test in Anatomy and Physiology modules

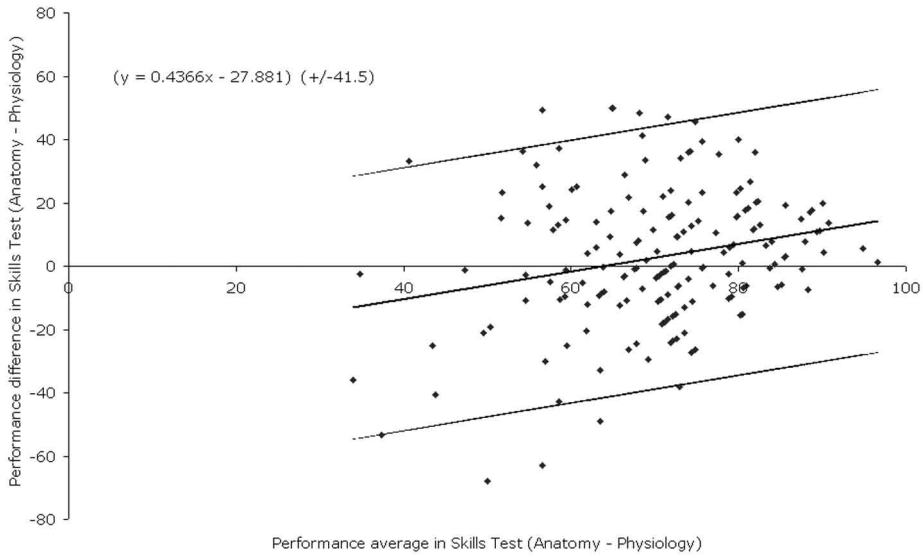
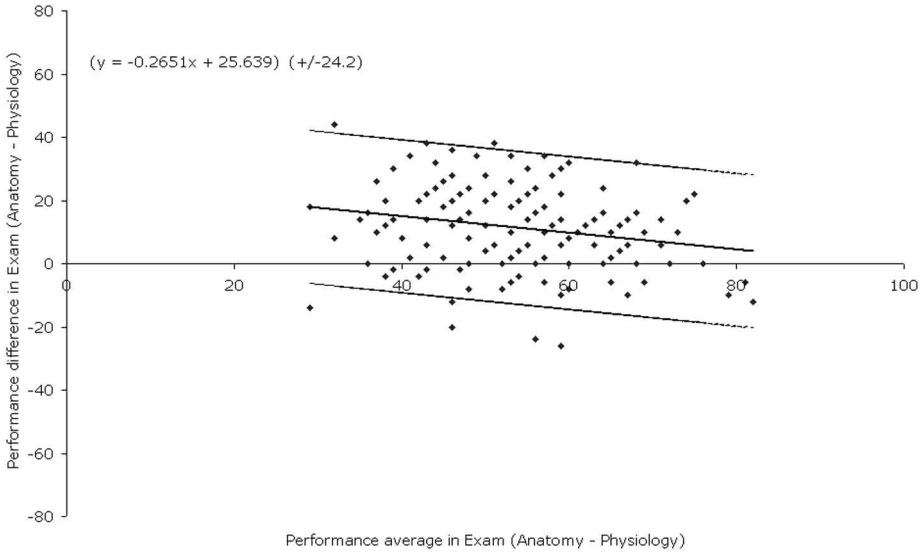


Table 5: Bias when comparing assessment performance between the Anatomy and Physiology modules

Performance level	**40%**	**50%**	**60%**	**70%**
Skills Test (Anatomy – Physiology)	-10.4%	-6.1%	-1.7%	2.7%
Examination (Anatomy – Physiology)	15.0%	12.4%	9.7%	7.1%

Figure 4: Agreement in student performance in an examination in Anatomy and Physiology modules



Examination

Agreement between the student performance in the Anatomy and Physiology Module is shown in Figure 4. The general slight trend was that as students' overall performance deteriorated, students tended to perform relatively worse in Physiology. Through the range of student performance, however, performance tended to be relatively better in the Anatomy Examination (i.e. positive bias). The limits of agreement plot shows an agreement of ±24.2% between the Anatomy and Physiology modules. Interestingly, this agreement is considerably better than that for the Skills Test.

Discussion

Within UK higher education, students may be exposed to a range of assessment tasks within a programme of study, including examinations of various types, report writing, essay writing, poster presentations and oral presentations. Anecdotally, it is often claimed that, regardless of knowledge and understanding, performance of an individual student may vary according to the particular type of assessment task (Yorke *et al*., 2000). It is also claimed that certain types of assessment are more difficult for all students, and even that students may select modules on the basis of the assessment tasks involved. If claims about lack of agreement in student performance

between assessment tasks are true, students might be supported differentially and appropriately depending on the assessment task they struggle with. Alternatively, the assessment tasks themselves might require revision and the performance of the assessors might require investigation. Before any action may be recommended, such claims need to be investigated systematically.

The present study examined the agreement in performance in different assessment tasks within a module, and the same assessment tasks between modules. In order to control potential confounding variables, in making comparisons across modules, similar discipline modules were selected, assessment tasks were well matched (e.g. multiple choice examination in both cases) and modules took place at the same level of study, but within different semesters. In making comparisons within modules, the same assessors were involved in different assessment points. A particularly useful feature of the present study was the large data set involved in each analysis, resulting in meaningful findings for the population under consideration.

Contrary to the view that students do consistently better in one form of assessment compared with another (Yorke *et al*., 2000), the findings from the present study suggest that this is not the case. When comparing performance in two assessment points within each module, relative student performance varies as a function of the average mark from the two assessment points. For example, a student with a low level of performance in the Anatomy module performed relatively better in the Examination than in the Practical File. The converse is true for a student with a high level of overall performance in the module. Within a module, the only comparison of two assessment points that yielded a consistent bias across the assessment range was when the performance in the Skills Test and Examination was compared in the Physiology module. In this case, students consistently scored better in the Skills Test. The notion that examinations yield lower levels of performance than other forms of assessment (Yorke *et al*., 2000) is not evident from the present study. The relative performance in the examination appears, in general, to be a function of student level of performance.

A further commonly made claim, that strong students score relatively better in an examination (Elton & Johnston, 2002), also appears not to be the case. Students with a high level of performance scored relatively worse in the Examination in the Anatomy module, regardless of which other assessment point the Examination is compared with, whereas students with a high level of performance scored relatively better in the Examination in the Physiology module when compared with the Report.

A claim that students generally perform consistently in the same types of assessment may be challenged based on the findings of the present study. For example, when comparing the performance in the Skills Test, it was clear that students with a lower level of performance scored better in the Physiology Skills Test, whereas the students with a higher level of performance scored better in the Anatomy Skills Test. Whilst student performance in the Anatomy Examination tended to be better than performance in the Physiology Examination, relative performance still varied as a function of average student performance. For example, a student scoring an average of 70% in the two assessment points would score 7.1% higher in the Anatomy Examination, whereas a student scoring an average of 40% would score 15.0% higher in the Anatomy Examination.

It is not possible to claim that one module was more challenging than another in the present study. It is not even possible to claim that students with lower levels of performance found one module more difficult than another, since performance varies differently as a function of level of student performance, depending on the form of assessment examined.

Conclusion

The results of the present study, which have so far been discussed in terms of the bias in performance, challenge several commonly-held beliefs. First, students do not consistently perform better in one form of assessment compared with another. We have shown that relative performance in assessment is generally a function of student level of performance. Second, students with a higher level of performance do not tend to do better in examinations. We have shown that this was the case in one module (Anatomy) but not in another module (Physiology). Third, whether students are low or high level performers, performance in a form of assessment is not consistent, even within the same broad discipline. In other words, the performance of the same student is neither consistently good in examinations, nor consistently bad in examinations. Examination performance appears to be a function of the discipline, as well as the student level of performance.

When examining the degree of performance agreement between assessment points, we found that agreement between the three assessment points in the two modules examined was broadly similar. So within a module of the discipline examined in the present study,

it may be claimed that student performance between assessment points agreed by about ±33%. The only clear exception to this was the lower level of agreement between the Practical File and Skills Test in the Anatomy module (±43.1%). The level of agreement is not a function of the student level of performance, so no claims about students with a higher level of performance showing greater levels of agreement may be advanced.

When examining the degree of performance agreement between similar assessment points in different modules, we found that agreement between the assessment points varied according to the type of assessment. Agreement was better for the Examination (±24.2%) than for the Skills Test (±41.5%). It is interesting to note that there is generally no less agreement when comparisons are made between assessment points within a module, compared with similar assessment points between modules.

In summary, despite some of the prevalent beliefs about assessment in higher education, in the modules examined in the present study, students did not perform consistently better in one particular form of assessment. Students who showed different levels of performance (e.g. high versus low) did not appear consistently to do better in a particular form of assessment. Finally, performance was extremely variable, with agreement in most comparisons not being better than ±30%. Further research is required to examine agreement in performance in different disciplines, and between different levels of study. Once a comprehensive examination of agreement in student performance has been conducted, researchers and practitioners will be better placed to ask informed questions. Such questions might include:

- is performance agreement a useful indicator within and between modules?
- are interventions necessary to influence performance agreement?
- should the variety of assessment modes be determined by student choice?
- should assessment of performance agreement be part of routine evaluation of modules and courses?

Acknowledgements

This research was funded by a University of Gloucestershire Faculty Scholarship of Learning and Teaching (SoLT) Project Grant. We are grateful to Dr Dan Wood for his advice on statistical methods employed in this study, and to Professor Clare Morris for her constructive comments on an earlier draft.

Technical note

Traditionally, agreement between methods of measuring something (in this case student knowledge or competence) has been assessed inappropriately by using product moment correlation coefficient (*r*) and significance tests. Correlation is appropriately used to assess the strength of a relationship between two variables. However, such a relationship provides little useful information about agreement. Correlation is inappropriate for assessment of agreement between methods, for the following reasons (adapted from Bland & Altman, 1986):

1. A perfect relationship, as indicated by an *r*-value of 1.00, may be attained with extremely poor agreement. For example, when viewing a scatter plot of one method of measurement plotted against another, it is only the extent to which the data points fall close to the line of identity that indicates agreement. A high *r*-value may be achieved with data points far away from the line of identity.

2. The strength of a relationship is influenced by the range of numerical values in a sample. For example, if student marks in a sample ranged between 40% and 70%, the strength of the relationship would be very different from a sample with a mark range of 0% - 100%, regardless of the degree of agreement.

3. The statistical significance of a relationship indicates little about agreement. It is highly likely that two methods of measurement of the same thing (in this case student knowledge or competence) will be related, as demonstrated through a statistical significance test.

An appropriate approach for the assessment of agreement between methods is to plot the difference between the methods (y-axis) against the mean value of the two methods (x-axis), as in Figure 1. For example, if one student scored a mark of 65% in a physiology report, and 71% in a physiology exam within the same module, the

difference is reported as 6% and the mean is reported as 68%. A data point is then plotted for this student. Once data points have been plotted for all students in the sample (i.e. on the module), the mean and standard deviation of the differences is calculated. The mean of the differences represents the 'bias', and the standard deviation of the differences represents the 'agreement'.

It is suggested that the degree of agreement is expressed as a 95% confidence interval, and illustrated on the plot. The 95% confidence interval is calculated by multiplying the standard deviation by 1.96, providing the data are normally distributed. However, should the data not be normally distributed, a multiplication by 2.00 is recommended (Bland & Altman, 1986). The 95% confidence intervals illustrate that one can be 95% confident that in the population from which the sample was drawn, agreement will be contained within these limits. Having undertaken this procedure, the researcher or practitioner should normally then ask the question 'Is this level of agreement appropriate?'

Notes

1. For a comprehensive annotated list of assessment modes and methods, see Brown (2001).
2. There is already a sophisticated network of websites providing students with the opportunity to buy and download written essays, e.g. http://www.termpapers4u.com/ and http://www.papersheaven.com/.
3. Validated in 2000, the University of Gloucestershire's portfolio of sport and exercise-related provision includes three named BSc (Honours) awards in Sport & Exercise Sciences, Sport Science & Exercise and Health Sciences.
4. A 95% confidence interval is derived from a sample of normally distributed data points, and defines the interval within which 95% of data points are contained.

References

Atkinson, G. & Nevill, A.M. (1998) Statistical methods in assessing measurement error (reliability) in variables relevant to sports medicine, *Sports Medicine*, vol. 26, pp. 217-238.

Baume, D. & Baume, C. (1992) *Assessing Students' Work*, Oxford: Oxford Brooks University, Oxford Centre for Staff Development.

Bland, J.M. & Altman, D.G. (1986) Statistical methods for assessing agreement between two methods of clinical measurement, *Lancet,* i, pp. 307-310.

Bland, J.M. & Altman, D.G. (1990) A note on the use of the intraclass correlation coefficient in the evaluation of agreement between two methods of measurement, *Computers in Biology and Medicine*, vol. 20, no. 5, pp. 337-340.

Bland, J.M. & Altman, D.G. (1999) Measuring agreement in method comparison studies, *Statistical Methods in Medical Research,* vol. 8, pp. 135-160.

Brown, G. (2001) *LTSN Generic Centre Assessment Series No.3: Assessment: A Guide for Lecturers*, York: Learning & Teaching Support Network.

Brown, S. & Knight, P. (1994) *Assessing Learners in Higher Education*, London: Kogan Page.

Brown, S., Race, P. & Smith, B. (1996) *500 Tips on Assessment*, London: Kogan Page.

Buswell, J. (2002) Examinations assessed! Diversity in assessment, *Link,* vol. 5, pp. 17-19.

Cantwell, J. (2002) Formative feedback, *Link,* vol. 5, pp. 15-16.

Elton, L. & Johnston, B. (2002) *Assessment in Universities: A Critical Review of Research*, York: Learning and Teaching Support Network.

Fullerton, H. (1995) Embedding alternative approaches in assessment, in P. Knight (ed) *Assessment for Learning in Higher Education*, London: Kogan Page, pp. 111-123.

Habeshaw, S., Gibbs, G. & Habeshaw, T. (1993) *53 Interesting Ways to Assess your Students*, 3rd ed, Bristol: Technical and Educational Services.

Knight, P. (ed) (1995) *Assessment for Learning in Higher Education*, London: Kogan Page.

Larkham, P. & Manns, S. (2002) Plagiarism and its treatment in higher education, *Journal of Further and Higher Education,* vol. 26, no. 4, pp. 339-349.

Miller, A.H., Imrie, B.W. & Cox, K. (1998) *Student Assessment in Higher Education: a handbook for assessing performance*, London: Kogan Page.

Race, P. (1995) The art of assessing 1, *New Academic,* vol. 4, no. 2, pp. 3-6.

Race, P. (1996) The art of assessing 2, *New Academic,* vol. 5, no. 1, pp. 3-6.

Race, P. & Brown, S. (2001) *The Lecturer's Toolkit*, 2nd ed, London: Kogan Page.

Rust, C. (2002) The impact of assessment on student learning: how can research literature practically help to inform the development of departmental assessment strategies and learner-centred assessment practices, *Active Learning in Higher Education,* vol. 3, no. 2, pp. 145-158.

Swann, J. & Eccelstone, K. (1999) Improving lecturers' assessment practice in higher education: a problem-based approach, *Educational Action Research,* vol. 7, no. 1, pp. 63-87.

University of Gloucestershire (2002) *University Policy on Teaching, Learning and Assessment*, Cheltenham: Centre for Learning and Teaching.

Webber, J., Donaldson, M., Allison, S.P., & MacDonald, I.A. (1994) A comparison of skinfold thickness, body mass index, bioelectrical impedance analysis and dual X-ray absorptiometry in assessing body composition in obese subjects before and after weight loss, *Clinical Nutrition,* vol. 13, pp. 177-182.

Yorke, M. (2001) *Assessment: A Guide for Senior Managers*, York: Learning and Teaching Support Network.

Yorke, M., Bridges, P. & Woolf, H. (2000) Mark distributions and marking practices in UK higher education: some challenging issues, *Active Learning in Higher Education*, vol. 1, no. 1, pp. 7-27.

Assessment of Student Learning: promoting a scholarly approach

LORRAINE STEFANI
University of Auckland, New Zealand

Introduction

Assessment of student learning matters more than ever in the changing world of higher education and with changing expectations society has of its university graduates. For this reason, it is imperative that all staff involved in supporting student learning (particularly new academic staff starting their lecturing careers) are enabled to understand the fundamental principles of assessment of student learning, so that they in turn can endeavour to enhance student learning through effective assessment (Stefani, 1998). The aim of this paper is to present a model for promoting the scholarship of assessment which highlights that assessment is an integral aspect of student learning.

Over the past decade or so, there has been considerably more focus on the nature and quality of university teaching. While it has always been the case that teaching in higher education matters greatly, there has been a tendency for the research element of an academic's career to take precedence over their teaching responsibilities. However, there are increasing pressures on academic staff to provide an effective teaching and learning environment for all students. The advent of new technologies, for example, has resulted in 'the knowledge explosion', with the consequence that information is much more transient than was previously considered to be the case. While historically universities were primarily concerned with the 'transmission' of knowledge, culture and values through teaching (Hattie & Marsh, 2000), it is now the case, in a knowledge rich society, that a very different approach to teaching must be taken, to enable learners to develop a different skill set and be able to seek out, analyse and evaluate information (Breivik, 1998), rather than to simply accumulate knowledge.

There have been government demands for higher education to promote the concept of lifelong learning (Longworth, 1999), highlighting the need for a continuous ability to update one's

knowledge, skills and understanding in any career pathway. There have also been global shifts towards mass higher education and widening participation, putting pressure on universities to develop a greater understanding of what it means to facilitate the learning of a more diverse student population. This in turn has put pressure on academic staff to develop an understanding of what it means to design, develop and deliver an accessible curriculum which acknowledges student diversity (Stefani & Matthew, 2002).

While these changes have been dramatic, there have been other developments relating to higher education, which have focused on the promotion of teaching in higher education and are redressing the balance between the privileged status of disciplinary-based research and the facilitation of student learning (Stefani, 2004).

For example, in the UK, the report of the National Committee of Inquiry into Higher Education (1997), otherwise known as the Dearing Report, highlighted the growing need to explicitly reward excellence in teaching and proposed that greater attention be paid to the skills development of academic staff involved in teaching students. The Dearing Report also proposed the initiation of the Institute for Learning and Teaching in Higher Education (ILTHE), to provide a professional body specifically relating to higher education teaching.

To become a member of the ILTHE (which has now been incorporated into the UK Higher Education Academy) requires experienced staff to submit a reflective portfolio of their current teaching practice, covering a range of activities associated with the facilitation of student learning (ILTHE, 2004). For academic staff new to higher education teaching, it has become commonplace for universities to provide a major professional development opportunity in the form of postgraduate level programmes through an Educational Development Unit (see Gosling, 2001). Many of these programmes are accredited by the ILTHE and relate to academic practice in general, or to the facilitation of student learning in particular, and all that this encompasses. Many universities within the UK have made the completion of such accredited programmes a compulsory component of the probationary period for new academic staff and completion of such a course allows for automatic membership of ILTHE.

One of the very positive benefits of the rapid and significant changes in the nature of higher education has been the 'rediscovery' of the concept of the scholarship of teaching. While eminent academics such as Elton (1986; 1992; 2003) and Boyer (1987; 1990) have long argued that scholarship is at the heart of what the profession of higher education is all about, a criticism which can be justly directed

towards academic staff is that they too rarely apply the same kind of thought processes to their teaching as they apply to their disciplinary-based research.

However, the scholarship of teaching has now achieved a greater level of importance and it has become almost fashionable to quote in all papers and publications relating to academic practice or educational development Ernest Boyer's four categories of 'scholarship', namely the scholarship of *discovery, integration, application and teaching* (Boyer, 1990).

There is a danger though, that merely quoting the most popular definitions of scholarship passes for a deep understanding of the terms of reference and indeed signifies that a scholarly approach is now taken at all levels in the teaching and assessment of student learning. To avoid this potential pitfall, it is incumbent upon educational developers in particular to enable all staff engaged in supporting student learning to interrogate the terms of reference of the 'scholarship of teaching' and to apply the principles to their disciplinary-based classroom practice.

It is clearly a positive step when academic staff discuss the scholarship of teaching, particularly in light of the contradictory moves to continue to polarize teaching and research through the imposition of a much greater level of performance-based funding for research activities in universities, thus putting pressure on young or new academic staff to focus on research rather than teaching. However, a second potential pitfall in the current emphasis on the 'scholarship of teaching', particularly if the term is not 'unpackaged' and understandings of what it means in different disciplinary contexts are not developed, is that teaching becomes (or perhaps remains) uncoupled from student learning and from the assessment of student learning (Stefani, 1998).

In essence, it would be a greater step forward to hear academics discussing the 'scholarship of teaching, learning and assessment', as this would give greater recognition to the fact that teaching and learning are complementary activities and that assessment is an integral component of teaching and of learning. It is known from research into student learning that assessment essentially drives the curriculum (Ramsden, 1992; 2003; Biggs, 1999) and furthermore that assessment in itself should constitute an episode of learning (Stefani, 1998). Therefore it is fundamental to effective teaching that assessment is seen as an integral part of the teaching and learning contract, and thus crucial that a scholarly approach is applied equally to the facilitation of student learning and to the assessment of student learning.

In the remainder of this paper I will attempt to present a potential model for promoting among academic staff the underlying principles of a 'scholarly approach to assessment', while at the same time showing the linkage between teaching, learning and assessment.

The learning forum

As mentioned above, in UK universities, and indeed many universities in Australasia, it is commonplace for new staff to be encouraged to pursue a professional development programme relating to teaching and learning in higher education. Having gained knowledge of many of the UK programmes through my role as an accreditor for the ILTHE and through being invited on to Quality Assurance Committees in many higher education institutions, I have noticed that assessment is often treated as a separate entity from teaching and learning. This can, in my opinion, reinforce a notion that assessment is something which is bolted on after course content has been decided upon. This may be because the vast body of research literature now available on teaching, learning and assessment cannot reasonably be packed into a few workshop sessions or even within longer modules of postgraduate professional development programmes. Frequently there is also a tension between supporting new staff in the 'how to' of teaching in higher education (given the difficult circumstances they face) and potentially imposing a body of research and literature which many staff may consider to be somewhat divorced from their own subject area.

However, I believe that if we are to succeed in the quest to encourage, support and enable staff to take a scholarly approach to teaching, learning and assessment, even in the difficult circumstances which we currently face, then educational developers must themselves take a scholarly approach in this enabling task. This idea is affirmed by researchers such as Lueddeke (1997) and Middleton (1998). Much of the work I describe here stems from my experience of co-facilitating a module within the University of Strathclyde postgraduate accredited programme entitled Advanced Academic Studies; this module, 'Learning, Teaching and Assessment in Higher Education', accounted for 50% of the overall programme. The intention of the module was to encourage participants to compare their own ideas about their day-to-day practice with those in theoretically informed literature on learning, teaching and assessment in higher education (Soden & Stefani 2001). Participants were expected to attend four full day workshops over a period of twelve weeks and, to gain credit for the module, had to present an assignment of 4,000-5,000 words, reflecting on their learning and showing ways in which they had applied that learning to

their disciplinary-based practice. Between workshops, participants engaged in peer critique of each other's practice, which could include both their teaching and assessment strategies.

A logical model of curriculum development

To initially engage module participants in a discussion on curriculum design, development and delivery, I have found that engaging them in an interrogation of Cowan & Harding's 'Logical Model of Curriculum Development' (1986) is helpful. This shows staff that assessment is an integral part of course design, not something which is 'bolted on' at some convenient chronological moment — convenient that is, only to the staff member doing the bolting (see Figure 1).

In the original model of curriculum development presented by Cowan & Harding (1986), the educational terminology in vogue was that a course had aims and objectives. In compliance with current trends, it is now more common to talk about learning outcomes, so in the modification of the original model which I use, I have changed aims for learning outcomes. Using the model therefore also allows for dialogue on the setting of learning outcomes for any course or programme.

A further modification which has been made to the original model is to ask not only about the 'How' and the 'What' of the assessment strategy and the teaching and learning strategies, but also to ask the question 'Why?' I use this with participants in workshop settings to emphasize the need to interrogate our classroom practice and to reflect on our actions (see Figure 2).

This model allows for an interrogation of 'how' to assess and 'what' to assess. I have always interpreted 'how to assess' to mean 'how can we assess for the stated learning outcomes?' Asking this question allows an opportunity to encourage module participants to pause and reflect on the learning outcomes for any course in which they have a teaching role, and to think through the assessment devices they use. Interactive discussion will invariably result in a list which includes essays, reports, case studies, laboratory-based practical or creative work, individual and group projects, written tests and examinations, etc. This list is not exhaustive, but drawing it out does allow for some discussion on the appropriateness of assessment strategies and the contexts in which different assessment methods might be used, and for references to be made to case study material and other resources which might be useful.

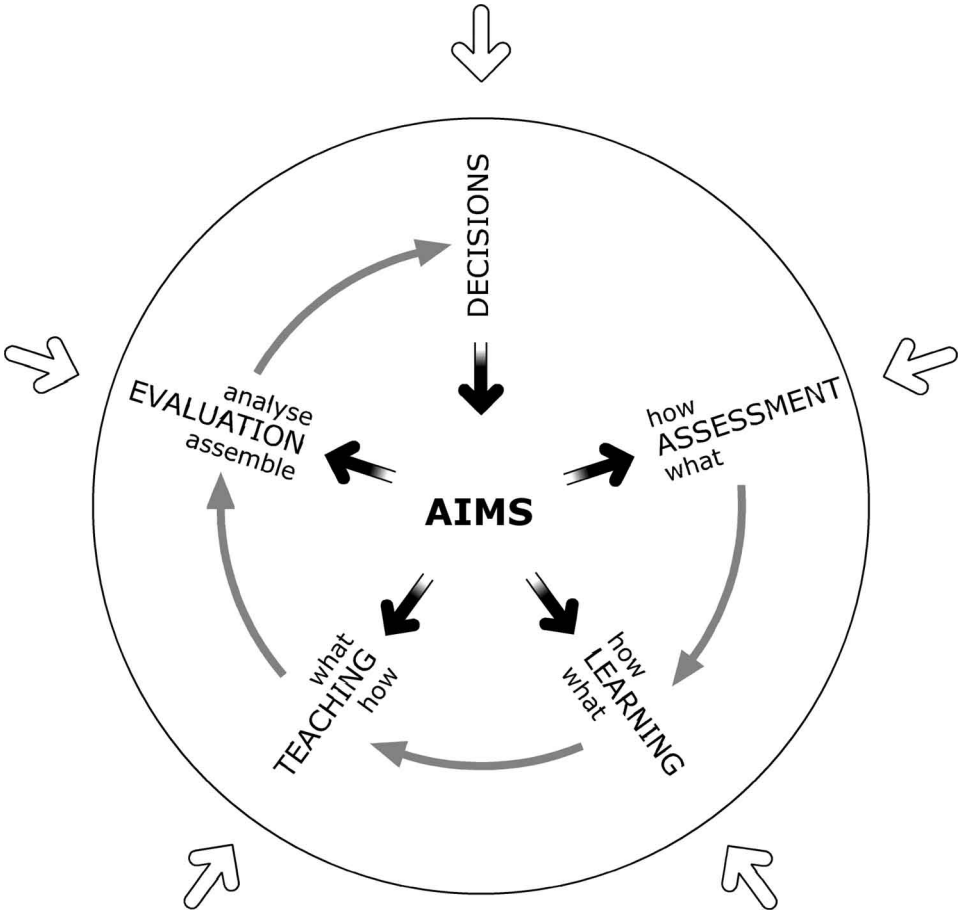


Figure 1: A Logical Model of Curriculum Development (from Cowan & Harding, 1986). Diagram used for the purpose of introducing academic staff to the concept of using a model to inform the design, development and delivery of the curriculum

The more interesting question which the model encourages is 'what am I assessing?' or 'what knowledge, skills, understanding, creativity, attitude and aptitudes, etc., will we assess?'

To open up assessment in this way allows for deeper level dialogue with staff relating to the goals of higher education. How can we assess student learning fairly, objectively and rationally without developing our own conceptions of the learning outcomes or the overarching goals of higher education? While to enter into a dialogue on the goals of higher education may appear to detract from discussion on teaching, learning and assessment, it is important on the one hand to recognize that the key domain of academic staff is within their disciplinary base

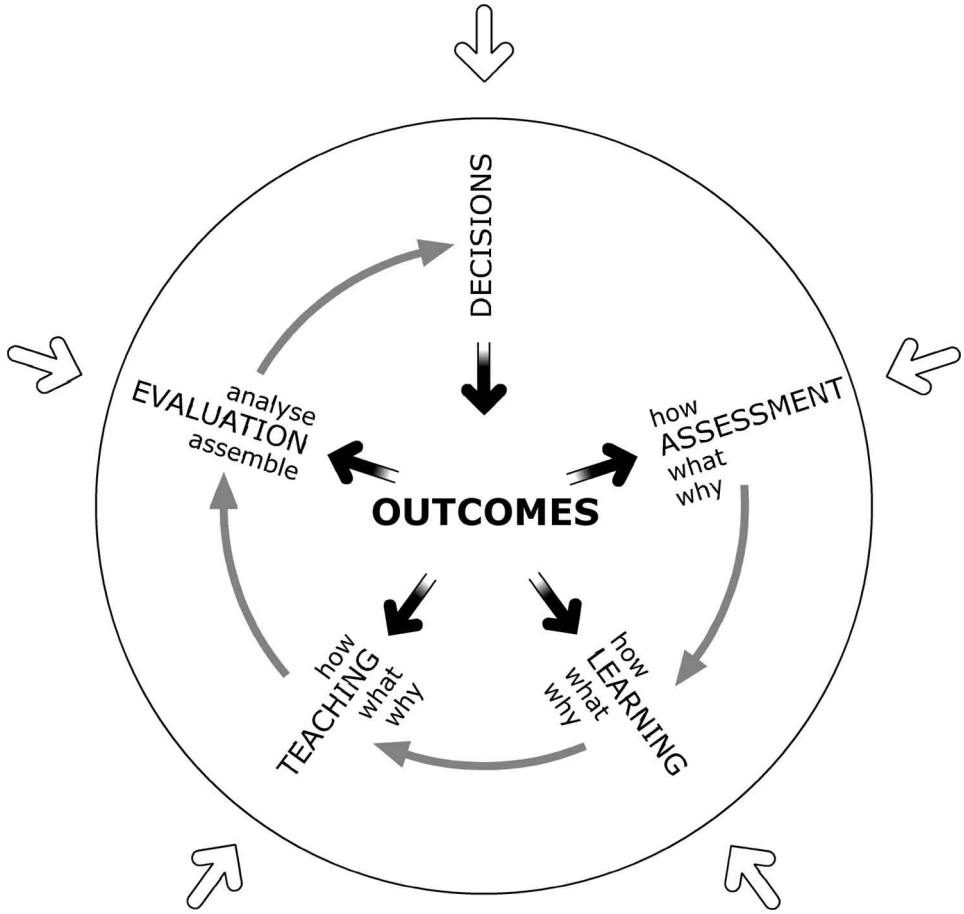


Figure 2: A simplified version of Cowan & Harding's Logical Model of Curriculum Development (1986). The model places learning outcomes at the centre and includes 'why' in addition to 'what' and 'how', relating to assessment, learning and teaching.

(Becher, 1989; Becher & Trowler, 2001), but on the other hand, that the changing needs of the graduate employment market (Harvey *et al.*, 1997; Stefani, 2004) require university staff to take a broader view of the goals of higher education. We must recognize that in an era of mass higher education, pursuing a university degree programme is, for most students, merely a step (albeit a major one) on the pathway of lifelong learning (Longworth, 1999). Thus, while we don't want to detract from the integrity of the disciplinary domain, we must take into account that we are preparing students for a dynamic employment market which requires graduates to be flexible, adaptable and able to take responsibility for their own learning and development.

What are the goals of a 'higher' education?

While the contemporary higher and further education curriculum is a highly contested arena (Soden & Stefani, 2001), there seems to be a general consensus that it should enable students to think for themselves or to become 'critical thinkers' (HEQC, 1995; 1996). In working with academic staff new to teaching in higher education, it is fair to say that many of them begin their career with a staff-centred view of the curriculum, considering only the needs within the discipline, which often do not go much beyond course content and 'traditional' modes of assessment.

In broad terms, one of the key goals of higher education can be expressed as enabling students to become autonomous, independent learners (Boud, 1995). However, we then have to address the question of what characterizes the autonomous, independent learner? Prior to the recent emphasis on 'the scholarship of teaching', there was already a growing focus on the development of the 'reflective practitioner' in higher education. This terminology came in to vogue with the publication of Donald Schön's seminal work entitled 'The Reflective Practitioner' (1991) and was a call for 'education for reflective practice' which, as Schön argued, is crucial to the professions and to professionalism. The concept of reflection has now become a more fundamental aspect of teaching and learning in higher education and has had an impact on the 'what' and the 'how' of assessment of student learning.

Promoting the goals of a 'higher' education at disciplinary level

Working with academic staff accustomed to thinking in their disciplinary language (Becher, 1989; Becher & Trowler, 2001), 'reflection' as a concept must be unpackaged into different terms, terms which may be seen to have meanings compatible with the knowledge understanding, skills and attitudes which are key to different disciplinary domains.

In working with staff through such issues, I have found it to be important to encourage them to take ownership of such concepts as 'critical reflection' and 'critical thinking' and to support them in translating or transforming these concepts to give them a meaning in disciplinary terms. It is after all a tall order to shift one's thinking towards teaching and assessing for the development of these higher

order thinking skills, when essentially this is at odds with a 'traditional' content-driven curriculum and assessment strategies which are primarily based on knowledge content (Fisher, 2003).

When first exposed to a discussion of the goals of higher education and to the ways assessment strategies can either enable students to develop higher order thinking skills or can in fact impede the development of such skills, many staff members in workshop settings give some indication that they feel uncomfortable with such notions and in fact may want to reject the ideas altogether. In this situation, I work back to first principles and revisit the five different understandings of learning discerned by Säljö's (1979) interviews with adult learners, which were expressed as follows:

1. Learning as a quantitative increase in knowledge, i.e. acquiring information or knowing a lot
2. Learning as memorization, i.e. storing information that can be reproduced
3. Learning as acquiring facts, skills and methods that can be retained and used as necessary
4. Learning as making sense or abstracting meaning. Learning involves relating parts of the subject matter to each other and to the real world
5. Learning as interpreting and understanding reality in a different way. Learning involves comprehending the world by re-interpreting knowledge, or *transforming* knowledge.

(reproduced in Ramsden, 1992)

This almost always achieves its intention, allowing a deeper discussion relating to what the module participants hope to achieve through their teaching and it allows us to ask the question 'do we necessarily facilitate student learning and then assess that learning in a way which recognizes transformation of knowledge?'

Assessing for the goals of higher education

While it is easy enough to assert that learning is an active process and that at the level of higher education it should result in a process of transformation of knowledge, there is also ample indication that curriculum design, as it is currently conceived, does not necessarily support the development of critical thinking and reflection. For

example, the work of Stephen Brookfield (1995) suggests that critical reflection requires an environment where the self-worth of the learner is respected, where the curriculum is built around the needs and aspirations of learners and where learners are willing to have their own views challenged and can feel safe to challenge others.

While it may be possible to create this type of environment in small group teaching, we need to question the complexities of achieving this in the context of mass higher education and the concomitant large classes which academic staff must deal with. It is important to work with staff in a manner which recognizes the constraints they work with and to reassure them that achieving and assessing for the higher level goals of higher education is not a simple task, but if they themselves develop a greater understanding of what they are in fact trying to achieve, then the chances of success should be greater.

When I work with staff in participative workshops on assessment strategies, I am constantly moving back and forth between the model of curriculum development previously discussed (to remind staff where 'assessment' is positioned in this model) and discussions on research into teaching and learning (to affirm the linkages between these three processes).

Returning to the question of why we are not always assessing for higher order skills, the reviews of research into teaching and learning carried out by Kember (1997) suggest that curriculum design may influence lecturers to focus on subject matter in their teaching rather than on the development of critical thinking. Soden & Stefani (2001) suggest that this is because curriculum content is usually specified far more fully than key skills, attitudes and attributes. It is also the case, of course, that lecturers are offered little help in clarifying what is encompassed in the notion of good 'thinking'. Thus they are not clear what it is they are supposed to be helping students to develop. It should not be a surprise therefore that the lack of clarity about the nature of thinking leads to the confusion about how good thinking and critical reflection might be assessed.

While weaving in a theoretical backdrop to the scholarship of assessment, when working with staff to encourage them to consider the structure of assignments which they set for the students, I often quote my experience of being invited into a department to discuss with a senior academic 'the appalling set of essays' he has just received from a class of postgraduate students. My role as an educational developer involved acting as a consultant for staff to support them in dealing with teaching and learning issues. In response to the question, I suggested that this academic provide

me with the set of essay questions which had been presented to the students and that I organize some workshops with the students to explore the gap between the academic expectations within the course and the student achievement.

In facilitating a series of workshops with the student group, I was much less interested in the actual subject matter and course content than in the students' conceptions of what was expected of them.

During these workshops, I discovered that the problem lay primarily with the 'stem' of the essay questions. While the academic staff involved were indeed trying to promote 'critical thinking' by setting assignments in the form of essays which asked students to 'compare and contrast', to 'interpret', to 'critically evaluate', etc., it turned out that the students were very poor at defining these terms. Their tendency was to 'describe' or to merely 'present' what they considered to be the appropriate content relevant to the topic of the essay. After further discussions with the irate staff member, it turned out that while academic staff wanted to promote critical thinking and to assess for critical thinking and reflection through the device of essays, the methods of teaching or facilitating learning did not in any way promote such critical thinking in the classroom. The teaching methods were primarily transmission-based and content-driven with few opportunities for students to engage in critical inquiry.

This cameo which I share with the module participants is one of many in a similar vein which I have experienced while working as an educational developer. It highlights and affirms that academic staff themselves need to critically analyse the assignments they set for their students and ensure that there is a shared understanding between staff and students of what is expected within the teaching and learning contract (Stefani & Nicol, 1997). In other words, enhancing students' capacity for critical reflection requires us to offer clear guidance about what is required for critical reflection, give feedback on how reflective capacities can be improved and model critical reflection throughout the courses we offer and present (Brookfield, 1995; Harvey & Knight, 1996).

After sharing such anecdotes, I invite staff within workshop sessions to reflect on 'critical thinking processes' relevant to their discipline and how they might build such processes into their courses, model them for students and also assess for them. From the discussions and feedback I receive from staff, the sharing of anecdotes helps them to understand more clearly what I mean by asking them to take ownership of the terminology of 'reflection' and 'critical thinking' in their own disciplinary-based terms.

Assessing for critical thinking and critical reflection should be important whatever the learning level in any university course, but it must be clear to the students what is expected of them and every effort must be made to ensure that students understand the assessment criteria in the way that staff intended them to be understood (Orsmond *et al.*, 1996; 1997).

There is a natural lead in here to discussions on how academic staff set assessment criteria and whether they consciously, explicitly, link assessment criteria to the intended learning outcomes for the courses being offered. There is research to suggest that lecturers often assess students on easily assessable matters, such as memorization of large bodies of factual material (Boud, 1995). In other words, staff sometimes actually reward a surface approach to learning whereas the focus should be on how students use, interpret or criticize material to do something with it, taking a deep approach to learning (Marton & Säljö, 1976; Ramsden, 1992; Entwistle, 1995).

The language of assessment

Sharing the underlying principles of teaching, learning and assessment within workshop settings takes up a lot of time. It is a facet of the Advanced Academic Studies programme, upon which the work described is based, that it must be interactive and must allow for the sharing of experiences amongst the participants. The intention of this is to model peer learning and to encourage staff to promote peer learning within their own classes.

It is also the case that the programme is based upon the concept of 'situated learning', whereby the staff are being presented with new ideas about teaching, learning and assessment which are theoretically underpinned, but they are not being bombarded with ideas on 'how to teach in your classes'. Rather they are being encouraged to reflect on the class discussions, the recommended readings which are a standard aspect of any accredited programme and their classroom actions, to try to reinterpret their practice in the light of the insights they have gained (Soden & Stefani, 2001).

In addition to promoting a scholarly approach to teaching, learning and assessment through an interrogation of the goals of higher education, it is of course important to ensure that the participants understand the basic language of assessment. Terms such as *summative* and *formative* are clearly defined with ample scope for discussion on the importance of formative feedback to affirm student learning. There is always a considerable level of discussion relating to the difference between *criterion-referenced assessment* and *norm-*

referenced assessment. There has, for example, always been ample anecdotal evidence that academic staff have a 'sense' of what they are looking for in student assignments, but do not necessarily fully articulate the criteria. Having this 'sense' of what they are looking for in an assignment has led to a preponderance of norm-referenced assessment, comparing one student's performance with another or, put another way, assignments are put into a rank order (Biggs, 1999). This is an ineffective means of assessing student learning and affirms the notion of 'the hidden curriculum' (Miller *et al*., 1998).

Many subject areas favour the use of essays as a means of assessing student learning. After having the discussion relating to 'what exactly are we assessing?' as described above (i.e. are we assessing content knowledge or are we assessing critical thinking?), it is highly productive to ask course participants to reflect on how they assess essays. Do they have clear criteria? Do they share these criteria with their students? Are they sure that their students understand the criteria in they way they meant these criteria to be understood? This exercise is intended to encourage reflection on how assessment is carried out and how fairly staff are treating their students.

The key point to convey is that the assessment of student learning entails making a judgement as to how well a student's performance matches the intended learning outcomes for a given course. This in turn requires the setting of criteria relating to these learning outcomes. The task for staff is to apply the scholarship of assessment within their own disciplinary contexts, and to work towards assessing for higher order thinking skills.

The examples of discussions, exercises and reflections which I have presented are not exhaustive. Indeed, within the module it is not possible to cover every facet of assessment, but the hope is that participants take more of an interest in the pedagogical underpinnings of their practice and will follow up on the module by engaging in further professional development opportunities available to them. To end the module, I generally present the following simplistic guide to the assessment process. I present it to staff as an aid to their practice which can be used whatever the format of the assessment and whatever the medium being used.

The key processes associated with assessment are:

- setting the criteria for assessing student learning in accordance with the learning outcomes
- ensuring a shared understanding between staff and students of the assessment criteria

- selecting the evidence that would be relevant to judge against these set criteria
- ensuring students understand the nature of evidence to be provided
- making a judgement about the extent to which the assessment criteria have been met
- ensuring transparency of these judgements
- communicating assessment outcomes to students
- providing useful feedback to the students on the assessment outcomes.

Summary

Clearly there is no definitive way to introduce or raise awareness among academic staff to the scholarship of teaching, learning and assessment. The intention of this paper is to give some insights into how staff can be supported in developing an understanding of the relationship between student learning and assessment of student learning. While I choose to use the model of curriculum development presented by Cowan & Harding (1986), others may choose to use the model of alignment of learning, teaching and assessment proposed by Biggs (1999). On the other hand, educational developers working with disciplinary-based staff may choose to use neither of these models, but rather to use a case study approach, for example on facilitating understanding of learning and assessment.

While this paper does not touch upon many other aspects of assessment, the modular structure of the Advanced Academic Studies programme allows for further inputs on assessment in other modules. The course assignments require participants to reflect on their current practice and to engage with other easily accessible literature on teaching, learning and assessment.

There is often insufficient time to go into particular types of assessment in great depth, for example peer- and self-assessment, group work assessment and online assessment. However, these modes of assessment are almost always brought up by participants in the course of discussion and there is always scope to provide extra resources for the participants on request. However, the ethos of the Advanced Academic Studies programme from the outset was to promote scholarship and a scholarly approach to the facilitation

of student learning. If the underlying principles of assessment are better understood, it is hoped that in the long term, academic staff will recognize that assessment is integrally linked to learning and that they will develop the skill of transforming their theoretical understandings into pragmatic classroom action.

References

Becher, T. (1989) *Academic Tribes and Territories: intellectual enquiry and the cultures of discipline*, Buckingham: Society for Research into Higher Education & Open University Press.

Becher, T. & Trowler, P.R. (2001) *Academic Tribes and Territories*, 2nd ed, Buckingham: Society for Research into Higher Education & Open University Press.

Biggs, J. (1999) *Teaching for Quality Learning at University*, Buckingham: Society for Research into Higher Education & Open University Press.

Boud, D. (1995) *Enhancing Learning Through Self-assessment*, London: Kogan Page.

Boyer, E.L. (1987) *College: the undergraduate experience in America*, New York: Harper & Row.

Boyer, E.L. (1990) *Scholarship Reconsidered: priorities of the professoriate*, Princeton: Carnegie Foundation for the Advancement of Teaching.

Breivik, P.A. (1998) *Student Learning in the Information Age*, American Council on Education: Onyx Press.

Brookfield, S. (1995) *Becoming a Critically Reflective Teacher*, San Francisco: Jossey Bass.

Cowan, J. & Harding, A. (1986) A logical model of curriculum development, *British Journal of Educational Technology*, vol. 2, no. 17, pp. 103-109.

Elton, L. (1986) Research and teaching: symbiosis or conflict, *Higher Education*, vol. 15, pp. 299-304.

Elton, L. (1992) Research, teaching and scholarship in an expanding higher education system, *Higher Education Quarterly,* vol. 46, pp. 252-268.

Elton, L. (2003) Some thoughts on scholarship, *Educational Developments*, vol. 4, no. 4, pp. 7-8.

Entwistle, N. (1995) The use of research on student learning in quality assessment, in G. Gibbs (ed) *Inspiring Student Learning through Assessment and Education*, Oxford: Oxford Centre for Staff Development.

Fisher, K. (2003) Demystifying critical reflection: defining criteria for assessment, *Higher Education Research and Development,* vol. 22, no. 3, pp. 313-336.

Gosling, D. (2001) Educational development units in the UK: what are they doing five years on? *International Journal of Academic Development*, vol. 6, no. i, pp. 74-90.

Harvey, L., Moon, J., Geall, V. & Bower, R. (1997) *Graduates Work: organisational change and students' attributes*, Birmingham: Centre for Research into Quality, University of Central England.

Harvey, L. & Knight, P. (1966) *Transforming Higher Education*, Buckingham; Society for Research into Higher Education & Open University Press.

Hattie, J. & Marsh, H. (2000) *The Relation between Research Productivity and Teaching Effectiveness: complimentary, antagonistic or independent constraints?* Australian Research Council.

Higher Education Quality Council (1995) *The Graduate Standards Programme: interim report*, London: HEQC.

Higher Education Quality Council Enhancement Group (1996) *What are Graduates? Clarifying the Attributes of 'Graduateness'*, London: HEQC.

Ilthe (2004) [online]. Available from http://www.ilt.ac.uk/ [accessed 9 January 2004].

Kember, D. (1977) A reconceptualisation of the research into university academics conceptions of teaching, *Learning and Instruction,* vol. 7, no. 3, pp. 255-275.

Longworth, N. (1999) *Making Lifelong Learning Work: learning cities for a learning century*, London: Kogan Page.

Lueddeke, G.R. (1997) Preparing academics for teaching in higher education: towards an institutional model of professional practice, *Reflections on Higher Education*, vol. 9, pp. 51-75.

Lueddeke, G.R. (2003) Professionalising teaching practice in higher education: a study of disciplinary variation and 'teaching scholarship', *Studies in Higher Education*, vol. 28, no. 2, pp. 213-228.

Marton, F. & Säljö, R. (1976) On qualitative differences in leaning: 1. outcome and process, *British Journal of Educational Psychology,* vol. 46, pp. 4-11.

Middleton, S. (1998) The Scholarship of Teaching [online]. Available from http://www.uoguelph.ca/guelph/98-02-27/sanchy.html [accessed 9th January 2004].

Miller, A.H., Imrie, B.W. & Cox, K. (1998) *Student Assessment in Higher Education: a handbook for assessing performance*, London: Kogan Page.

National Committee of Inquiry into Higher Education (1997) *In The Learning Society: report of the National Committee of Inquiry into Higher Education* (Dearing Report), London: HMSO.

Orsmond, P. Merry, S. & Reiling, R. (1996) The importance of marking criteria in the use of peer assessment, *Assessment and Evaluation in Higher Education,* vol. 2, no. 3, pp. 239-250.

Orsmond, P. Merry, S. & Reiling, R. (1997) A study in self-assessment: tutor and students' perceptions of performance criteria, *Assessment and Evaluation in Higher Education*, vol. 2, no. 4, pp. 357-369.

Ramsden, P. (1992) *Learning to Teach in Higher Education*, London: Routledge.

Ramsden, P. (2003) *Learning to Teach in Higher Education*, 2nd ed, London: Routledge.

Säljö, R. (1979) Learning in the learners' perspective. 1: some common sense conceptions, *Reports from the Institute of Education*, University of Gottenburg, no. 76 (quoted in Ramsden, P., 1992).

Schön, D.A. (1991) *The Reflective Practitioner: how professionals think in action*, London: ARENA.

Soden, R. & Stefani, L.A.J. (2001) *Developing Academic Practice*, Glasgow: University of Strathclyde [Internal Handbook].

Stefani, L.A.J. (1998) Assessment in partnership with learners, *Assessment and Evaluation in Higher Education,* vol. 23, no. 4, pp. 339-350.

Stefani, L.A.J. (2004) Encouraging scholarship in higher education teaching and learning, in P. Ashwin (ed) *Changing Higher Education: the development of teaching and learning*, London: Routledge.

Stefani, L.A.J. & Matthew, R. (2002) The difficulties of defining development: a case study, *International Journal of Academic Development*, vol. 7, no. 1, pp. 41-50.

Stefani, L.A.J. & Nicol, D.J. (1997) From Teacher to Facilitator of Collaborative Enquiry, in S. Armstrong, S. Brown & G. Thompson (eds) *Facing up to Radical Changes in Universities and Colleges*, London: Kogan Page.

Unfulfilled Promise: formative assessment using computer-aided assessment

MARTIN JENKINS
University of Gloucestershire, UK

Introduction

> 'Assessment is one of the most powerful drivers of innovation and change in education, as it defines the goals for both learners and teachers.'
> (DfES, 2003, p.32)

Assessment is recognized as probably the most influential factor related to how students learn (Brown & Glasner, 1999), directly linked to effective teaching and learning by rewarding understanding and achievement. Garrison & Anderson state:

> 'Successful learners most often rely on assessment deadlines and activities to both pace and direct their learning efforts. Effective teachers use assessment activities strategically to motivate learners to engage successfully in productive learning activities.' (2003, p.95)

The strategic use of assessment as an extrinsic motivator for learning is closely linked with assessment for summative purposes; any assessment forms part of a formative–summative continuum (Brown & Glasner, 1999). At the formative end of the continuum, assessment is regular and provides advice and guidance on progress. At the other end of the continuum is 'end of learning event' assessment. However, assessment can be both formative and summative though it has been argued that over-reliance on assessment as an extrinsic motivator can lead to surface learning approaches (Bull *et al*., 2002).

Assessment can be used for a variety of reasons, including:

- providing feedback
- grading

- enabling students to correct errors
- motivating students
- consolidating learning
- applying abstract ideas to practical examples
- estimating students' potential
- guiding selection or option choice
- giving staff feedback on how effective their teaching has been
- providing statistics for internal and external uses.

(Brown & Glasner, 1999)

Whilst it is possible to identify varied reasons for assessment, the range of methods deployed remains limited. Brown & Glasner (1999, p.8) claim that 'around 80% of assessments the world over are in the form of exams, essays and reports.' Trehan & Reynolds observe that while:

> 'examples of critical pedagogies, including those situated online are accumulating, they seldom exhibit corresponding changes in assessment practices'.
> (2002, p.280)

A similar observation was made by the author during a recent Association for Learning Technology (ALT) study tour in the Netherlands, looking at innovative uses of e-learning (Calverley *et al*., 2003). While innovative examples of e-learning were prevalent, assessment practices had changed little.

Providing feedback to students is a valuable part of the learning process and should be ongoing, frequent and comprehensive (Garrison & Anderson, 2003). The traditional practice of formative assessment is text-based feedback, providing advice and guidance on how to improve. But anecdotal evidence suggests that students do not fully exploit this feedback. Elton & Johnston (2002) indicate in their review that students do not take it seriously and are only interested in 'what counts'.

This paper focuses on the use of formative assessment through the deployment of Information and Communication Technology (ICT). Examples are presented and consideration is given to issues that need to be addressed to help fulfil the potential of formative assessment using ICT.

Computer-aided assessment

Computer-aided assessment (CAA) has been increasing within UK higher education, in part driven by the changing higher education environment (increased student numbers, lower unit of resource, modularisation, increased flexibility, etc.). CAA can be used in a wide range of contexts; Bull *et al*. (2002) summarize these as:

- *diagnostic* — ascertaining students' skills levels prior to learning events, i.e. audits
- *formative* — carried out during a learning event
- *summative* — carried out at the end of the learning event.

The common perception of CAA is that of multiple choice questions (MCQs); this approach undoubtedly exploits some of the strengths of using computers, for example providing consistent delivery, immediate grading and feedback and (once the questions are produced) saving time for academic staff. The range of MCQs, or objective question types, has increased markedly. Commercial systems such as Questionmark provide a large number of question types, allowing the integration of images and video, and hence moving beyond the MCQ. Focusing solely on MCQs limits the possibilities of how ICT can be used for formative assessment.

Web technologies provide huge potential for promoting more imaginative applications of CAA. They enable the use of different assessment methods within a range of approaches, including peer-assessment, self-assessment, group-based assessment and objective testing. Assessment methods involving ICT include case studies, mock exams, group projects and the creation of authentic learning tasks (Brown *et al*., 1999; Peat & Franklin, 2002; Herrington *et al*., 2002). CAA can encourage collaborative and reflective styles of learning. It is also possible for CAA to be adaptive, in that the outcomes of an assessment can be used to determine further questions or information that the student needs to address. Thus it is possible to guide students through a programme based on their responses at key stages, in a way that is appropriate for their specific learning requirements.

Charman (1999) identifies the following advantages of using CAA for formative assessment:

- repeatability
- immediacy of response to the student — providing a close connection between the activity and the feedback

- immediacy of marks to staff for monitoring and adaptation
- reliability and equitability
- increasing the diversity of assessment
- markers are not influenced by presentation
- timeliness — potential for assessments to be used at the most appropriate time, for example weekly tests
- flexibility of access, especially using the web
- student interest and motivation
- student-centred skills and learning — open access can encourage students to take responsibility for their own learning.

Such advantages do not apply uniformly to all forms and applications of CAA. Potential disadvantages or barriers associated with CAA must also be recognized, such as development time, potential risks (hardware, software and administrative) and the necessity for students to possess appropriate ICT skills and experience.

The changing environment

The higher education environment is continually changing: higher education institutions in many parts of the world are now addressing issues of modularisation, high student numbers and greater diversity in the student population, and the demands faced by many students who combine their studies with employment. Consequently, conventional forms of face-to-face teaching, learning and assessment are unlikely to meet student needs. (A project at the University of Gloucestershire investigating students' use and perception of flexible delivery methods, including online delivery, revealed that students defined as being 'on campus' were in fact travelling considerable distances — in some cases over 100 miles — to attend the University.)

Peat & Franklin (2002) have identified that, coupled with the need for more support, students have higher expectations for quality formative feedback. This is particularly significant as the demands on students make it increasingly difficult for them to attend on-campus sessions in the 'traditional' manner. Pressures on students' time and changing expectations mean they are demanding more immediate feedback. Alternative means of formative assessment (compared with traditional 'text-based' comments on assignments) therefore need to be considered.

Whilst ICT in learning and teaching has become widespread in the UK, it has yet to make a significant impact on changing learning and teaching practices. In an international survey of ICT use, Collis & Wende conclude:

> 'ICT use … has become common place but in a way that only gradually is stretching traditional on campus [teaching and learning] practices.'
> (2002, p.7)

The UK Government defines e-learning very broadly as 'learning in a way that uses information and communication technologies' (DfES, 2003, p.4). Higher education teachers will be familiar with students using ICT to research assignments using the Internet or online databases, to collect information delivered through their virtual learning environment (VLE) courses, or to engage in online discussions. Yet in the vast majority of cases, students continue to be presented with assessments consisting of essays, exams and reports. This is one reason why the UK Government, in its e-learning consultation document, emphasizes the need to re-align assessment with how students learn (DfES, 2003), and to develop means of formative assessment which provide individualized feedback and recognise the e-oriented skills students acquire through e-learning.

Using CAA for formative assessment: examples and discussion

Table 1 illustrates ways of employing CAA, which are expanded upon in the following sections.

Computer-based formative assessment

Multiple choice questions (MCQs)

Brown *et al*. (1999, p.11) argue that CAA allows students to monitor progress at regular intervals, thus addressing one of the key elements of formative assessment. According to Peat & Franklin:

> 'Formative computer based assessment can produce improvement in student learning outcomes and this can lead to positive attitudes to learning.'
> (Peat & Franklin, 2002, p.516)

Table 1: Forms of computer-aided formative assessment (after O'Reilly, 2001)

Approach	Method	Example Uses
Computer-based assessment (objective testing)	MCQs	Audit of knowledge and skills, with results and feedback immediately available to students and staff Use in class, allowing groups to discuss responses and get immediate feedback To provide a basis for discussion, useful for technical points such as Law
	Case studies	Allows testing of knowledge against authentic task
	Online portfolios	Students maintain their own web pages detailing their development. Also useful for providing feedback to staff
Group-based assessment	Group discussion for assignment preparation	Feedback on work in progress Weblogs
	Case studies	
Peer-assessment	Weblogs	Feedback on reflective journals
	Peer-assessment	Students provide feedback on other students' assignments
Self-assessment	Online mock exams	Allows students to make a comparison against 'exemplar' material
	Audits; MCQs	Allows students to monitor their own progress

Zakrzewski & Bull (1999) demonstrated a significant grade point increase in final results for students who worked through formative tests.

MCQs can provide feedback in a range of contexts. At the University of Gloucestershire, MCQs are used in weekly tests as part of a first year Marketing module (a popular module with over 300 students). Providing formative feedback on this basis using traditional approaches would be prohibitive. The MCQ tests were introduced to provide students with regular feedback on their understanding of the key principles being introduced throughout the module. Eight

tests, each consisting of ten questions, are made available at weekly intervals during the module, delivered via the University's VLE — WebCT. Initial evidence suggests that students have responded positively to receiving feedback in this way. As an added incentive to completing the formative tests, the best five scores contribute to the summative assessment of the module.

The Genetics programme at Wageningen University and Research Centre in the Netherlands uses Questionmark Perception for formative assessment. Weekly tests, linked to lectures and required reading, are made available for formative feedback; these tests are repeated as part of the final examination and account for 10% of the marks. Staff have found that students welcome doing the formative tests and that it frees up staff time to focus on problems.

At the University of Gloucestershire, MCQs have also been used to encourage discussion and understanding, both in class and with online groups. As part of a university-wide study skills module, MCQs are presented to a whole class using a web-based assessment package. They provide a focus for discussion and promoting shared understanding, with students having to justify their choice of answer and if necessary putting it to a vote. In a Local Administration Law module, delivered online, students are presented with a number of MCQs, which are subsequently used as the basis for discussion on the course bulletin board. Students are again encouraged to explain their choices and argue their case with peers. In the first case, the use of MCQs provides the students with immediate feedback on their group decisions, located within their discussions and decision making. In the latter, the use of MCQs provides students with immediate feedback on their choices, allowing them to then engage in an online debate with their peers.

Case studies

The Wageningen University & Research Centre has also utilized MCQs within a Marketing module. Students are presented with a weekly case study on which they are asked a series of questions. This approach tests the students' application of their knowledge against authentic examples, rather than testing recall of key facts (Calverley *et al*., 2003). The use of MCQs to augment inquiry-based learning has also been successfully used by Honey & Marshall (2003) in nurse education. In a Pathophysiology course students were presented with case studies to help develop their critical thinking skills. Honey & Marshall found that the use of MCQs in this way encouraged active learning, and that students particularly valued the timeliness of the feedback.

Group-based assessment

Collaboration

Online communication facilities can promote peer support, the creation of authentic and applied tasks, and an environment for reflection and continuous engagement in iterative assessment tasks. Iterative and continuous involvement in discussion areas provides a potentially valuable means for students to develop ideas and understanding, and to receive formative feedback from peers and tutors within a safe environment.

A recent development in web-based technology that supports collaborative learning is weblogs (also known as 'blogs'). A blog is a web page containing a series of short frequently updated postings in chronological order, in effect providing a personal publishing tool. This allows individuals to post reflective messages to which peers can then respond, so providing feedback, questioning and so forth — in other words, it can be used as a group tool. Robinson (2003) reports examples where teachers provide information for students to submit assignments and where students provide feedback to each other using weblogs.

Elton & Johnston (2002, p.15) note that 'formative assessment thrives on students' openness, while summative discourages it.' Formative assessment requires students to put forward their initial ideas and understanding for 'critique' and sharing. Students need to be encouraged to make public their ideas and the development of their thinking. Online methods can actively encourage this approach, and at different levels. The use of MCQs provides a more closed environment for students to test their ideas. The use of online discussion areas or online journals exploits the collaborative potential of the web. Research shows that submissions to online discussion areas encourage more reflective contributions (Garrison & Anderson, 2003). Participation is a key component of e-learning — engaging in the sharing of ideas and experiences and recognizing the value of 'exposing' students to the sharing of 'written' ideas. If formative assessment is a continuous process of feedback and support, it does not only have to come from the tutor. Participation can be encouraged through the use of well-designed, authentic tasks; indeed, Garrison & Anderson (2003) believe that education should be a collaborative constructive experience, where understanding is developed in a critical community of inquiry.

Peer-assessment

Involving the student in peer-assessment can encourage motivation, both through students looking at peers' work and knowing their own work will be peer reviewed; this can be a valuable experience, allowing students to measure themselves against others and compare their own assessments with those of their peers. A peer's comments may also have more impact than a tutor's; the author's personal experience suggests that peers can be more critical. Brew (1999) makes a distinction between peer marking and peer feedback, indicating that peer feedback can help focus learning and be positive for a group of learners, whilst peer marking can be disruptive. It is obviously important that the process is managed to ensure consistency; this means ensuring clear criteria and encouraging the students to develop the skills to evaluate and provide feedback.

In an online environment, bulletin boards are used as a medium for the iterative exchange of work to support peer-assessment, providing a forum for social engagement among students and the opportunity for students to test out and explore ideas: 'the ebb and flow of constructive criticism is important for learning' (Robinson, 1999, p.95).

The potential for using online bulletin boards as a forum for personal reflection, combined with collaboration, creates an ideal environment for personal development. Examples of this include the MEd in Networked Collaborative Learning run by the University of Sheffield. Here participants working in small groups engage in individual projects and activities, within a larger collaborative group which provides support and guidance. This culminates in a process of peer-assessment and review on the individual projects, prior to final submission (McConnell, 2000).

At the University of Gloucestershire, the module 'Collaborating with Communities' practises community development in its delivery[1]. A group online assignment forms part of the assessment for the module and students work in small groups on a negotiated assignment to produce a short report. This requires the students to share ideas, resources and information. Whilst the final report forms part of the overall summative assessment, the process provides a forum for students to collaborate and give each other feedback on their contributions to the completed project.

Self-assessment

Audits

At the University of Gloucestershire, online 'quizzes' are used to assess students' subject knowledge as part of a Primary Science module[2]. A range of tests assessing understanding and knowledge is made available to students as they progress through the module. Students can access the tests whenever they need and they are able to repeat them, obtaining feedback on all their attempts. The rationale behind this approach is to encourage students' confidence and interest in the ideas they will ultimately be teaching in school. In addition, as trainee teachers, they have to provide a portfolio of evidence, and the tests provide them with evidence which can be included. Comments from students highlighted that they welcome feedback in a more flexible way, being able to monitor their own progress as and when they wish. The formative tests also provide grounding for a summative assessment: at the end of the module formative tests are replaced by a summative 'exit' audit, which is completed under exam conditions. This approach provides the students with ongoing feedback whilst also preparing them for summative assessment.

Mock exams

Peat & Franklin (2002) employed a range of approaches in a Biology course, including weekly quizzes, a mock exam (which students could self-mark against web-based information), and the development of self-assessment modules. They believe this approach has led to 'significant benefits for staff and students', allowing more time for face-to-face interaction and providing students with more opportunities to gain valuable feedback at a time to suit them.

Institutional issues

The introduction of online learning requires a significant culture change and investment in staff development over and above the investment in technology. Many academic staff have not experienced any form of online and/or collaborative learning and may be inclined to perpetuate teaching and learning in a form with which they are familiar; likewise, when they were students they probably experienced a limited range of assessment methods (Alexander *et al*., 2002). Two factors are required to facilitate assessment online: explicit valuing

of collaborative learning in communities; and cultural change in academic communities (Angelo, 1999; O'Reilly, 2001). The utilization of CAA, as with any form of e-learning, must be accompanied by investment in staff development.

The use of CAA does have potential problems: the risks associated with using technology; the assumption of computer skills in students; and the potential for plagiarism. For students there may be feelings of isolation due to lack of face-to-face contact. The impact of these concerns can be reduced, depending on how online assessment is implemented; if the full potential of online learning is used to encourage collaboration, reflection and formative assessment, the feelings of isolation are reduced and 'managed' out of the system. Plagiarism through the use of the Internet is now widely recognized as a serious area of concern in higher education. As a consequence, the Joint Information Systems Committee (JISC) set up the Plagiarism Advisory Service (http://online.northumbria.ac.uk/faculties/art/information_studies/Imri/Jiscpas/site/jiscpas.asp), which focuses on detecting plagiarism in traditional written assignments.

The impact of student profiling must also be considered by universities in the UK; this is recognized in the Department for Education & Skills (DfES) e-learning strategy and coincides with the development of Managed Learning Environments (MLEs). Indeed, as Ryan *et al*. state:

> 'the movement towards the integration of assessment and teaching applies across all forms of web-based teaching and learning … the 'electronic record of student activity', or profiling, is also significant … . The IMS project will lead to the establishment of common standards for such student profiles that will be linked to the student as they move from one piece of courseware to another.' (2000, p.139)

Conclusions

> 'Online assessment in the 21st Century has come to mean many things to a range of people in various educational contexts.' (Northcote, 2002, p.623)

The examples in this paper illustrate the range of methods CAA can bring to formative assessment, and show how they meet some of the advantages identified by Charman (1999). Whilst online assessment provides benefit to students, teachers and administrators, it also brings challenges and potential risks, thus it demands to be managed and invested in appropriately.

Does introducing online formative assessment improve outcomes, i.e. raise student marks? There is mixed research on this; Peat & Franklin (2002) show that it did not raise outcomes and Charman (1999) cites examples where evidence is inconclusive, yet Zakrzewski & Bull (1999) provide evidence that it does. In this paper I have attempted to provide examples showing that online formative assessment can bring benefits in terms of flexibility and immediacy of feedback. However, introducing online learning on its own will not produce results; Alexander & McKenzie (1998) and Laurillard (2002) emphasize that it is also necessary to change the assessment. Without this culture change and recognition that CAA is more than MCQ then there will be unfulfilled promise.

Notes

1. This case study is one of the Learning Environment and Pedagogy (LEAP) Case Studies on the LTSN Generic Centre web site and can be viewed in full at: http://www.ltsn.ac.uk/genericcentre/index.asp?docid=18375.

2. This case study is one of the Learning Environment and Pedagogy (LEAP) Case Studies on the LTSN Generic Centre web site and can be viewed in full at: http://www.ltsn.ac.uk/genericcentre/index.asp?docid=18375.

References

Alexander, S., Kandlbinder, P., Howson, E., Lukito, L., Francois A. & Housego, S. (2002) Simassessment: enhancing academics understanding of assessment through computer simulation, in A. Williamson, C. Gunn, A. Young & T. Clear (eds) *Winds of Change in the Sea of Learning: charting the course of digital education. Proceedings of the 19th ASCILITE Conference held at UNITEC, Auckland, New Zealand, 8-11th Dec 2002*.

Alexander, S. & McKenzie, J. (1998) *An Evaluation of Information Technology Projects for University Learning*, Canberra: Australian Government Publishing Service.

Angelo, T. (1999) Doing assessment as if learning matters most, *AAHE Bulletin*, May 1999 [online] [accessed April 2004]. Available from http://www.aahebulletin.com/public/archive/angelomay99.asp.

Brew, A. (1999) Towards autonomous assessment: using self-assessment and peer assessment, in S. Brown & A. Glasner (eds) *Assessment Matters in Higher Education: choosing and using diverse approaches*, Buckingham: Society for Research into Higher Education & Open University Press.

Brown, S. & Glasner, A. (eds) (1999) *Assessment Matters in Higher Education: choosing and using diverse approaches*, Buckingham: Society for Research into Higher Education & Open University Press.

Brown, S., Race, P. & Bull, J. (1999) *Computer-assisted Assessment of Students*, London: Kogan Page.

Bull, J., Conole, G., Davis, H.C., White, S., Danson, M. & Sclater, N. (2002) Rethinking assessment through learning technologies, in A. Williamson, C. Gunn, A. Young & T. Clear (eds) *Winds of Change in the Sea of Learning: charting the course of digital education. Proceedings of the 19th ASCILITE Conference held at UNITEC, Auckland, New Zealand, 8-11th Dec 2002*, pp. 75-86.

Calverley, G., Corley, L., Creanor, L., Littlejohn, A., Jenkins, M., Newland, B. & Roberts, G. (2003) *Making Connections: British perspectives on learning technology developments in Netherlands higher education*. Report of the ALT/ SURF Study tour to the Netherlands in April 2003.

Charman, D. (1999) Issues and impacts of using computer-based assessments (CBAs) for formative assessment, in S. Brown, P. Race & J. Bull (eds) *Computer-assisted Assessment of Students*, London: Kogan Page.

Collis, B. & Wende, M. van der (2002) *Models of Technology and Change in Higher Education: an international comparative survey on the current and future use of ICT in HE*, University of Twente: Centre for HE Policy Studies.

Dfes – Department for Education and Skills (2003) *Towards a Unified e-Learning Strategy: consultation document*, Nottingham: DfES Publications.

Elton, L. & Johnston, B. (2002) Assessment in Universities: a critical review of research [online] [accessed April 2004]. Available from http://www.ltsn.ac.uk/application.asp?app=resources.asp&process=full_record&section=generic&id=13.

Garrison, D.R. & Anderson, T. (2003) *E-learning in the 21st century: a framework for research and practice*, London: RoutledgeFalmer.

Herrington, J., Oliver, R. & Reeves, T. (2002) Patterns of engagement in authentic online learning environments, in A. Williamson, C. Gunn, A. Young & T. Clear (eds) *Winds of Change in the Sea of Learning: charting the course of digital education. Proceedings of the 19th ASCILITE Conference held at UNITEC, Auckland, New Zealand, 8-11th Dec 2002*, pp. 279-286

Honey, M. & Marshall, D. (2003) The impact of on-line multi-choice questions on undergraduate student nurses learning, in G. Crisp, D. Thiele, I. Scholten, S. Barker & J. Baron (eds) *Interact, Integrate, Impact. Proceedings of the 20th Annual Conference of the Australasian Society for Computers in Learning in Tertiary Education (ASCILITE), held at the University of Adelaide, Adelaide, Australia 7–10 December 2003*.

Laurillard, D. (2002) *Rethinking University Teaching: a conversational framework for the effective use of learning technologies*, 2nd ed, London: RoutledgeFalmer.

McConnell, D. (2000) *Implementing Computer Supported Cooperative Learning*, 2nd ed, London: Kogan Page.

Northcote, M. (2002) Online assessment: friend, foe or fix? *British Journal of Educational Technology*, vol. 33, no. 5, pp. 623-625.

O'Reilly, M. (2001) Improving student learning via online assessment, in C. Rust (ed) *Improving Student Learning using Learning Technology. Proceedings of the 9th International Symposium on Improving Student Learning, Heriot-Watt University, September 2001*, Oxford: Oxford Centre for Staff and Learning Development, pp. 269-280

Peat, M. & Franklin, S. (2002) Supporting student learning: the use of computer-based formative assessment modules, *British Journal of Educational Technology*, vol. 33, no. 5, pp. 515-523.

Phillips, R. & Lowe, K. (2003) Issues associated with the equivalence of traditional and online assessment, in G. Crisp, D. Thiele, I. Scholten, S. Barker & J. Baron (eds) *Interact, Integrate, Impact. Proceedings of the 20th Annual Conference of the Australasian Society for Computers in Learning in Tertiary Education (ASCILITE), held at the University of Adelaide, Adelaide, Australia 7–10 December 2003*.

Robinson, G. (2003) Blog in the limelight, *Times Educational Supplement Online*, 7th November 2003, p. 14.

Robinson, J.M. (1999) Computer-assisted peer review, in S. Brown, P. Race & J. Bull (eds) *Computer-assisted Assessment of Students*, London: Kogan Page.

Ryan, S., Scott, B., Freeman, H. & Patel, D. (2000) *The Virtual University: the Internet and resource-based learning*, London: Kogan Page.

Trehan, K. & Reynolds, M. (2002) Online collaborative assessment: power relations and 'critical learning', in C. Steeples & C. Jones (eds) *Networked Learning: perspectives and issues*, London: Springer-Verlag.

Zakrzewski, S. & Bull, J. (1999) The mass implementation and evaluation of computer-based assessments, *Assessment & Evaluation in Higher Education*, vol. 23, no. 2, pp. 141-152.

Assessment for Learning

Sally Brown

Independent Higher Education Consultant; Visiting Professor at the Robert Gordon University (Aberdeen) and at Buckinghamshire Chilterns University College, UK

The changing nature of assessment

Assessment is probably the most important thing we can do to help our students learn. We may not like it, but students can and do ignore our teaching; however, if they want to get a qualification, they have to participate in the assessment processes we design and implement. For that reason I believe it is worth thinking through, individually and collectively, what we currently do and exploring how we can do our best to ensure that our assessment practices help rather than hinder learning. In this paper I will explore these issues, play with a negative exercise about what we can do to hinder learning and conclude with some pointers towards integrating learning and assessment.

Internationally, assessment is changing as the nature of teaching and learning in post-compulsory education changes. The student population in many countries is becoming diverse, with increasing numbers of part-time students, mature students and students coming from non-traditional backgrounds, particularly in the UK, where there is a political imperative to widen participation to students from socio-economic groups who previously had little or no access to higher education. A diverse population of learners necessitates a change in practice in post-compulsory education, with less focus on didactic tutor-led approaches and more concentration on the learning outcomes that students can hope to achieve (Miller *et al.*, 1998; Rust, 2002).

Fit-for-purpose assessment

I have long argued that assessment needs to be 'fit-for-purpose'; that is, it should enable evaluation of the extent to which learners have learned and the extent to which they can demonstrate that learning (Brown & Smith, 1997). We need to consider not just *what* we are assessing and *how* we are doing it (particularly which methods and approaches), but also *why* — our rationale for assessing on

any particular occasion and in any context. Our different reasons (to motivate students, to encourage activity, to provide guidance and feedback for remediation, grading and selection) will impact on our choice of assessment instruments, which may include the wide diversity of under-used methods which are suitable in different contexts. Rather than continuing to over-use unseen time-constrained exams, essays and reports, for example, we can consider using portfolios, in-tray exercises, posters, annotated bibliographies, reflective commentaries, critical incident accounts, reviews, role-plays, case studies and many of the other available means of assessment that are widely used in higher education institutions in the UK and internationally (Brown & Knight, 1994).

We also need to think about the agents *who* undertake the task. For example, if we want to assess group work, using intra-peer group assessment seems sensible in order to access group process, whereas if we want to assess employability, involving placement supervisors or clients would give us a better understanding of how students engage in a working environment than a hastily scribbled post-hoc work placement report could do. In some cases only the tutor will do, but these occasions I believe are fewer than is often posited.

We also need to consider *when* is the best time to assess (not, I would suggest, all at the very end of a learning programme if we want students to have a chance to learn from early errors). Is it possible to give students a choice about when they are ready to be assessed? How far can we (or should we) allow multiple attempts at assessment over a period of time? Why is UK higher education so wedded to a model of assessment that means that an Honours degree can normally only be undertaken over three years? Why not much longer? Or shorter?

To ensure that assessment is part of the learning process, I would argue that it should be learner-centred assessment and should reflect a learner-centred curriculum. Assessment methods and approaches need to be focused on *evidence of achievement* rather than the ability to regurgitate information. Inevitably this means a lesser concentration on traditional written assessments, particularly time-constrained unseen exams, and a greater emphasis on assessment instruments that measure not just recall of facts, but also the students' abilities to use the material they have learned in live situations. To be *valid*, the assessment needs to focus as well on what is intended to be learned. If we want our students to demonstrate employability when they graduate, our assessments need to be designed to be *practice-orientated,* whether in terms of the practice

of being a researcher or applications to professional contexts such as being an artist, an accountant, a health practitioner or a quantity surveyor. Rather than assessing a learner's ability to write about good practice, an effective assessment strategy would seek to measure how the student can put into practice the learning achieved. The methods used need to be *authentic,* that is, assessing what they claim to assess, not just what is easy to assess.

Any assessment strategy needs to be *efficient* in terms of staff time, cost-effective for the organizations concerned and should ensure that learners find the tasks they are set manageable, relevant and developmental. We cannot simply expect our students or ourselves to just keep working harder and harder; where possible we must make best use of the available technologies to make assessment more efficient (Brown *et al*., 1994). The assessment tasks need to be *integral* to the learning process, rather than a subsequent bolt-on and, to ensure this, tutors should be able to concentrate equally strongly on giving feedback and on making evaluative decisions about performance. Timing of assessment is also a key issue, since the responses given to assessed work need to allow opportunities for amendment and remediation of errors.

To enable a synthesis of learning achieved, asking students to write a reflective account can be very helpful. This provides an opportunity for students to review their experiences of the programme of learning as a whole, describe how they have developed over the period of study, reflect upon the literature that has influenced and guided their practice, and indicate how they plan to develop their work and themselves into the future.

Current literature on assessment argues strongly that the process should be a *transparent* one, with criteria that are explicit and clear to all concerned (assessors, those being assessed and moderators reviewing the process) from the outset. Assessment can become *valid* when the assessors use evidence of achievement, clearly matched against the criteria (Brown & Glasner, 1999; Gibbs & Rowntree, 1999; Thorpe, 2000).

Any assessment strategy that aims to be *inclusive* should deploy a variety of methods for assessment (for example written assignments, presentations, reflective accounts and so on), so that the same students are not always disadvantaged. All participants need to be provided with *equivalent opportunities* to demonstrate their abilities and maximize their potential.

It is imperative to clarify tutors' and students' expectations at the time of giving the assessment brief to the students. This means that the assessment criteria need to be clear, explicit, framed in language that is meaningful to staff and students and available well in advance of the commencement of activities that will subsequently be assessed.

The programme of assessment chosen needs to be *reliable*, so that different assessors derive the same grade for similar work (inter-assessor reliability) and individual assessors mark reliably to a defined standard (intra-assessor reliability). This can only be assured when the criteria are clearly understood by all who undertake assessment.

Creative subjects like music and art often provide particular challenges when it comes to assessment. Where possible, it may be helpful to involve students in establishing or negotiating the criteria for assessment, so that they fully understand what is expected of them. The degree of subjectivity involved in evaluating artefacts and productions needs to be recognized and articulated, so that everyone concerned understands the rules of the game. Assessors need to be sure that where students are involved in assessed work of widely divergent types, they can be assured of the equivalence rather than the identicality of the assessment experience. For example, students involved in group activities (such as drama and dance productions, or the production of installations) will necessarily take different roles, so assessment criteria must be designed to ensure that all students have an equal chance of achieving high grades.

Feedback

This for me is the principal area in which we can influence the extent to which our assessment practices are developmental, rather than solely judgmental. If assessment is to be integral to learning, feedback must be at the heart of the process. Even though it is time-consuming, I would argue that significant energy must be devoted to helping students to understand not only where they have gone wrong, but also what they need to do to improve. They also need feedback when they have done well, to help them understand what is good about their work and how they can build on it and develop further. No one can pretend this is an easy task. Summative feedback, which enables judgements to be made for progression and completion, needs quite clearly and overtly to relate to the assessment criteria and to be strongly aligned to the curriculum objectives.

Formative feedback is crucial. It needs to be detailed, comprehensive, meaningful to the individual, fair, challenging and supportive, which is a tough task for busy academics. We must consider using the whole range of means available to us to make this possible, including computer-aided assessment and strategies for giving feedback efficiently such as assignment return sheets, assignment reports, in-class collective feedback and other means (Brown *et al.*, 1994).

We can also use self-assessment, peer-assessment and group assessment, none of which should be regarded as a 'quick fix', because they take considerable briefing, training and rehearsal if they are to be effective, but can, when properly managed, save some staff time and they are extremely valuable in helping students interpret criteria. As these methods also encourage students' metacognition (that is, a means of learning about their own learning), they are also very effective in encouraging deep rather than surface learning.

How to use assessment to prevent learning!

These tongue-in-cheek tips are designed to make you think about some of the behaviours that can actually get in the way of students' learning. They could be shown to students, who could be asked why the advice presented here is seriously misguided; such a discussion might help students better understand what is really expected of them in assessment.

- **Keep students in the dark about the rules of the game**. Brighter students will intuitively understand the criteria and naturally excel. You will thereby get a good range of marks, from the truly appalling to the really outstanding. This will make your external examiner happy.
- **Do all the assessment at the end of the learning programme**. You can't assess students until you have taught them everything, so all the assessment needs to take place at the very end of the process. If they then fail, it just shows that they weren't very good.
- **Make sure you know the identity of the student who has done each piece of work**. Then you can check whether the standard is what you would expect of that student. You can then correct marks if you think that a poor student has over-performed or a strong student has not done themselves justice. You can normally tell early on what kind of a degree a student will get and expectations are rarely proved wrong.

- **To be fair to all students, give each an identical test**. If they have problems with it because of so-called 'special needs', that's their problem.
- **For coursework assessments, stick firmly to your deadlines**, regardless of the plausible excuses students come up with. The real world works on deadlines. If you show any flexibility, students will just take advantage of you.
- **Don't be soft on any students who claim that they don't do well in exams**. Even if their coursework marks are good, remind them that they have to get their act together for exams, or else they shouldn't be in higher education in the first place.
- **Don't indicate how many marks go with each of the parts of your questions**. This just causes trouble when you mark the scripts, and stops you being fair to the students who have worked out what was important in their answers. The really good students will know where the marks lie.
- **Don't give students any written feedback**. They will only argue with you about the marks you have given them and ask you to justify how the comments link to the scores you have awarded. You can't be expected to do that. In fact it is probably safest just to give them the mark and never give them back the original work.
- **Always plan at least some questions on material that you haven't covered with the class**. This sorts out those students who read around the subject and those who don't. But don't actually tell them that this is what you're planning to do, or the strategic students may read more extensively.
- **Only look at student scripts once**. Mark them as soon as you see them and never attempt to revisit earlier scripts because you might be tempted to change marks once you have seen other students' work. Rely on your innate ability to make fair and accurate judgments.
- **When designing assessments, trust your first instincts**. Don't show your draft assignments to other staff, who may interfere by making comments about the type of questions, wording, balance and coverage of the material.

- **Stick to tried and tested methods like unseen exams**. Any fancy innovative methods will be just too much hard work for you and won't test what you really want to find out, i.e. whether students can remember the facts.
- **When you set coursework essays, don't set a word limit**. The good students will naturally have a lot to say, and will deserve good marks. You can usually tell at once from the length of an essay how much thought has gone into it.
- **Don't make your questions too straightforward**. You want to be able to see who can make sense of the questions, and give these students the higher marks. Students who can't make sense of a question are demonstrating their ignorance, and don't deserve high marks.
- **When you know that some particular bit is really important, hit it in several different questions in parallel**, so that there is no escape for those students who have not mastered that bit.
- **Don't get into discussions with a class about how they will be assessed**. Just remind them that they're here to learn, and you're here to teach them, and then they've got to prove to you what they've learned.
- **Don't be tempted to include self-assessment elements**. Students would simply give themselves over-high marks or grades, and would probably feel that it wasn't their job to assess their work.
- **Don't get students peer-assessing each other's work**. They would learn too much from each other's mistakes, and you want to be able to see who makes which mistakes for yourself.
- **If you design a question paper that really works well, use it year on year**. You can save yourself a lot of work by using the same questions again and again. There is no need to worry about students from previous years talking to the next cohort of students as they all tend to lead quite separate lives.

Some conclusions

The ways we assess our students can really make a difference to how students learn. There are multiple and complex problems to resolve and solutions are not easy to find (or the brightest minds in the world would have done so already), permanent (as we have to deal with an ever-changing environment), or universal (assessment is an area where context is of paramount importance; what works well in a medical environment probably doesn't work equally well in a poetry workshop, although there might be some interesting cross-overs). So we are left with the need for professional higher education practitioners to take the lead in ensuring that we do not allow the process to slip out of our hands. We cannot let bureaucratic regulations (whether from within our institutions or nationally) to skew our effective assessment processes. If we find our systems do not allow us to implement a really valuable assessment innovation, for example, then we must find ways to change the system. We need to ensure that decisions about assessment strategies are based on the best available evidence-based research on assessment, rather than on custom and practice or what is easy to do.

So we need to keep abreast of new developments, evaluate tried and tested ones and experiment with our own initiatives, preferably within a supportive learning community of fellow practitioners. Inevitably, I would therefore argue that we in the UK need to participate in our subject communities, through the Higher Education Academy (http://www.heacademy.ac.uk/).

Bibliography and suggestions for wider reading

Baume, D. (2001) *LTSN Generic Centre Assessment Series No.6: A Briefing on Assessment of Portfolios*, York: Learning & Teaching Support Network.

Boud, D. (1988) *Developing Student Autonomy in Learning*, London: Kogan Page.

Boud, D. (1994) Keynote speech at *SEDA Conference on Assessment,* Telford, May.

Bridges, P., Cooper, A., Evanson, P., Haines, C., Jenkins, D., Scurry, D., Woolf, H. & Yorke, M. (2002) Coursework marks high, examination marks low, *Assessment & Evaluation in Higher Education,* vol. 27, no. 1, pp. 36-48.

Brown, G. (2001) *LTSN Generic Centre Assessment Series No.3: Assessment: A Guide for Lecturers*, York: Learning & Teaching Support Network.

Brown, G., Bull, J. & Pendlebury, M. (1997) *Assessing Student Learning in Higher Education*, London: Routledge.

Brown, S. & Glasner, A. (1999) *Assessment Matters in Higher Education*, Buckingham: Open University Press.

Brown, S. & Knight, P. (1994) *Assessing Learners in Higher Education*, London: Kogan Page.

Brown, S. Race, P. & Bull, J. (1999) *Computer Assisted Learning in Higher Education*, London: Kogan Page.

Brown, S., Race, P. & Smith, B. (1996) *500 Tips on Assessment*, London: Kogan Page.

Brown, S. Rust, C. & Gibbs, G. (1994) *Strategies for Diversifying Assessment*, Oxford: Oxford Centre for Staff Development.

Brown, S. & Smith, B. (1997) *Getting to Grips with Assessment*, Birmingham: SEDA Publications.

Erwin, T.D. (1991) *Assessing Student Learning and Development*, California: Jossey-Bass.

Gibbs, G. (1998) *Teaching in Higher Education: theory and evidence. Chapter 4: Marking and Giving Feedback*, Milton Keynes: The Open University.

Gibbs, G., Morgan, A. & Northedge, A. (1998) *Teaching in Higher Education: theory and evidence. Introduction*, Milton Keynes: The Open University.

Gibbs, G. & Rowntree, D. (1999) *Teaching in Higher Education: theory and evidence. Chapter 3: Designing Assessment*, Milton Keynes: The Open University.

Holroyd, C. (2000) Are assessors professional? Student assessment and the professionalism of academics, *Active Learning in Higher Education*, vol. 1, no. 1, pp. 28-44.

Knight, P. (1995) *Assessment for Learning in Higher Education*, London: Kogan Page.

Knight, P. (2001) *LTSN Generic Centre Assessment Series No.7. A Briefing on Key Concepts*, York: Learning & Teaching Support Network.

Miller, A.H. Imrie, B.W. & Cox, K. (1998) *Student Assessment in Higher Education: a handbook for assessing performance*, London: Kogan Page.

Prosser, M. & Trigwell, K. (2001) *Understanding Learning & Teaching*, Buckingham: Open University Press.

Race, P. (2001a) *The Lecturer's Toolkit*, 2nd ed, London: Kogan Page.

Race, P. (2001b) *LTSN Generic Centre Assessment Series No. 9. A Briefing on Self, Peer & Group Assessment*, York: Learning & Teaching Support Network.

Rowland, S. (2000) *The Enquiring University Teacher*, Buckingham: Open University Press.

Rust, C. (2002) The impact of assessment on student learning: how can the research literature practically help to inform the development of departmental assessment strategies and learner-centred assessment practices?, *Active Learning in Higher Education*, vol. 3, no. 2, pp. 145-158.

Rust, C. (2001) *LTSN Generic Centre Assessment Series No.12. A Briefing on Assessment of Large Groups*, York: Learning & Teaching Support Network.

Schwartz, P. & Webb, G. (eds) (2002) *Assessment: case studies, experience and practice from higher education*, London: Kogan Page.

Thorpe, M. (2000) Encouraging students to reflect as part of the assignment process: student responses and tutor feedback, *Active Learning in Higher Education*, vol. 1, no. 1, pp. 79-92.

Yorke, M., Bridges, P. & Woolfe, H. (2000) Mark distributions and marking practices in UK higher education: some challenging issues, *Active Learning in Higher Education*, vol. 1, no. 1, pp. 7-27.

Case Studies

A Presentation: a preferred route for assessment in Counsellor Training

Di Brothers
University of Bristol, UK

Assessing students on Counselling Training programmes is complex (Johns, 1998). Our postgraduate Diploma in Counselling attracts students aged from 23 to 79, with widely diverse backgrounds and cultural and educational experiences. This provides us with a rich student group with some intrinsic challenges; this is especially evident within the writing of assignments.

As a practice-based training programme that provides the training element for professional counsellors, we aim to pay attention to theory, ability to practise and personal development, and consequently must include these three elements as key components within all the assessed pieces of work. In the case of students whose formal education has been limited, writing assignments that encompass all of these requirements has proved to be daunting and, in some cases, paralysing, as described beautifully by Maybank in 'An academic invalid in a world of academic excellence' (Johns, 1998).

An additional complication is that as trainers we aim to model what we teach and this is evident in our very student-centred approach to teaching and learning. We aim to pay full attention to each student's personal struggles to reach their potential and in this we reflect the core conditions of person-centred practice (Rogers, 1961). Through our teaching we provide many different and creative ways of learning, all of which are transferable into students' counselling practice. We expect our students to transfer *their* learning into their practice and yet it became obvious that we were excluding a wide range of potential methods of creative assessment from our assignments and that our emphasis on 'the written word' was disadvantaging a proportion of our less confident students. Feedback from the students and External Examiner echoed this point and so for these reasons we devised a new practical assessment to replace one essay.

Students are now asked to give a presentation on the theory of an area of life-span development that

has particular relevance for them and link this with their application to practice and personal experiences. They work in groups of three for the planning and the presentation and we encourage them to be as creative as they wish to be.

The presentations are delivered to the whole student group plus all tutors and there is an opportunity for feedback from the audience. The tutors meet separately and assess each presentation and detailed collaborative feedback is written. We do not assess presentation skills as this is not an area covered by our syllabus.

This assignment seems to give sufficient flexibility for all students to find their strengths so those of an academic frame of mind produce standard didactic pieces, whilst those who find creativity a route into learning have provided some amazing experiences whilst still retaining the necessary elements of theory, practice and self. A good example of the latter was a Sikh woman who rarely spoke in the group, taking the challenge and exploring through (and about) her veils the cultural transitions that she had had to make and the implications and complications of these transitions in our multicultural society for herself and her clients.

By working in small groups there is often some cross-fertilization so that the 'standard' presenter finds some small movement towards the creative and those with a creative mind find some more discipline. This too seems like additional personal learning.

One of our struggles was about the external examination of this assignment and because of regulations we had, finally, to include a request for students to write a brief personal reflection on their experience. This would also offer the opportunity to include any elements that were missed during the presentation (e.g. through mismanaging their time) and thus provide a safety net and, more importantly, would model the self-reflective writing that counsellors may use after seeing their clients.

The outcomes of this have been surprising. Feedback shows that it is the students' most powerful learning experience of the whole course and they appreciate the opportunity for their personal and professional development to be seen and valued by the group. The group itself becomes a safer learning environment and the quieter and less confident group members become more active. As tutors, we find that the live evaluation of this assignment mirrors our practice without losing any of the quality and we are always deeply moved by the courageous and personal presentations given by the majority of the students. We learn too.

Our mutual excitement about the success of this change of assessment method has widened our vision and deepened our experience. We plan to review other assignments to ensure that we offer a range of methods of assessment that are more inclusive and fit more comfortably with our broad cultural base.

References

JOHNS, H. (ed) (1998) *Balancing Acts: studies in counsellor training*, London and New York: Routledge.

ROGERS, C.R. (1961) *On Becoming a Person*, Boston: Houghton Mifflin.

Keywords

Counsellor training, creative assessment

Biography

Di Brothers has been working as a counsellor and counsellor trainer since 1976 and has been a core trainer for the University of Bristol's post graduate Diploma in Counselling (BACP Accredited) since 1990.

Demystifying Marking: reflections on developing and using grade descriptors

Ben Calvert
University of Gloucestershire, UK

Marking is a necessarily contentious activity. It is not always possible to arrive at bull's-eye conclusions about the merit of an individual piece of work, particularly in the arts, social sciences and humanities where studies of culture, society and history do not yield principles or general rules that can be tested for accuracy. If it were possible, double marking would be unnecessary and regulations allowing students to appeal their marks would become more or less redundant. But I use the words 'necessarily contentious' because marking should compel tutors to engage with one another about this difficulty, to keep on asking what it is that they are looking for in student work. Although we may not agree on the finer detail, we should welcome common ground where we can find it. It is vital to double-check that we know which skills and knowledge we are testing, that these are set at an appropriate level, that we all know when they have been met, and that they cohere within a curriculum. However, in modular programmes it is challenging to reach shared understandings when working lives can be isolated and disconnected.

In 2003, tutors in the Field of Media Communications at the University of Gloucestershire, attended an away day to improve marking. The session combined reflection on existing practice with planning for improvements. A key aim was to place marking in context. It was hoped that staff would better understand the subject's response to nationally disseminated standards in UK universities (QAA, 2002) and to locally-derived position statements on them (i.e. the Programme Specification for our course). The significance of these processes would become apparent by making them relevant to 'coalface' delivery, in this case assessment and marking. Throughout the day the idea that setting and grading assessments should be an expression of wider curriculum aims, as outlined in our Programme Specification, was reinforced. However, provisional discussions revealed that recently-recruited staff were largely unaware of national and institutional initiatives and saw them as rather distant and mysterious. Related developments, such as the Quality Assurance Agency's (QAA) guidelines for assessment (QAA, 2000) were also unfamiliar. A principal function of the session was for staff to familiarize

themselves with them by reading a document that tracked connections between benchmarking, programme specifications, level descriptors, learning outcomes, assessment criteria and grade descriptors.

The away day was organized as a mock QAA inspection meeting. Questions ranged from issues about student progression (e.g. 'What would you expect a model first year, second year and third year student to demonstrate in terms of skills and knowledge?'), to concerns about specific theoretical content (e.g. 'How do students' skills of textual analysis progress across three years?'). Staff were also asked how they differentiated between skills and knowledge within various levels (e.g. 'What would an A grade student in year three demonstrate in textual analysis compared to a B, C, D or fail student?'). This revealed information on student attainment that is often kept 'in the heads' of staff as implicit knowledge. Varied understandings also led staff to recognize the need for a degree of consensus in order for the curriculum to be coherent.

The team then provided examples of how our expectations are articulated by writing learning outcomes for a mode of textual analysis at all three levels, demonstrating progression. They then derived marking criteria related to these, informing students of the elements against which they would be assessed. Finally, grade descriptors were composed detailing the skills and knowledge required in order to score grades in each category. An example of practice in another subject was disseminated to assist in this. Previously, bespoke grade descriptors had been absent from our modules in respect to specific items of assessment. This is not unusual, with Rust (2002, p.147) arguing that although the use of learning outcomes is common, 'the linkage between the outcomes and the coursework essay, exam or whatever is tenuous at best, and almost always implicit'. For us, such descriptors existed only in generic form in the course guide, with statements applying across all three levels. The team agreed that these were inadequate for mapping attainment levels in individual assessments testing a variety of skills.

The team composed descriptors according to a grid system, with sets of statements applying to each marking criteria. An example of format is given in Figure 1, relating to an examination on a compulsory level two module. It is worth noting that individual written comments on student performance are also included on the definitive descriptor sheet.

At the end of the session a holistic understanding of the relationship between benchmarks and our curriculum, flowing down to individual assignments, outcomes, criteria and grade descriptors, had been achieved. It was agreed that from September 2003 module guides would include statements linking the module to wider programme aims, along with bespoke grade descriptors for all assessments. An agreed template for these guides secured consistency of practice. The

Figure 1: Example of Grade Descriptors

Marking Criteria: The exam answers will be assessed against the following criteria. All are given equal weighting:

i. Clarity of and specificity of the response to the question with evidence of a sustained and balanced argument

ii. Ability to evaluate and critique theoretical concepts and perspectives effectively and persuasively

iii. Ability to arrive at an informed and persuasive conclusions

iv. Clarity and fluency of expression

Grade Descriptors: marks will be awarded according to the following grid:

Marking Criteria	70+	60-70	50-60	40-50	30-40	Below 30	Mark
i	Response fully addresses the question and draws on a range of advanced sources to construct a balanced and sustained argument	Response addresses the question and draws on relevant sources to construct a balanced argument	Response addresses the question using relevant sources with evidence of a basic ability to construct an argument	Response addresses the question but may lack relevancy in places, may not fully utilize relevant academic sources and lacks a sustained argument	Response may address the question indirectly and arguments tend to be intuitive or descriptive with few relevant academic sources consulted	Response to the question partial or unclear with unconvincing, poorly sourced arguments	55
ii	Theoretical positions fully and clearly explained, demonstrating an ability to critique persuasively on the basis of evidence. May contain evidence of original thinking or synthesis	Theoretical positions clearly explained, demonstrating ability to critique persuasively using evidence	Theoretical positions explained and critiqued but explanations may not be full or may have some inaccuracies, or critique lacks detail or persuasiveness	Theory explained at a basic level, but little evidence of critique. May contain some theoretical inaccuracies	Partial or inaccurate exposition of theory with only an intuitive ability to critique demonstrated	Little or no engagement with relevant theory	62
iii	Theoretical debates and evidence carefully drawn together to construct a focused argument that convinces the reader fully of the writers understanding. May contain evidence of original thinking	Theoretical debates and evidence summarized effectively. Evidence of an ability to construct a persuasive argument	Theoretical debates summarized clearly, with a basic argument or position taken	Theoretical debates summarized but with little evidence of an informed argument or a convincing summary position being taken	Theoretical debates inadequately summarized. Any arguments likely to be intuitive or descriptive	Little or no evidence of a theoretical conclusion. Entirely intuitive or descriptive	64
iv	Impressive command of language, including spelling and grammar. A fluency of style and an ability to make complex debates clear is in evidence. Work well structured with clear bridges and links built between key points	Clear command of language, including spelling and grammar. Style is generally fluent and an ability to make points clearly is in evidence. Work is soundly structured	Good command of language, including spelling and grammar, although style may lack fluency. Soundly structured, but may lack some clear bridges and links between key points	Sound command of language, including spelling and grammar. May contain some basic mistakes or lack fluency or a clear structure in places	Basic command on language but several mistakes in spelling and grammar are evident, and work lacks fluency or a clear structure	Poor command of language with many basic mistakes of spelling and grammar and little fluency or structure	58

template was derived to ensure that curriculum aims and the learning and assessment strategy of modules spoke to each other in a process of 'constructive alignment' (Biggs, 2003). It was also hoped that descriptors would be useful for students in formative assessment by directing attention to the meeting of module outcomes. For staff they would draw attention to common issues within assessments and focus minds on whether or not agreed outcomes had been met. In both cases, the aim was to demystify marking.

Currently, our use of this approach is being monitored. Preliminary views have revealed a balance of positive feedback and genuine concerns. Whilst all staff felt that descriptors focused their attention more clearly when marking, and in particular made double marking more transparent, some felt that this was at the expense of an overall evaluation of a particular piece of work and its merit. Some would read the work and allocate an 'instinctive' mark and then apply the descriptors to fit this impression. They believed that certain assignments 'felt like' a 65% or 50%, for example, but argued that this mark might not emerge via strict adherence to the grade descriptors. As a result, they ended up marking in much the same way as before and adapted the descriptors to their existing marking style. The view that marking (and perhaps teaching and learning more widely), is a complex activity that cannot be adequately reproduced in a tighter regime of marking was also expressed. In an appraisal of the use of learning outcomes, Hussey & Smith (2002) have argued that, when misused, they perform a managerial function as an aspect of the commodification of learning. This reflects less their use for deepening learning and more 'the desire to audit and monitor the performance of those involved'. It is possible that some staff may see the use of grade descriptors in the same way. Whether or not this was the case within the team, the exercise of judgment when marking had been confronted and in some cases changed.

This led to further discussions about the way in which the criteria were written. Was it better for them to be tight or loose in order for some discretion to be allowed? The provision of an additional catch-all descriptor, based on a general impression of the work was also debated. This could allow staff to weight the work according to both the tightly written descriptors and their impressionistic sense of its merit. Others felt that this would remystify the process again and would run against the spirit of using the descriptors in the first place. Student responses have also been mixed. In module evaluations some have welcomed the use of the 'shaded grid' method as it focuses clearly on what they need to do to score particular grades. However, they do not see tutors shading the grid as 'feedback', which they perceive only as a personal written statement. Although the team do write individual feedback as well, some students felt that there was an issue of balance between generic responses (albeit useful ones) and those that are more personalized.

At the moment decisions about whether to modify their use are on hold as it is felt that time is needed to adjust to a new mode of marking. Staff recognized that they have the power to determine the phrasing of the descriptors as long as they reflect the appropriate marking criteria and learning outcomes. If staff wished to write descriptors that encouraged perhaps more elusive factors (e.g. originality or a creative approach to a piece of work) then this was fine, as long as students were aware that this was an expectation and as long as marking criteria and module outcomes required those elements to be assessed. At the end of the meeting it was emphasized that descriptors do not necessarily diminish the professional judgement of staff, they just require them to exercise it differently, in a way that is more transparent and meaningful for students. In summary, the team is continuing to use grade descriptors and is generally persuaded by their value, contingent upon modifications being discussed at the end of a full academic year.

References

Biggs, J. (2003) *Teaching for Quality Learning at University*, 2nd ed, Buckingham: Society for Research into Higher Education and Open University Press.

Hussey, T. & Smith, P. (2002) The trouble with learning outcomes, *Active Learning in Higher Education*, vol. 3, no. 3, pp. 220-233.

Qaa (Quality Assurance Agency for Higher Education) (2000) *Code of Practice – Quality and Standards in Higher Education. Section 6 'Assessment'* [online]. Available from http://www.qaa.ac.uk/public/Cop/COPaosfinal/contents.htm.

Qaa (Quality Assurance Agency for Higher Education) (2002) *Benchmark Statement for Communication, Media, Film and Cultural Studies* [online]. Available from http://www.qaa.ac.uk/crntwork/benchmark/phase2/communications.htm.

Rust, C. (2002) The impact of assessment on student learning, *Active Learning in Higher Education*, vol. 3, no. 2, pp. 145-158.

Keywords

Grade descriptors, marking, benchmarking, constructive alignment

Biography

Ben Calvert's background is in the social sciences and he has recently published work on television studies. He has undertaken a funded project on effective dissertation supervision and is currently undertaking research on developing a research culture in first year undergraduate students.

An Aligned Assessment to Promote Learning About Collaboration Between Health and Care Professionals

Colleen Connor
University of Wales Institute Cardiff (UWIC), UK

The MSc in Inter-professional Studies at the University of Wales Institute Cardiff was launched in 1992, and currently offers three pathways in 'Health', 'Learning Disability' and 'Quality Assurance'. The students attending the MSc programme are part-time and have professional posts in e.g. nursing, social work, dentistry, occupational therapy, etc.

All students take a core module entitled 'Inter-professional Workshops' (IPW). This module aims to develop a critical understanding of, and relevant skills for, collaboration across the health and social care professions.

The principal aims of the module are:

- to provide students with relevant knowledge about collaborative activity and the skills to up-date such knowledge independently. Developing skills in the presentation of such knowledge is also important
- to enable students to analyse the processes and actions that occur when diverse professional groups collaborate and to apply such understanding to the development of appropriate procedures to facilitate collaboration
- to teach students how to evaluate their personal contribution to collaborative working experiences, in order that this can inform their working practice.

When designing the module it was evident that more 'traditional' methods (e.g. essay, exam) would not be sufficient to test whether students had achieved the required knowledge and skills. Moreover, the assessment, in addition to being a good indicator of achievement, needed to be integrated in a way that was meaningful to the students and furthered their learning rather than being viewed as simply burdensome and a necessary evil. Past experience showed that, owing to their considerable commitments, students often withdrew from the programme at the point of assessment if not purposefully engaged in the activity.

After much deliberation the eventual assessment strategy employed four different elements. Although this may, at first sight, appear

rather onerous, their integration with one another and with the learning strategy compensated for the number of elements. The overall approach sought to achieve a 'constructive alignment' between learning and assessment. This refers to ensuring that '… the teaching methods used and the assessment tasks are aligned to the learning activities assumed in the intended outcomes' (Biggs, 2002). The 'trick' to achieving such alignment appears deceptively simple; according to Biggs it is 'to make sure the assessment tasks mirror what you intended them to learn'.

Briefly, the module is delivered over a period of eight months. During this time students attend two weekend workshops and participate in a number of exercises and discussions related to inter-professional collaboration and teamwork theory. In the first of these workshops they are divided into two workgroups and set a group task to be undertaken between workshops. Each group is required to agree an inter-professional topic that all group members find relevant and to research this topic throughout the module and finally present an analysis of the topic as a group.

The module is assessed through each group presenting a portfolio of evidence of how they worked together (20%) and a joint presentation of their topic (30%). Each student also completes an individual essay reflecting on their group experience and critically discussing some key issues in inter-professional collaboration (40%). Finally, within their groups, students undertake a peer review of one another in terms of their contribution to the group activity (10%). This is carefully structured and facilitated by the module leader to ensure that it is a constructive learning experience for everyone.

The portfolio allows an ongoing record and final assessment of the process of the workgroup and thus demonstrates how procedures are employed to improve collaboration between participants. The presentation demonstrates their ability to collate, interpret and convey this knowledge and to do so co-operatively. The reflective essay allows an analysis of group processes and consideration of their significance to individual work situations. Finally the peer review, although only 10% of the assessment, is extremely useful in sensitively discussing individual contributions and learning needs when collaborating with others. Thus, although all focused on the same task, each element tests a different aspect of the learning outcomes.

The module has been delivered in this form for four years with a recent development being the introduction of computer-mediated communication to supplement the work groups' collaborative processes (see Connor, 2003, for further details). Some groups have used transcripts of their online discussions to provide evidence of reflection and analysis. The use of the discussion board has allowed greater flexibility for students (owing to the asynchronicity afforded) and greater transparency of the group process for the tutor. The next stage

will be to develop the assessment criteria within the portfolio element to formally utilize the visibility of the online communication.

Implementation

As with all innovations in assessment, implementation issues have been numerous and within this short account only a few can be briefly mentioned. They concern students' own doubts about an unfamiliar approach, issues of group size, student absence and 'freeloading'.

Students were initially somewhat sceptical about the benefits of the group project, particularly when told that they were expected to decide their own topic, and earlier groups put considerable pressure on the tutor to be 'given' the topic to explore. However, all groups have eventually recognized the value of the learning achieved through participation in the process of negotiating a topic with colleagues. Far greater explication of the reasons for the topic not being set by the tutor together with an explicit input on negotiation processes and skills has helped.

Student withdrawals, although infrequent, also cause some alarm with students being concerned about group imbalances. For example, a group that has an eventual size of 4 sometimes feels disadvantaged compared to a group that has an eventual size of 6. A 'size isn't everything' reassurance that quality, rather than volume is the key to portfolio evidence together with discussions of the pros and cons of relating in small or larger groups has been necessary to allay fears.

The thorny problem of how to deal with a student who is not available for the presentation itself but has contributed fully to the group process and preparation has not arisen in actuality, but has been discussed by the course team and with external examiners. It has been agreed that should this occur, and provided the role the student would have taken can be adequately demonstrated, they would still receive the group mark.

A related issue is that of the possibility of a student 'freeloading' within the group. This hasn't occurred, but regular progress reviews, the visibility of online work and the possibility of tutor intervention probably influence this.

The reliability of group assessment has been questioned (e.g. Race, 2001) and this has, in part, influenced the decision to diversify the assessment tasks. However, there is a strong feeling in the course team that the benefits of including group assessment far outweigh any such concern.

On a positive note, although time intensive in the design stage, the group project is a more efficient means of assessing in the long term, a significant factor given tutors' concerns about the assessment workload. More importantly the project, through its requirement to work in an inter-professional group, increases assessment validity.

Both students and external examiners have consistently evaluated the learning and assessment as valuable. The success of the module owes a great deal to the alignment of the assessment strategy with students' learning needs and the validity of the methods used.

References

Biggs, J. (2002) Aligning the curriculum to promote good learning. A paper written for the *Constructive Alignment in Action: Imaginative Curriculum Symposium, 4th November 2002* [online]. Available on the LTSN Generic Centre website http://www.ltsn.ac.uk/.

Connor, C. (2003) Virtual learning and interprofessional education: developing computer-mediated communication for learning about collaboration, *Innovations in Education and Teaching International*, vol. 40, no. 4, pp. 341-347.

Race, P. (2001) *LTSN Generic Centre Assessment Series No. 9. A Briefing on Self, Peer & Group Assessment*, York: Learning & Teaching Support Network.

Keywords

Inter-professional, assessment, collaborative, integrated

Biography

Colleen Connor is Head of the Centre for Inter-professional Studies at UWIC and also manages quality enhancement activity across the university. She is an Accreditor for the Higher Education Academy (formerly the Institute for Learning & Teaching in Higher Education — ILTHE) and has been involved in a number of research projects concerned with inter-professional education on both undergraduate and postgraduate programmes.

Student Self-evaluation of Coursework Assignments: a route to better perception of quality

Averil Macdonald
University of Reading, UK

The initial problem

Lecturers routinely put in hours of work marking assignments and providing individual feedback only to find that the next piece of work submitted by the same student shows precisely the same qualities — and the same flaws. I became aware that my students seemed only to notice the 'bottom line', i.e. the grade allocated, with a perfunctory scan of the rest of the document in case it was apparent they could increase the mark allocated through appeal. Rarely was there evidence of students taking the advice offered and making an effort to address the particular areas of weakness. Some seemed unable to understand how one piece of work was worthy of a higher grade than another; taking it as just a fact of life that other students would score more highly than they. After all this is what they had experienced for more than a dozen years of education thus far, so why should anything change now?

Some students have the ability to perceive what is required of them and the quality of work that warrants a high grade from the beginning. It is those others, who seem not to be aware of the difference between what they submit and what was deemed worthy of commendation, whom I wanted to address. By engaging the students in the grading process, thereby making them aware, very clearly, of what was missing from their work, I hoped to encourage them to be more critical of their own work, to change their practices and to enable them to generate assignments of a higher quality.

I initially devised a self-evaluation document to supplement an assignment in which students were asked to generate a 250-word précis of a 4-page article on the 'Acoustics of Concert Halls'. The intention of the assignment was to make them read the article critically and extract from it the most pertinent facts and redraft them into a readable piece not exceeding the word limit.

The self-evaluation document (Appendix 1) listed 15 important points that students may have included. Students were required to identify which points their précis contained plus any other points they had included that were not on the original list. A grade was allocated depending on the total number of points included. However, beyond this students had to write a brief

evaluation of their work to say why it was worth the grade they allocated. This was particularly pertinent if they were seeking credit for points they had identified but which were not on the list. It was at this point that students could argue for a higher grade than simply counting the points would warrant.

How practice was changed

The initial trial worked well overall with over 80% of students allocating themselves the grade I considered appropriate. However, this was a very new experience for the students and they were most suspicious of how it could work and very wary that less honourable students may try to lie to achieve a higher grade. There is no doubt they have a innate sense of justice coupled with complete trust in the ability of lecturers to mark fairly and are most concerned by the possibility of fraud when faced with a new system. The fact that the self-evaluation itself is subsequently 'marked' has to be made clear.

Some of the students took exception to having to do what they saw as 'my work' (even though I went through all the work to check their evaluation subsequently!), so the task has to be presented to them as an integral part of their learning process with the benefits to them spelt out clearly.

It was apparent also that the wording of the self-evaluation document had to be less confrontational; e.g. changing the phrase 'points you should have included' to 'points you may have included', as this wording immediately caused some students to see it as a matter of principle to challenge the mark allocation criteria (something they rarely do when faced with a lecturer-marked assignment).

Using the same format to enable students to evaluate an essay they write later in the year (see Appendix 2) proves more successful both in terms of their ability to allocate the appropriate grade and their attitudes to the process as a whole, which would indicate that their increase in experience facilitates the effectiveness of the process.

Gains and losses

The biggest gain is the fact that students are provided not only with a grade for their work but a clear indication of why it was worth the grade awarded and, therefore, what they could have done differently to achieve a higher grade. This does not, however, guarantee that they will make the necessary changes to their working practices next time around, but it does increase their awareness of why they, perhaps, are not achieving as highly as their colleagues.

I also discovered another interesting element regarding the psychology of valuing oneself. I commonly find that the female students will underestimate their grade and will be very modest in how they evaluate the quality of their work, waiting for me to tell them it's much better than they have claimed. I (as a woman) put this down to the fact we are taught from a very early age that it is not polite to sell oneself overtly and that

modesty is a virtue. Male students, on the other hand, seem to consider the whole process as a competition in which their role is to acquire the best grade possible by working out what the rules are and playing them to their best ability. I'm sure that experts in psychology will be able to tell me if this observation is more general than simply an undergraduate physics class. If it is more general then there is likely to be a benefit for girls who may realize there is no good reason to be overly modest in this very competitive world.

The biggest loss is the fact that it is harder work to persuade the students that this self-evaluation exercise is valid and fair. You have to be prepared to talk it through with them both as a group and as individuals and be prepared to argue the case that it is a useful educational experience. The fact that they will be required to undertake self-appraisal in the workplace is, in my opinion, a valid 'real world' reason for such exercises beyond the benefits they gain educationally.

Future development

The same approach has already been applied to an extended essay that students are required to write. Students are certainly more accepting of the process when they meet it for the second time. We are now working on developing a similar process whereby students will evaluate the quality of their practical write-ups. It has been observed for a long time that lecturers find themselves writing the same comments week after week while first year students seem immune to the idea of including errors, for example, no matter how often it is pointed out to them. The plan is to require students to indicate what has or has not been included in their write up, as per the exercises described above. However, they will also have to draw out particular pieces of information from their logbooks, such as the units and uncertainty of a reading, the gradient of a graph, the final value obtained and the error in the final value. If the student has kept a log of reasonable quality then these facts should be easily found whereas those who are more lax about keeping note of experimental details or who have not completed their write-up will be unable to provide the information. In using their log directly this way they should see the need to ensure their work is both legible and sufficiently detailed.

Keywords

Assignments, feedback, self-evaluation

Biography

Averil Macdonald works part-time in the Physics department of University of Reading and part-time as an educational consultant. She has contributed to 16 school textbooks and written web-based teaching resources making university research accessible for use in GCSE and A-level Physics classes. She was awarded the 1999 Bragg Medal by the Institute of Physics for her distinguished contribution to Physics Education and popularising Physics.

Appendix 1

Professional Skills **Name**

Executive Summary of 'Concert Hall Acoustics': self-evaluation

The aim was to include as much pertinent information from the paper within the word limit. Below are listed some of the most significant points which you may have included:

Introductory sentence:

1. Paper considers the association between the physical characteristics of a room and the subjective appreciation of sound quality

Background:

2. Early work identified relationship between reverberation time and sound absorption and volume of room
3. Studies in 1950s showed importance of early reflections
4. Early reflections reinforce direct sound and hence increase clarity
5. Reverberation time defined as time for total sound amplitude to decrease by 60 dB
6. High reverberation time (or high level of reverberant energy relative to direct sound) tends to blend sound — preferred for music
7. Reverberation time should be constant throughout hall
8. Shape of hall determines direction and time of arrival of reflected sound
9. Square hall has first reflections from side walls
10. Fan shaped hall has weak reflections from side walls but has first reflections from ceiling
11. Studies in 1960s showed early side (lateral) reflections increase audiences' perception of spaciousness (feeling of being in a room)
12. Strong, early lateral reflections increase apparent width of sound source (as source position is ambiguous)
13. Strong, late side reflections increase sense of listener envelopment

Conclusions/final paragraph:

14. Designers now include large side wall reflectors in new halls to increase the number of possible hall shapes
15. Surround sound entertainment systems have to consider late lateral reflections to recreate concert hall effects.

Evaluate your report. Firstly tick off the listed points you included. On the reverse of this sheet write an evaluation of your work. You may also have included additional points which you consider equally important. Justify any additional points for which you wish to claim credit. Consider also the layout of your work, accuracy of punctuation and comment on the overall quality of the work and why it is worth the grade awarded.

Grades as below:

12+	9 - 11	6 - 8	3 - 5	1 - 2	points
A	B	C	D	E	

Appendix 2

Professional Skills **Name**

Self-evaluation of essay based on 'In Search of Schrodinger's Cat' by Gribbins

'The Bohr model of the atom has been useful despite, or perhaps because of its flaws.' Discuss.

Below are some of the most significant ideas which you may have used in your essay:

Introductory paragraph:

1. Definition of 'classical' Rutherford atom with randomly placed electrons as basis for Bohr model
2. Bohr noted it should not work — electrons should spiral in radiating energy
3. Bohr postulated :
    - allowed energy shells where electrons just do not radiate
    - particular values of angular momentum
    - quantum leaps — electrons move between energy levels emitting radiation as photons

    (may list all postulates here)
4. Result of Bohr's postulates is an atom linking quantum theory to the classical atom

   (N.B. there was no theoretical justification for these ideas — they were just ideas)

Successes of the Bohr atom:

1. Explained spectral lines especially of hydrogen
2. Explained chemical interaction, k.l.m. shells etc
3. Predicted unknown elements

Flaws in Bohr's atom:

1. Predicted too many spectral lines
2. Doesn't explain energy BANDS and broadening of spectral lines
3. Assigns quantum numbers randomly to fit observations — no underpinning theory
4. Needed too many adjustments — Somerfield spent ages refining the model to fit each new observation (the Bohr – Somerfield atom)

Usefulness of Bohr atom despite its flaws:

1. The classical orbital model is easy to visualize — especially useful as first version of atom to teach to younger students
2. Made people think seriously about bringing quantum theory into the model of atom — a very brave thing to do

Usefulness of Bohr atom because of its flaws:

1. Flaws are obvious so demand criticism and improvement — set other scientists thinking and led to further models

Concluding paragraph stating personal view:

Evaluating your essay. Tick off the ideas overleaf that you included. You may also have included other ideas which you consider important. Identify these on your essay. You may justify inclusion of any additional ideas for which you wish to claim credit in your evaluation below.

Write an evaluation of your work taking into account how many of the points you included from the list overleaf and any additional ideas you raised. Consider, also, the layout of your work and the accuracy of punctuation.

Grades as below for number of ideas included:

13+	12 - 10	9 - 7	6 - 4	3 - 1	points
A	B	C	D	E	

Here write your justification for the grade you award:

GRADE AWARDED

Triadic Peer Review in Scenario-based Assessments

Debra Nestel, Roger Kneebone & Jane Kidd
Imperial College London, UK

Initial prompt/problem

In undergraduate medical education, students are expected to achieve competence in procedural skills (e.g. suturing a wound, inserting a urinary catheter). In the United Kingdom, most medical schools have clinical skills laboratories in which students practise technical procedural skills on simulated models before performing them under supervision on real patients. Although traditionally the focus of assessment, technical skill is just one component of the procedure and when students work in clinical settings they need to be able to integrate technical with communication skills and to respond to contextual stimuli (e.g. interruptions by staff, equipment unavailable).

Scenario-based assessments provide a means by which students can practise merging these complex skills in simulated and real settings (Kneebone *et al*., 2002a; 2002b; Nestel *et al*., 2003). By linking simulated models (e.g. suture pads, pelvic models) with actors who are trained to portray patient roles and give feedback to students on their communication skills, students are provided with a safe environment in which to develop their clinical skills as they will be performed in real work settings.

The key elements of scenario-based assessments are:

1. *Preparation*: Immediately before the procedure, students are prompted to think about the technical and communication skills required for the procedure and contextual factors that may influence performance (e.g. time frame).
2. *Performing the procedure*: The student performs the procedure on the actor in the simulated setting. Each procedure is observed in real time by technical and communication skills experts and recorded for later review by the student.
3. *Reflection*: The student is encouraged to write brief and immediate reflections on what worked well and what could have been improved during the procedure.
4. *Feedback*: Students receive focused feedback from the actor and communication and technical skills experts.

Although students have found scenario-based assessments highly realistic and powerful learning experiences, these formative assessments are not sustainable for large cohorts of medical students because of the extent of expert involvement required. Therefore, we proposed a triadic peer review process in which students worked in groups of three rotating through each of the three roles: technical and communication skills observers and student performing the procedure. Collaborative learning theory emphasizes group participation in which students share knowledge, attitudes and skills based on their prior experiences to explore, discuss and evaluate a problem (Brufee, 1999).

For each observer role, students are provided with rating forms to focus their observations. Each item on the rating form is accompanied by explanatory behavioural markers. The student performing the procedure is given a clear statement of the task and all students are provided with a protocol for giving feedback.

What the practice was trying to achieve

Triadic peer review addresses the resource issue of two expert faculty working with three students during a 2-hour session. The students involved in this project had all previously received technical and communication skills training relevant to the procedures and were familiar with the feedback protocol although they had not formally facilitated a learning session as part of their clinical skills programme. Rather than deny students the opportunity of this unique experience because of resource limitations, we conducted a pilot study to evaluate the feasibility of triadic peer review in formative assessment.

Identification of the gains and losses

Potential gains include: student engagement in each role; enhancing awareness of the specific skills required for procedures; providing a structure for formative assessment; developing critical observation skills; providing an insight into educational approaches and ensuring receipt of feedback from the *patient's* perspective.

Potential losses include: incorrect and/or inappropriate behaviours are left unchecked; students may reinforce inappropriate behaviours and students may find peer review threatening, which could compromise the quality of feedback. These potential losses are minimized by the use of the rating forms, the feedback protocol which prompts sensitivity in the order and balance of feedback and the presence of the actor. All procedures and feedback are recorded so can be reviewed if required.

Method of evaluation

Two qualitative evaluation methodologies were used:

1. Semi-structured interviews were conducted after each scenario-based assessment

Senior medical students were recruited to participate in the study using a convenience sampling strategy based on availability and practical constraints. Twenty-two students participated in sessions between September 2002 and June 2003.

In response to being asked about giving and receiving feedback from peers, students largely found the process constructive.

> Always useful listening to colleagues' criticisms ... they know you and can relate better, therefore it's useful.

> It's good to get feedback from peers. That together with being able to review the procedure would be very valuable. You're not always aware of things that you might be doing wrong or well.

While most students reported familiarity and value in peer-assessment, one student reported a limitation.

> If your knowledge is a little sketchy you might not pick up on another's faults.

Another student suggested peer-assessment was without benefit.

> It would take someone to be really dreadful before saying something, and would feel awkward telling someone that they were dreadful. Therefore peer review seems pretty pointless.

Students were asked what they would do if they disagreed on any aspects of the assessment.

> We'd just ask a clinician. That's what we would usually do.

One trio discovered they had been taught differently to each other and asked for clarification. This alone was a valuable learning experience. Some students suggested working with mixed cohorts so that seniors could work with juniors.

2. Observations by the research team

Faculty observations of procedures did not identify any unsafe skills that went unchecked. However, the feedback provided by students was less articulate than that delivered by experts.

Plans for future development

Consider implementation for additional clinical skills and for the entire cohort of students. Develop evaluation strategies that target measurement of professional skills other than those required for conducting the procedure (e.g. critical observation, facilitation and feedback).

References

Brufee, K. (1999) *Collaborative Learning: higher education, interdependence and the authority of knowledge*, Baltimore: Johns Hopkins University Press.

Kneebone, R., Kidd, J., Nestel, D., Asvall, S., Paraskeva, P. & Darzi, A. (2002a) An innovative model for teaching and learning clinical procedures, *Medical Education*, vol. 36, no. 7, pp. 628-634.

Kneebone, R., Nestel, D. & Darzi A. (2002b) Taking the skills lab onto the wards, *Medical Education*, vol. 36, no. 7, pp. 1093-1094.

Nestel, D., Kneebone, R. & Kidd, J. (2003) Teaching and learning about skills in minor surgery – an innovative course for nurses, *Journal of Clinical Nursing*, vol. 12, no. 2, pp. 291-296.

Keywords

Peer review, peer assisted learning, collaborative learning, integrated assessments

Biography

Debra Nestel (Monash University), Roger Kneebone (Imperial College London) and Jane Kidd (Warwick University) developed scenario-based assessments at St Mary's Hospital, London and continue to work on projects that integrate communication with technical and contextual skills using simulation at undergraduate and postgraduate levels in medical and other health care professional groups.

A Framework for Formative Assessment: initiating quality learning conversations

Alan Robinson & Mark Udall
Southampton Institute, UK

Problem

This case study is based on a level 2 unit that forms part of the curriculum of a BEng (Hons) Engineering programme. The unit introduces a software engineering theme and draws on knowledge and skill areas that are different from those which form the focus of the other units on the course. A survey indicated that students perceived the unit as not being central to the specific engineering discipline in which they were interested and judged the unit to be of less relevance to them than the other units studied. This was particularly the case for the part-time students who were already practitioners in specific engineering roles. These students saw less need for the 'broadening subjects' that comprise the curriculum. Students had the tendency to 'slow start', not completing early formative tasks designed to aid completion of the summative assessments later in the unit. This resulted in a lack of understanding of the early material thus forming a poor foundation for later, much more complex, concepts. Conversations about the quality and quantity of student learning in process were mainly initiated by the tutor answering questions that the students had not themselves even formulated.

Intervention

The intervention described here uses a redesigned formative assessment strategy that encourages and enables learners to instigate conversations about their learning.

The intervention has a number of key features:

1. All 'classroom-based' learning activities have stated intended outcomes, which articulate the specific knowledge, understanding and skills (both cognitive and practical) associated with that activity. The activities are designed to allow the outcomes to be delivered at different levels

2. These activities require learners to undertake specific preparation, participate in the sessions and make a self-assessment of whether they have met the intended outcomes at a threshold level or above

3. The students record on a single 'progress record' the preparation, activities and post-activity self-

assessment and identify the extent to which the outcomes have been met

4. The students record specific questions for the tutor based on outcomes that they feel they have not yet met.

The activities and the ensuing learning conversations between peers and between learner and tutor focus strongly on what the learner is 'doing', as well as the output from what they are 'doing'.

The progress record is an important aspect of this as it makes the learning tasks explicit and provides a visualization of the extent to which each learner is actually engaging with their own learning.

The specific questions (in 4, above) are formulated and recorded by the learner. The intention here is that by 'owning' these questions, the answers themselves have greater meaning. Further, the learner is instigating the conversation which forms the formative assessment activity. The use of language and structure adopted when the learner articulates these questions can be used as an indicator of the extent of learning. From a SOLO Taxonomy perspective (Biggs, 1999), some learners construct and articulate highly relational questions whereas others, with a less sophisticated understanding of the subject area, tend to pose unistructural or multistructural questions. This provides a quick diagnosis that allows the tutor to explore answers in different ways with different learners to reflect their level of development. Multistructural questions formulated by the students were identifiably short and focused on the acquisition of facts or the clarification of definitions, 'what is verification?' More relational questions were longer, more sophisticated and dealt with the application of higher level ideas and concepts, '… so what sort of verification techniques would be used in these circumstances?'

Evaluation

The intervention is based on the Mexican Hat Approach[1] (Robinson & Udall, 2003) and has been evaluated by the authors and two external researchers, using inventories and a variant of the Delphi technique[2] (Linstone & Turoff, 1975). For this particular evaluation exercise, the Delphi technique used was computer-mediated. This maintained the advantages of Delphi as a feedback technique but allowed the approach to be more quickly conducted.

From a pragmatic viewpoint, the result of this intervention was improved unit pass rates and higher quality learning outcomes leading to all learners achieving at least a threshold pass.

From a learner perspective, the evaluation results show the importance of clear intended outcomes for activities. These were seen as key to identifying why things were being done, rather than just what was to be done, which was their prior experience of very task-oriented activities. Learners also

felt that they had a much better understanding of how well they were doing as they progressed through their studies and the summative assessment did not come as a big surprise. Some of these learners felt that the intervention reduced anxiety and improved their motivation for study generally.

From a tutor perspective there was clear evidence of a higher quality, learner-driven, dialogue about learning. The conversations were more readily instigated by the learners and comprised richer and deeper questioning.

Developments

As a student explained as part of the evaluation 'More units should adopt this technique as it makes it far clearer what is expected of me and I find it easier to learn the subject.' The framework is currently being used and developed across a number of different units and courses, both at Southampton Institute and in other higher education institutions.

The framework has now been adapted to form part of a Virtual Learning Environment (VLE) and the outcomes of this project will be evaluated later in the academic session.

Notes

1. **Mexican Hat model**
   The Mexican Hat model promotes the design of an aligned teaching, learning and assessment strategy with a particular focus on increasing formative assessments, but within a manageable overall assessment workload. The model provides a visualization for students and teachers of the quantity and quality of learning in process. Therefore, providing an ongoing indication of the likelihood of which students will achieve success so that the appropriate type and level of support can be targeted.

2. **Delphi method**
   After an initial individual non-collaborative brainstorming session in the classroom on a specific question, the facilitator collates the group's views and emails them to each student as a questionnaire. The students' reflections are processed and a revised questionnaire is sent asking students to assess the merit of each idea, using a scale that ranges from '0' (no potential) to '7' (very high potential for dealing with the issue) or 'N' (no judgement).

References

Biggs, J. (1999) *Teaching for Quality Learning at University*, Buckingham: The Society for Research into Higher Education & Open University Press.

Linstone, H. & Turoff, M. (1975) *The Delphi Method: techniques and applications*, Reading, Massachusetts: Addison-Wesley.

Robinson, A. & Udall, M. (2003) Developing the Independent Learner: the Mexican Hat approach, in *Conference Proceedings of the IEE 3rd International Symposium on Engineering Education*, *Southampton, UK, January 6-7, 2003*.

Keywords

Formative assessment, student self-assessment.

Biography

Alan Robinson is Associate Dean (Operations) and Mark Udall is Principal Lecturer in Computing and Educational Developer for the Faculty of Technology at Southampton Institute. Their research interests currently focus on formative assessment and the related issues of student participation and retention. The Mexican Hat Approach project has been developing for 5 years and has received funding from both Southampton Institute and Learning & Teaching Support Network (LTSN) - Engineering.

Supporting 'HOCS Learning' via Students' Self-assessment of Homework Assignments and Examinations

Uri Zoller
University of Haifa – Oranim, Israel

The essence of the current reform in science education, worldwide, is the shift from the contemporary dominant traditional algorithmic lower-order cognitive skills (LOCS) teaching, to the higher-order cognitive skills (HOCS), promoting *learning*; that is, the development of students' capabilities of question-asking, critical/system-*evaluative thinking*, decision-making, problem (not exercise) solving and *transfer* within both the science disciplines and real life interdisciplinary situations in the science-technology-environment-society (STES) context (Zoller, 1999). Clearly, such a paradigm shift in the educational goals of both teaching and learning of science requires not only the application of new teaching strategies but also *alternatives* to traditional assessment strategies; that is, new examination and evaluation methods which are not only consonant with the new goals, but also foster them (Zoller, 1993; 1999; 2001).

A major issue of concern is the development and implementation of appropriate teaching and assessment strategies which would indeed support and promote effective 'HOCS learning' in traditional disciplinary science courses at all levels (Zoller, 1997). Within our longitudinal research-based and teaching experience-based efforts to promote our science major students' HOCS, we have incorporated, among others, the following four teaching-learning assessment methods in our freshman introductory, general and organic chemistry courses:

1. Self-study of pre-class lecture material; namely, students who have the course outline, scheduling, objectives, requirements and assignments in their hands, self-study the relevant 'material' *before* it is 'covered' in the class, to which they bring *their* questions for discussion

2. No specific course textbook(s) are assigned; rather, the students are provided, at the beginning of the course, with a list from which they can choose textbooks and reference books, to use for the study of any relevant topic, as they find appropriate for their needs, during the course

3. Several homework assignments — mainly problems (not exercises) which require HOCS for their 'solution' (e.g. Zoller *et al*., 1997) — to be worked out by the students (preferably in groups) and submitted, *individually*, for feedback and grading by the teaching assistants, former 'graduates' of these courses

4. Students' self-assessment (Zoller *et al*., 1997) of their home assignments, pre-guided by the course professor. Needless to say, all the exams in these courses are of the open-book/HOCS-type (Zoller, 1997) which not only are consonant with this approach, but also promote 'HOCS learning'.

A pre-condition for successful HOCS-promoting teaching, is the provision, — to be made for students — to actually experience and practise these strategies. Indeed, this is the essence of the approach described here. The implementation of these strategies was closely, formatively and summatively, followed by action-type research. The results of one related 'case study' are given below.

The average students' grading of one of their mid-term HOCS-oriented homework assignments, in a freshman Introduction to Modern Organic Chemistry course for biology majors, was 82.8 (out of 100), whereas that of their teacher assistants in the same course was 83.8. Similarly, the average students' self-assessed score on their final oral exam, in this course, was 79.4 (out of 100), compared with 78.8 of their course professor (author of this article). The above, as well as other related studies, suggest that although the road to 'HOCS learning', in terms of conceptual learning, critical system thinking, problem solving, evaluative thinking and transfer in the higher education context, is rocky, it is attainable (e.g. Zoller 2003). It can and should be done.

References

Zoller, U. (1993) Lecture and learning: are they compatible? Maybe for LOCS; unlikely for HOCS, *Journal of Chemical Education*, vol. 70, no. 2, pp. 195-197.

Zoller, U. (1997) The traditional-to-innovative switch in college science teaching: an illustrative case study on the reform trail, in*:* M. Caprio (ed) *From Traditional Approaches Toward Innovation*. Society of College Science Teachers' Monograph Series, pp. 3-10.

Zoller, U. (1999) Teaching tomorrow's college science courses – are we getting it right?, *Journal of College Science Teaching*, vol. 29, no. 6, pp. 409-414.

Zoller, U. (2001) Alternative assessment as (critical) means of facilitating HOCS: promoting teaching and learning in chemistry education, *Chemical Education Research and Practice in Europe*, vol. 2, no. 1, pp. 9-17.

Zoller, U. (2003) HOCS problem solving vs. exercise solving: what do college science students prefer?, in*:* D. Psillos, P. Kariotoglou, V. Tselfes, E. Hatzikraniotis, G. Fassoulopoulos & M. Kallery (eds) *Science Education Research in the Knowledge-Based Society*, Dordrecht: Kluwer Academic Publications, pp. 201-207.

Zoller, U., Fastow, M. & Lubezky, A. (1997) Student self-assessment in chemistry examinations requiring higher- and lower-order cognitive skills, *Journal of Chemical Education*, vol. 76, no. 1, pp. 112-113.

Keywords

Cognitive skills, teaching strategies, assessment strategies

Biography

Uri Zoller is Professor of Chemistry and Science Education in the Faculty of Science and Science Education – Chemistry, at the University of Haifa – Oranim, in Israel. His areas of scientific research, teaching and national and international activities are: Science Education: Science, technology and environment in the social context; teaching, learning and assessment of higher-order cognitive skills/critical thinking; Environmental Chemistry: Surfactants and PAHs in surface and groundwater; and Organic Chemistry: Synthesis and chemistry of strained, small rings containing sulphur. He has published 6 books and over 200 publications.

Book Reviews

A Handbook for Deterring Plagiarism in Higher Education

Jude Carroll

Oxford: Oxford Centre for Staff and Learning Development, 2002, 96pp. ISBN: 1-873-57656-0 (pbk)

There can be very few colleagues currently teaching in higher education establishments who have not had any brushes with the extremely tricky subject of plagiarism. Even if it has not happened yet on a course you are teaching (well, actually, as this book points out, it may have and you did not detect it), you may well have experienced it in your department, either by it having happened to a colleague, or by taking part in a plagiarism disciplinary committee or appeal hearing. One way or another, it seems, there is no escaping it, even if you teach a subject, as I do, which is practical in nature and where the lack of written resources makes plagiarism more difficult than many theoretical subjects. So the timeliness of this book is indisputable.

The book encourages a certain pick and mix attitude to the subject and the chapter headings, which start at 'Reviewing the Issues' and end with 'Institutional Policy and Culture', passing through 'Designing Courses for Deterring Plagiarism' to 'Detection' followed by 'Punishment', certainly cover all the possible angles to the question. The book is laid out text book style. Little boxes with bullet points abound; supporting statements (almost journalistic in style) occupy the margin. There are even exercises to complete with answers at the back of the book. It is all admirably clear and informative.

So does it do the job the title says it sets out to do, i.e. does it tell you how to deter plagiarism amongst your students? Well, yes, it does. But, like most things, it is not as simple as that. The core of the problem it seems is that it demands a tremendous amount of work detecting plagiarism and many of us just cannot face what is involved or do not have the time to do it. So, to my mind, the real focus of the book is not really about what you can do to avoid plagiarism in the first place or how to detect it, but an ever so slightly hectoring appeal (supported with a lot of evidence) to us all to get our act together and develop policies and strategies for dealing with it. The present situation seems fairly chaotic without any consistent policies across institutions. The book paints a fairly grim picture of current practice,

which I saw absolutely no reason to disbelieve. It is maybe not the message the reader who chooses to turn to this book wants to hear, but it is undeniable that the status quo cannot be justified in this respect. The book is first and foremost a call to arms. Whether the weary troops will heed the battle cry remains to be seen. If they do not, the book suggests, catastrophe may be not too far away. The book is certainly persuasive on this point.

It is extremely hard to write a book about this subject and make it relevant to so many disparate academic disciplines and I think the book scores very highly in this regard. You may not be able to use all the suggested strategies, but at least some of them will be helpful, no matter what your subject area. You may of course be already doing a lot of it, purely as a matter of good pedagogical practice (asking for essay outlines, changing essay questions year by year, making questions and projects as up to date as possible), but you may not have considered how useful these strategies are as a deterrent to plagiarism. Some of us have yet to dip our toes into the murky waters of peer-assessment and this book reinforces the lessons learned at the recent GWAMP[1] conference, which supported the notion that you do have to change the way you teach when you are dealing with large numbers, and that it can be productive and helpful to do so.

Another very helpful aspect of the book is that it is not at all narrowly focused. We may think of plagiarism largely in terms of non attribution of sources culled either from books or increasingly from the Internet, not to mention the awful spectre of 'guns for hire' writing essays for cash or 'paper mills' full of past essays which you are urged to change just a little bit and present as your own work, but of course there are many other aspects to it. Collusion (especially problematic in group work) is dealt with at length and the author does not duck the difficult questions it raises. Nor does she avoid the extremely touchy subject of international students and the fact that they appear to be very over represented in the plagiarism statistics. She makes the rather chilling point that this may simply be because their plagiarism is easier to detect. Food for thought there. She gives us a case study on informing students about collusion which reproduces a handout from a UK postgraduate course. The handout pulls no punches. It tells students that it is not acceptable to ask someone outside the course to read and correct written work they intend to submit as their own, even if the corrections are *only confined to the English language components of the work*[2]. It certainly gave me something to think about, if only the notion that we fudge quite a lot of this with regard to students who do not have English as their first language and that giving out a draconian handout may make things clearer, but will it improve the work submitted? Or just up the drop out and failure rate?

I came away from reading the book feeling a bit depressed rather than empowered to go out and fight my corner on this. The book certainly convinced me that the problem is bigger than most of us think it is and that the currently envisaged solutions do not really work very well. I think most of us are hoping rather pathetically that simply putting everything through some sort of electronic gateway will magically make plagiarism disappear. It is clear that so far, the electronic plagiarism detectors are pretty rough and ready and not all that reliable. The main hope seems to lie in redesigning courses and assessments so that plagiarism is rendered more difficult. There is also the task of how we get the message across to the students. There is no clear answer to this, of course. The book seems to suggest that we simply have to do it by every means possible, but the author acknowledges it is an extremely uphill struggle. Students plagiarize for a multitude of reasons and the gulf between us is never larger perhaps than on this issue. All credit to the author of this book therefore that she does not duck any of the difficult issues raised, including the issue of academics themselves plagiarising.

If you are in any way involved in teaching students, this book will provide you with plenty to think about and it is unlikely that you will not find ways to improve and alter what you do after reading it. The main point is that there is no quick fix to this undoubtedly increasing problem, and doing something about it represents an enormous amount of work facing those of us working in the sector.

Susanna Capon

Royal Holloway, University of London, UK

Notes

1 GWAMP (Group Working Assessment in Media Production). A HEFCE-funded project which ran from 2000 to 2003. There's a very useful website at: http://www.gwamp.bournemouth.ac.uk/

2 op.cit p.44. Italics not in the original text.

Assessment, Learning and Employability

Peter Knight & Mantz Yorke

London: Society for Research into Higher Education & Open University Press, 2003, 224pp.
ISBN: 0-335-21228-X (pbk)

As a distinguished contributor noted at a recent conference on e-learning, 'in addressing the challenges of learning and teaching, academics have become too preoccupied with the *how*, to the neglect of the *why*'. It is therefore particularly refreshing to review a book that addresses both, and does it well.

There is no doubt that 'employability' is the critical pedagogic leitmotif and there is no shortage of copy addressing this agenda. However, as far as I'm aware, this is the first book to undertake a concerted exploration of the policy and practical implications of addressing the employability agenda for the curriculum and, in particular, for assessment. The determination of the appropriateness of assessment practices is placed firmly in the context of a consideration of assessment purposes and assessment theory.

The authors set their exploration in the context of what some might regard as the antipathetic relationship between employability and traditional academic values. They argue strongly and convincingly, in my view, that there is no necessary opposition between these two goals. They exemplify this in the text itself with a rigorous, scholarly exploration of the policy and intellectual issues, applied in a professional way to a thorough consideration of the implications of employability for the instruments of practice.

The preface and introductory chapter provide a succinct and informative (yet familiar) overview of the increased demands faced by higher education, the role of higher education in the economy and a rationale for the focus on skills and employability. Models for an 'employability-aware curriculum' are reviewed and the challenges that enhancing employability places, particularly on modular schemes, are considered. The implications for summative assessment are identified and it is suggested that the concept of employability focuses the spotlight on the inherent problems of assessment. If higher education is expected to deliver a wider range of achievements, then how can we ensure that assessment meets this greater challenge and supports learning?

A review demonstrates clearly the fragility and limitations of current higher education assessment practices, highlighting the unreliability of assessment as grading, and raising serious questions about our over-reliance on summative assessment. It seems unsurprising, on the basis of the evidence presented, that teaching, learning and assessment gives the Quality Assurance Agency for Higher Education (QAA) some cause for concern! Add to this the increased complexity that developing employability necessitates, and we place further pressure on what might already be deemed a process at risk. This is not wholly a counsel of despair, however; the authors identify a critically-informed but nonetheless sensible and practical way in which programme teams can help to ensure a more intelligent application of assessment. Indeed, one of the merits of this text is its abiding concern for the practical payoffs of its theorizing.

The authors cogently argue that the potential of formative assessment, in promoting the kinds of complex learning and development associated with employability, is often unrealized. They suggest that 'structural threats' and resource constraints mean that formative assessment is often (and has been) a casualty of the changes placed on the sector. The clear message of the book is that if we take seriously the implications of employability, and indeed traditional academic values, we need to realize the potential of formative assessment. We also need to undertake a more thorough review of our assessment practices at programme and institutional level, in terms of their purpose and their research base. In this respect the book is both a call for action and a prompt as to how one might set about tackling this critical agenda.

In repositioning and opening up the problematic topic of assessment in the context of the employability agenda, the authors deploy three organising principles:

1. assessment must be recognized as involving judgement rather than measurement

2. the tension between formative and summative assessment must be reduced

3. a systemic approach to assessment at the programme-level, with an emphasis on agreeing goals and helping others to achieve them, is needed to increase reliability.

The authors draw on an impressive range of research to facilitate and enrich the application of these principles, in order to consider which methods should be used. They provide a helpful summary of a diverse range of methods and suggest how one might assess the complex achievements associated with employability. They also

explore the implications of the concept of authenticity, identify ways in which reliability might be improved and examine how formative assessment might be more effectively used. In addition, they address the implications of progression in terms of 'slow learning', introduce the concept of assessment as 'claims making' and the more radical notion of students making their own claims. Finally, they explore the implications for assessment systems at academic department and institutional level.

Key themes emerge across these chapters, illustrating how assessment can corrode complex learning and highlighting the difficulties of assessing 'fuzzy' contexts. It is repeatedly emphasized that 'low stakes' assessment (formative) would be more effective for learner development than the 'high stakes' summative assessment, as the former lowers the risk for the learner and enhances the opportunity for learning and development.

The chapters provide a stimulating, professional and engaging resource for the practitioner and policy maker alike. The educational issues are coherently teased out and the authors provide us with an accessible means of working through, in a more rigorous, self-conscious and sustainable way, the implications of the choices we make in deploying our assessment methods.

A fascinating, but in my view under-developed, section in chapter three explores the role of higher education in developing 'moral action'. The authors acknowledge that this is 'contentious territory', but I was disappointed that they didn't explore this aspect more fully. The spate of recent corporate scandals, e.g. Parmalat, and growing unease about Third World Debt and environmental issues, underline the importance of values in vocational education and business. A consideration of moral action could also open up the prospect of developing a curriculum and assessment that would challenge received notions of employability and could promote a more critical alignment between academic values and employability. Barnett's concept of 'critical being' is, I feel, particularly relevant here, as is his recent work on 'pernicious ideologies' (Barnett, 1997; 2002). This comment is not meant to detract from the overall merit of the authors' thesis, as taking 'moral action' seriously would, I'm convinced, underline even more strongly the importance of 'low stakes' formative assessment.

In drawing the book to a close, the authors summarize their main arguments. Higher education might be more complex but assessment problems are not new and we need to focus on them in a more scholarly way in order to develop theory. We also need to understand more about the relationship between assessment and learning. Linked to this, we need to review how we measure employability and seek to develop curricula that maximize opportunities for students to develop. It will

be important to understand how students develop self-efficacy and meta-cognition, as well as knowledge and skills. If learning is complex, then assessment systems need to be equally complex and capable of addressing 'fuzzy' learning, remembering that assessment is not an exact science but a process that inevitably involves judgment.

The tension between formative and summative assessment makes the teacher's role difficult; the former requires facilitative engagement with the learner that is in conflict with the distance and detachment required by the latter. The authors suggest that this tension might be reduced if we abandoned our commitment to determining overall grades and reduced our concern with the exactness of grading. The learning process would need to include approaches to help the students make 'claim achievements', supported by evidence as well as certification. Implied in this approach would be a stronger emphasis on the facilitation of learning, greater value placed on 'soft' achievements and a curriculum that acknowledges the different starting points of the learner. There would be less concern with summative grades and broad acceptance of ungraded awards. In principle, the argument is logical and powerfully presented. However, getting employers and academics to 'buy in' to this radically different approach will present a challenge!

The book finishes by re-emphasizing the importance of a 'systemic' approach to assessment at the programme level, the implications for developing the curriculum and the importance of leadership in facilitating change.

I particularly enjoyed this book. There is no doubt that it will give the reader plenty to think about — whether individual lecturer, curriculum designer, programme leader, or senior academic. As I read, I found myself questioning my own personal and institutional practice, coming up with ideas for change and identifying others who should read this book. Some of us might be further down the road than others in developing assessment techniques to address employability, but there is much for all to learn. Knight & Yorke provide a well-argued position that will serve to raise questions, inspire research and bring about change.

Christine Shiel

Bournemouth University, UK

References

Barnett, R. (1997) *Higher Education: A Critical Business*, Buckingham: SRHE & Open University Press.

Barnett, R. (2002) *Beyond All Reason: Living with ideology in the University*, Buckingham: SRHE & Open University Press.

How to Survive Your Viva: defending a thesis in an oral examination

Rowena Murray

Maidenhead and Philadelphia: Open University Press, 2003, 160 pp. ISBN 0-335-21284-0 (pbk)

The title is the least satisfactory part of this book; its reference to 'surviving' is presumably a marketing ploy that plays on students' anxieties about the supposed gladiatorial nature of vivas — a perception that Rowena Murray acknowledges but does not over-play in the text itself. Before receiving the book for review, I had expected it would focus upon the time after the viva — see the Oxford English Dictionary definition of 'to survive' as 'to continue to live after the end or cessation of some thing or condition or the occurrence of some event: to remain alive, live on' — yet the book barely touches on this. The subtitle tells the real story and the title does the book no favours. This is a pity because it is a worthwhile contribution, both practical and analytical, to a developing literature on this particular form of assessment.

The book attempts to hold in balance the undeniable truth that the viva is, viewed from different points in time, both extremely important (in prospect and while it is happening) and of almost no account (the moment it is over). As the author observes, who ever asks you how you did in the viva once you are successfully through it?

The book is organized into ten chapters that cover pretty much all that students might need to know about the viva, including aspects of it that they might never have considered. It leaves little to chance. Each chapter is divided into very small parts with plentiful subheadings, while advice is presented in boxes throughout the chapters. I found this busy and distracting, but others may like its sound-bite format.

The introduction sets out the author's rationale for writing the book. She claims the book is relevant for all three protagonists in the examining of research degrees — candidates, supervisors and examiners. In practice it is primarily addressed to research degree students, with some explicit reference to supervisors and only occasionally to examiners, who are left to work out how the research, practical advice and skills practice best apply to them. It is conceived as an extension of her existing book *How to Write a Thesis* (Murray, 2002).

Chapter 2 looks at the nature of the viva and the ways in which this can vary by disciplinary subject and institution. It is a useful reminder to candidates and external examiners that there is no universal practice and that they both need to prepare for the experience, albeit in different ways. Chapter 3, 'Roles and Responsibilities', addresses who will be at the viva, what the student might expect of them and the kinds of questions a candidate might fruitfully ask about the people and their roles. 'Countdown to the Viva' (Chapter 4) sets out a 3-month schedule of preparation for the viva, with an assumption that the candidate will have this amount of time between submitting and being examined. Given that Murray is careful throughout the text to underline how custom and practice vary, this strikes me as an unwarranted assumption and inconsistent with some regulations that would require a shorter period between submission and viva. However, she needs such an assumption if the advice on preparation that is contained in the following chapters is to be viable.

Chapters 5, 6 and 7, 'Questions', 'Answers' and 'Interactions', respectively give brief accounts of the different kinds of questions and answers that might be encountered or expected during a viva, together with some suggested strategies for focused preparation — for example, drawing up a list of general and specific questions that might be asked about the thesis. These chapters also raise questions about what the candidate might do if they encounter inappropriate, hostile or apparently ignorant questions from the examiners, or if they experience a lack of feedback for their answers to examiners' questions.

Chapter 8 urges the candidate to undertake graduated practice, from re-reading the thesis through to a mock viva and practising the set of questions prepared as a result of the previous chapters. Chapter 9 touches on the possible decisions that can be made as a result of the viva, doing corrections and appeals against examiners' decisions. The final chapter is a very brief summary about the ambiguity of the viva and a reiteration of the author's view that facing it head on with time to practise skills is the best way to deal with that ambiguity.

Throughout, Rowena Murray is concerned to recognize that candidates may feel powerless, but that they can learn to be more powerful by making themselves knowledgeable about *local* regulations, procedures and customs, practising appropriate skills and preparing themselves as they would for any other form of assessment. She is clear that some things cannot be carefully predicted or controlled, but that the examination process can be made less precarious by conscientious research and intellectual application to the job in hand.

Murray's starting point, summarized towards the end of the book, is that: 'positioning the viva as a new type of communication event [for the student] helps to clarify the new expectations it brings and the new skills it calls for. It also repositions the student about to take their viva as — still — a learner' (p.142). Throughout, she urges that students should be willing, and helped, to prepare for the oral examination and that such preparation can and should be considered appropriate from the beginning of their project, not something that is left to the last moment. I would agree that leaving consideration of the issues she raises and practise of the skills she considers vital for this particular form of 'peer review' until the period around submission could be disastrous. Current students of mine and one recently graduated PhD whom I asked to read this book commented that the book highlighted things that could go wrong, that they felt worse rather than better after reading it and that they thought it might be discouraging because of its emphasis on so many new things to learn in such a short time. None of these observations is damning unless supervisor and candidate really do leave all this to the last few months.

Rowena Murray argues a strong case for research students to prepare systematically for the oral examination of their thesis, and an equally good case that commonly such preparation has tended to be sporadic, insufficiently rigorous or even considered unnecessary. While recognising that our knowledge of regulations, procedures and custom and practice in relation to this form of assessment is partial, she nevertheless argues that students and their supervisors need to be active, even proactive, in relation to the viva, so that they can take control where appropriate and reduce uncertainty where control is not possible. Underlying these arguments is reference to research and anecdotal evidence that in many respects the examination of research degrees can be a bit of a shambles — unpredictable, barely regulated and not seen through the lens of good professional practice. It is presumably in this sense of dealing with something that needs to be tamed that Murray would justify using such an emotive word as 'surviving' in the title of this book.

Mary Fuller

University of Gloucestershire, UK

Reference

Murray, R. (2002) *How to Write a Thesis*, Maidenhead: Open University Press.

Assessment: case studies, experience and practice from higher education

Peter Schwartz & Graham Webb (eds)
London: Kogan Page, 2002, 192pp. ISBN: 0-749-43623-9 (pbk)

What are the practical difficulties faced when introducing computer-based testing for the first time? Does the software work? Might the computer crash before saving the test results? Is the response time fast enough? Is it feasible to allow tests to be taken at any time during the term, given the human propensity to leave things to the last minute? These are issues addressed in the first case study on assessment in this book of case studies edited by Peter Schwartz and Graham Webb.

This book presents a selection of such case studies about various issues in assessment. Higher education practitioners from a variety of disciplines, practising in universities in the UK, USA, Australia and New Zealand have contributed. Despite the diversity of countries and disciplines, I found myself relating to many of the situations raised in the book, albeit perhaps experiencing slightly different manifestations. Though written by different people, each case study description has the same structure. The issues raised in a case study are described first, such as the issue of managing difficulties arising from replacing pencil and paper tests with computer-based tests for campus-wide assessment. The context of the case is then explained; for example the type of institution, the number of students and their backgrounds and the technology being used. The case is then described in detail with a reflective break whenever an action has to be taken or some decision made. At this point, questions are asked to invite the reader to think about what they would do in this circumstance, such as *What steps would you recommend taking to improve the performance of the computer-testing programme? What do you think was actually done?* Each case ends with a case reporter's discussion that examines the success of the assessment and analyses the lessons that have been learned from the experience. Thus the book encourages active participation by the reader and is ideal for stimulating discussion if used by a group.

The case studies are grouped into themes. The first case study was part of a section on using Information Technology to help cope with large classes. Other themes included reflective assessment techniques, perhaps using journals, portfolios or peer-assessment, the development of assessment that addresses the needs of the individual student, and general problems in implementing assessment. An interesting section looked at the practice of institution-wide assessment in the US. The aim of such assessment is to document and improve educational effectiveness. It may focus on general educational aims, such as critical thinking or information literacy, or on the actual discipline studied, with the aim of improving the curriculum or teaching. Such assessment appears to be closely related to what we would call quality assurance in a UK context.

I found this book interesting and useful for several reasons. The detailed description of the case studies enabled a more in-depth analysis than is often found in books that focus on the assessment strategies themselves. This book does not, nor does it claim to, give a comprehensive picture of the different types of assessment and how to use them effectively. Other books that capably fulfil this role include Knight (1995) and Brown *et al*. (1997). What this book of case studies did particularly well was to provide an appreciation of the wide range of factors that influence the success of a particular strategy. These include the personalities of the people involved, the culture and sensibilities of people who might be asked to alter their approach to teaching, the support or otherwise of management in an institution and the perceptiveness and open-mindedness of those involved when responding to certain situations — all these factors had an effect in different case studies. Some of the situations were messy, with people resigning at critical moments and objectives not being achieved after years of effort. Such all too common occurrences are not usually described in books about pedagogy that aim to give advice and ideas about assessment.

The encouragement to reflect was a particular strength of this book and hence would allow it to be used as a starting point for small group study and discussion about assessment. It would be a valuable text in an action learning group. I would imagine that this book would be useful for new lecturers in opening their eyes to the complexity and breadth of the problems that might arise in implementing what might appear to be a relatively simple assessment strategy such as computer-based assessment. For more experienced lecturers, the opportunity to compare situations in different higher education institutions and to see different approaches to assessment is invaluable. Despite the diversity of the case studies in discipline

and context, I invariably found something of relevance. For all those involved with assessment, the committed and thoughtful approach of all the case study reporters is an inspiration.

Vicky Bush

University of Gloucestershire, UK

References

Brown, G., Bull, J. & Pendlebury, M. (1997) *Assessing Students' Learning in Higher Education*, London: Routledge.

Knight, P. (1995) *Assessment for Learning in Higher Education*, London: Kogan Page.