enhancement
themes

Enhancing practice

Reflections on Assessment: Volume I

Preface

The approach to quality and standards in Scotland is enhancement-led and learner-centred. It has been developed through a partnership of the Scottish Higher Education Funding Council (SHEFC), Universities Scotland, the National Union of Students in Scotland (NUS Scotland) and the Quality Assurance Agency for Higher Education (QAA) Scotland. The enhancement themes are a key element of a five part framework which has been designed to provide an integrated approach to quality assurance and enhancement, supporting learners and staff at all levels in enhancing higher education in Scotland drawing on developing, innovative practice within the UK and internationally.

The five elements of the framework are:

- a comprehensive programme of subject level reviews undertaken by the higher education institutions themselves; guidance on internal reviews is published by SHEFC (www.shefc.ac.uk)

- enhancement-led institutional review (ELIR) run by QAA Scotland (www.qaa.ac.uk/reviews/ELIR)

- improved forms of public information about quality; guidance on the information to be published by higher education institutions is provided by SHEFC (www.shefc.ac.uk)

- a greater voice for students in institutional quality systems, supported by a national development service - student participation in quality scotland (sparqs) (www.sparqs.org.uk)

- a national programme of enhancement themes aimed at developing and sharing good practice to enhance the student learning experience, which is facilitated by QAA Scotland (www.enhancementthemes.ac.uk).

The topics for the themes are identified though consultation with the sector and implemented by steering committees whose members are drawn from the sector and the student body. The steering committees have the task of developing a programme of research and development activities, which draw from national and international good practice. Publications emerging from each theme are intended to provide important reference points for higher education institutions in the ongoing strategic enhancement of their teaching and learning provision. Full details of each theme, its Steering Committee, the range of research and development activities, and the outcomes are published on the enhancement themes website (www.enhancementthemes.ac.uk).

To further support the implementation and embedding of a quality enhancement culture within the sector, including taking forward the outcomes of the various enhancement themes, a new overarching committee has been established, chaired by Professor Kenneth Miller (Vice-Principal, University of Strathclyde). It will have the important dual role of keeping the five-year rolling plan of enhancement themes under review and ensuring that the themes are taken forward in ways that can best support institutional enhancement strategies. We very much hope that the new Committee, working with the individual topic-based themes' steering committees, will provide a powerful vehicle for the progression of the enhancement-led approach to quality and standards.

Norman Sharp
Director, QAA Scotland

Contents

Workshops 5-8 can be found in *Reflections on Assessment: Volume II*

Enhancing practice: assessment

A series of workshops and scoping studies provided a focus for much of the Scottish higher education's work on the enhancement theme of Assessment. Five key issues emerged, along with suggestions as to ways in which each might be tackled. Needless to say, there was an extent of overlap. The key issues were as follows.

1   The need to avoid over assessment and find ways to reduce the assessment load

The following alternatives might be helpful as means of addressing these issues:

- the use of alternative methods, such as self and peer-assessment
- a close inspection of assessment instruments to ensure that specified learning outcomes are only assessed once
- a substitution of summative assessments for more formative ones
- the application of long, thin modules that require end-of-session examinations rather than at the end of the first semester
- the use of 'synoptic' end of year tests, which have the additional advantage of reducing the tendency for students to see learning in bite-sized chunks.

2   The need to redress the balance between formative and summative tasks with the former to be increased at the expense of the latter

There was a widespread belief that a major step forward for assessment practices will be to provide more opportunities for students to learn from their mistakes through, for example:

- the progressive weighting of assignments so that at the start of a course the summative element is a relatively small proportion compared to the formative, to a situation at the end where the proportions are reversed
- the use of computer-aided assessment (CAA), which was seen as an unthreatening environment for students and one that can provide instant, high-quality feedback
- the application of self and peer-assessment, which are ideal for formative purposes
- the development of personal development planning (PDP), which offers an opportunity to embed the notion of an ongoing process, using a variety of sources as evidence, including formative assessment tasks.

3   The need to provide effective student feedback and develop methods for improving its quality

Inevitably, some of the proposals discussed for dealing with this issue overlapped with the first two outlined above. These included:

- various forms of CAA, which should include automated, instant feedback for incorrect answers
- personal response systems which can be deployed in lectures and other large-group teaching situations
- classroom assessment that can be incorporated during staff-student contact times

- self and peer-assessment might include feedback comments on the strengths and weaknesses of assignments.

4 The need to ensure that there is a better match between teaching, assessment and learning outcomes

Although the above principle, based on constructive alignment, is well known across the sector, there is sometimes a tendency to assess that which is easy to assess rather than the explicit intended learning outcomes. The use of portfolios and processes such as PDP may force a revision of positions, along with the use of a variety of tests and other tasks that reveal competence or capability, but which might be considered unconventional. These might include:

- the use of oral presentations and teamwork

- directly engaging students in the design and application of the ways they are to be assessed.

5 The need to develop and implement innovative assessment techniques

Innovative techniques can be used to reduce the assessment load or to switch from summative to formative tasks; they can offer better and quicker feedback and they can provide an improved match between teaching, assessment and learning outcomes. It is important, however, to remember that:

- efficient practice is not necessarily effective practice and vice versa

- a careful selection of methods is required to achieve an optimum balance

- what may be innovative in one discipline is well-established practice in others and so a rounded view must be taken.

More detailed information, along with possible ways of addressing each of these challenges, can be found later in this report including all the case studies and workshop directors' reports. They are available at www.enhancementthemes.ac.uk and in printed form.

General introduction

The enhancement themes initiative was launched in autumn 2003 as part of the unique, collaborative approach to quality that is being undertaken in Scotland. It forms one of the five pillars of the Quality Enhancement Framework that has been developed by a partnership of the Scottish Higher Education Funding Council (SHEFC), the Quality Assurance Agency for Higher Education (QAA) Scotland, Universities Scotland and the student body. The main aim of the enhancement themes is to support the sector in improving the student experience in Scottish higher education (HE) by focussing on certain areas identified by the partners as being in need of further development and enhancement. The first two themes were Assessment and Responding to Student Needs.

This paper summarises the main outcomes from the work on Assessment, including possible ways of addressing the challenges identified, as well as matters that merit further work and reflection. It is intended that this paper, in conjunction with the

individual workshop reports and case studies, will provide the sector, subject groups (eg HE Academy Subject Centres and Heads of Department groups), student and support services groups and appropriate professional and statutory bodies with a valuable resource that will enable the important issues raised to be handed over and taken forward, as appropriate.

The Assessment theme

In autumn 2003, a steering committee drawn from across the Scottish HE sector, including students, as well as representatives from the HE Academy and officers from SHEFC, QAA Scotland and Universities Scotland, was established under the chairmanship of Professor Simon van Heyningen from the University of Edinburgh. The membership is listed in Annex A.

The Steering Committee was charged with developing and implementing a sector-wide programme of work on Assessment. Following discussion within the Steering Committee and consultation across the sector, a number of sub-topics were identified as priority areas.

One major topic which the Steering Committee commissioned work on was reviewing the UK honours degree classification system and investigating what other systems were adopted by other countries. The report, written by Jane Denholm (Critical Thinking), was followed up by a seminar in May 2004 to discuss these issues and finally a discussion paper was issued in October 2004, which summarised the Steering Committee's findings and raised various issues for further debate. The work of the Steering Committee has fed into, and been cited by, the recently published findings of the English Measuring and Recording Student Achievement Scoping Group[1].

The remaining topics were, it was felt, sufficiently self-contained, yet of universal interest across the sector, to justify running a series of eight workshops from January to June 2004. Each workshop was jointly organised and facilitated by QAA Scotland and a director, drawn from across the Scottish HE sector. The workshop sub-topics are listed in Annex B.

The workshops themselves were deliberately structured to involve and engage colleagues, as well as providing high-level input from renowned thinkers and writers on assessment. As a result of the format, it was possible not only to benchmark practice in Scotland against international standards but also to establish networks across institutions and allay any feeling of isolation among assessment pioneers.

Although each of the workshops had a particular focus, it was soon apparent that there were common issues that applied to all. Another, rather more pleasing thread, was the

---

[1] In October 2003, Universities UK, the Standing Conference of Principals and the Higher Education Funding Council for England set up a group, chaired by Professor Bob Burgess, to review the recommendations from the English Higher Education White Paper relating specifically to recording student achievement issues. The findings were published in November 2004 and, at the time of writing (May 2005) a UK-wide group, also chaired by Professor Bob Burgess, has been set-up to investigate these issues further.

clear evidence that not only are large numbers of academics across Scotland confronting these challenges in order to improve the quality of teaching and learning, but also, in many areas, domestic universities are at least on a par with best practice worldwide.

The section that follows outlines the five key issues that emerged from discussions between practitioners at the one-day workshops and some possible ways in which each may be addressed. Inevitably they overlap to some extent and readers will recognise both the totality of the challenge in improving assessment in HE and the applicability of solutions.

Key issues

Avoiding over assessment and finding ways to reduce the assessment load

It is accepted that a number of factors have conspired to increase the number of assessment tasks students are required to complete. With some exceptions, such as medicine, there has been large-scale modularisation of the HE curriculum. One of the effects of this change has been the fractionation of learning into bite-sized chunks, which for students at least, is one of its great attractions, because it appears to define 'learning limits'. Quite apart from academic considerations, however, the downside is that at least the same level of assessment has to be undertaken now which previously would have occurred over a much longer period. The result is a significant increase in assessment load for students and marking load for staff.

The workshops provided some alternatives that might be applied in order to address these issues. They include the use of alternative assessment methods, such as self and peer-assessment; a close inspection of assessment instruments to ensure that specified learning outcomes are only assessed once; and a substitution of summative assessments for more formative ones.

Structural solutions are also possible, such as long, thin modules. These can be especially effective for first year students, because they require end-of-session examinations rather than at the end of the first semester. Another possibility, along similar lines, is the use of 'synoptic' end of year tests, which have the additional advantage of reducing the tendency for students to see learning in bite-sized chunks.

Redressing the balance between formative and summative tasks with the former being increased at the expense of the latter

There is widespread acceptance of the educational value of formative assessment as a way of improving learning. However, the reality is that on the one hand students tend not to value anything that 'does not count', and on the other that modularity and other structural factors have conspired to squeeze out formative tasks in favour of summative ones.

The provision of predominantly summative assessments clearly directs and motivates students' approaches to learning. This is because students see summative assessment as 'high stakes' or 'high risk'. To them, their future employment or further study opportunities may depend upon the marks or grades obtained in summative tests. Thus, even if there is space within one module for formative tasks, if a student is

required to work towards summative assignments in another, the formative will invariably be neglected. Therefore, if assessment practices are to improve, a major step forward will be to acknowledge the need for changes in attitude and behaviour by both staff and students, and in particular to provide more opportunities for students to learn from their mistakes. This means engaging more in formative (assessment) learning tasks.

Strictly formative assessment provides non-threatening, 'low stakes' assessment practice, often rich in feedback (to correct misunderstandings) and feedforward (suggestions for improvement). Furthermore, because students know that their errors will not count against them when a final grade is being assigned to the module or unit of learning, they are more likely to be open and unafraid of making mistakes.

The tension between formative and summative assessment has been exacerbated more recently because modularisation has forced more summative assignments to be scheduled over shorter learning periods, often depriving students of opportunities for learning from previous assessments. Furthermore, in order to achieve 'efficiency', continuous assessments or coursework assignments, which may in the past have been used formatively, are now used for summative purposes.

The possibly damaging effects of mixing formative and summative assessments may be partially offset by the creative use of progressive 'weighting' of such assignments that nevertheless still 'count'. This may still encourage students to adopt strategic approaches to maximise their marks, but the 'fear of failure' will necessarily be reduced, especially at the early stages, thereby enabling formative feedback to be fed into the educational process.

Other solutions to the formative/summative issue are inevitably the same as those suggested for over assessment. Computer-aided assessment, for example, is seen as an unthreatening environment for students and one that can provide instant, high-quality feedback. Self and peer-assessment are sometimes seen as too bold to be used in a summative context, but are ideal for formative purposes. Personal development planning offers an opportunity to embed the notion of an ongoing process, using a variety of sources as evidence, including formative assessment tasks.

Providing effective student feedback and developing methods for improving its quality

One of the major impacts that a shift from summative to formative assessment may have, will be the need to provide more prompt, helpful and timely feedback. Summative tasks are often conventional examination papers that require careful and time-consuming marking, perhaps by several people, hence the speed of feedback can be very poor and may focus the student's mind almost exclusively on the mark, rather than any developmental comment. In addition, it is often the case that students, who are only interested in the mark or grade, frequently ignore feedback or feedforward comments on summative assignments. To counter this tendency students can be supplied with written feedback on summative assignments and the mark or grade withheld until after a period of reflection. However, there is a need to progress beyond such sleights of hand if there is to be a fundamental change to formative learning and greater use by students of feedback on their work, be it written or verbal, given by tutors or indeed, by their peers.

Innovative, research-informed and efficient ways of providing fast, quality feedback do exist eg various forms of CAA; personal response systems; classroom assessment and self and peer-assessment. Specific use of information technology in online assessment emphasises not only the increased efficiency of using computer-based assessment, but also the provision of automated feedback for incorrect answers. Such CAA - with feedback for learning - could be used more freely, including deploying virtual learning environment quiz tools, which research suggests students find increasingly engaging. Such tools can utilise question databanks for formative purposes, and, as an incentive to learners, the same database can also be drawn on for use in summative assessments.

Furthermore, immediate (personal) response systems can be deployed in lectures and other large-group teaching situations to check knowledge and understanding instantly, as well as, for the lecturer, providing feedback on the effectiveness of their own teaching. 'Classroom' formative assessment could be incorporated during staff-student contact times, so that 'required' problem solving based tasks or short notice mini-presentations are designed into the teaching session, with feedback from tutors and peers provided instantly.

Student collaboration and sharing should be promoted. Peer marking, including feedback comments on strengths and weaknesses of assignments, either on drafts or even final submissions, should assist learning for an eventual summative assessment or at least provide feedforward for the next one. Openness to this type of approach may reduce the incidence of apparent or covert student collusion.

Ensuring a better match between teaching, assessment and learning outcomes

Research shows that correct curriculum alignment can change surface to more lasting, deep learning. The critical factor is the correct alignment between teaching, assessment and learning outcomes. However, while both the phrase and the principle of constructive alignment are well known across the sector, there remains a suspicion that theory does not always match practice. For instance, there is sometimes a tendency to assess that which is easy to assess rather than explicit learning outcomes. One way around this issue is to use a variety of tests and other tasks that reveal competence or capability, but which might be considered unconventional. Using oral presentations and teamwork as part of the teaching/learning process and then assessing the students on their performances fall into this category.

More unconventional still is the notion of directly engaging students in the design and application of the ways they are to be assessed. This is a practice that is not commonplace, and yet the importance of assessment to both parties surely demands some kind of dialogue. Although the explicit declaration of learning outcomes in a module or programme is likely to better define the most appropriate assessment task to determine whether students have acquired new knowledge or understanding or developed a particular skill or set of skills, there is the potential danger of being over-prescriptive and thus restricting the range of learning with which students might engage. There is still an issue of how students present longer term, implicit learning outcomes, including personal learning and learning how to learn. It may in the end be the use of portfolios and processes such as PDP that will force a revision of positions and attitudes and if that happens surely all will benefit.

Developing and implementing innovative assessment techniques

As with all the categories, this one has no unique boundaries. Innovative techniques can be used to reduce the assessment load or switch from summative to formative tasks; they can offer better and quicker feedback and they can provide a better match between teaching, assessment and learning outcomes. It is therefore self-evident that innovative assessment techniques offer both staff and students new possibilities to better judge both teaching and learning. However, the introduction of innovative assessment methods is not simply about choosing a different way of doing things or saving staff time, instead the driver must be because the particular innovation is best suited to what students are being asked to learn. It is important to remember that what is efficient may not be effective and vice versa and, therefore, that a careful selection of methods is required to achieve an optimum balance. It is also worth noting that what may be innovative in one discipline is well-established practice in others and so a rounded view must be taken.

From the case studies and discussion sessions at the workshops, there is ample evidence that new techniques are being tried and used successfully across the Scottish HE sector. There are those in the sector who argue that there are substantial institutional or managerial barriers to their introduction, and while this may be the case in some institutions (or departments within institutions), it is clearly not stifling innovation in others. It is important to point out, however, that perceived barriers to change are just as obstructive as real barriers.

Being innovative not only requires boldness, it also takes time, which includes not only that required to learn the new technique but also to evaluate, and if necessary, change it. In the end, however, this can be seen as an investment that will repay the extra effort. For example, writing challenging and appropriate multiple-choice questions for CAA is a lengthy process but it will be worthwhile, since the database will become a valuable resource for some years to come. Nevertheless, finding the time is not easy, as many delegates to the workshops explained. Competing tasks are often seen as being as either equally or indeed more important than innovations in teaching, learning and assessment. The most quoted example of such competition was the need to research, but managerial demands, larger classes containing students with ever widening abilities, were also part of the equation.

Conclusion

The workshops were widely praised by delegates and the sector generally. They brought together individuals who might otherwise not have met and forged informal links, not just between colleagues in different institutions, but also across subject boundaries. It was quickly apparent, given the huge demand for places at the workshops, that individuals do worry about assessment issues, mainly because they genuinely care about their students and want to do a professional job. They are also conscious of the need to match their in-depth theoretical subject knowledge with that centring on pedagogy.

Unsurprisingly perhaps, it was also quickly apparent that many assessment issues are common across the sector and it was stimulating to see how much innovation is currently being practised. Sharing practice was one of the major benefits of the workshop series.

The injection of external speakers was of considerable value, even if some of their contributions did little more than reassure the audiences that no great gaps exist between what is going on in Scotland and what is happening in the rest of the world. Refreshing too was the number of 'home grown' talents, who spoke knowledgeably and passionately about their assessment activities and demonstrated the practical benefits of engaging with assessment in this way.

That the Assessment workshops provided an excellent focus for one of the first enhancement themes cannot be questioned, however, the challenge now is what lies beyond. The creation of an active website is clearly one way, but more tangible objects, such as a journal or other published output, are likely to reach a wider audience, especially among technophobes. From 2005 onwards, the new enhancement methodology of a more flexible, five year plan supported by a systemic theme, as well as topic-based ones, offers the possibility of further in-depth work. Indeed, formative assessment has already been chosen as one of the new 'mini' themes for 2005.

What has already been achieved is the highlighting of a crucial area in teaching and learning and facilitating the encouragement and support of quality enhancement in this area across the sector. By prompting reflection, debate and action, there are already students across Scotland who are receiving direct benefit from this work and the workshops. There can be little doubt that in the future the difference to the student experience of many more will be positively enhanced as a result.

Annex A List of Steering Committee members

Professor Simon van Heyningen, University of Edinburgh (Chair)

Professor Richard Byrne, University of St Andrews

Mr Duncan Cockburn, student participation in quality scotland (sparqs)

Professor Bob Craik, Heriot-Watt University

Professor Morag Gray, Napier University

Mr Win Hornby, The Robert Gordon University

Dr Sarah Mann, University of Glasgow

Ms Alison Ryan, The Open University Students' Association

Professor Brenda Smith, The Higher Education Academy

QAA Scotland officers

Dr David Bottomley, Dr Alastair Robertson, Ms Elizabeth Anderson

Annex B List of Assessment workshops with associated web links and workshop directors

Workshop 1 Streamlining assessment - how to make assessment more efficient and more effective

www.enhancementthemes.ac.uk/defaultpage121c0.aspx?pageID=140
Workshop Director: Professor David Ross, University of Abertay, Dundee

Workshop 2 Using assessment to motivate learning

www.enhancementthemes.ac.uk/defaultpage121c0.aspx?pageID=141
Workshop Director: Dr Andrew Eadie, Glasgow Caledonian University

Workshop 3 Constructive alignment of learning outcomes to assessment methods

www.enhancementthemes.ac.uk/defaultpage121c0.aspx?pageID=142
Workshop Director: Professor Mike Osborne, University of Stirling

Workshop 4 Developing a variety of assessment methods, including self and peer-assessment

www.enhancementthemes.ac.uk/defaultpage121c0.aspx?pageID=143
Workshop Director: Professor David Lines, The Robert Gordon University

Workshop 5 Assessing online

www.enhancementthemes.ac.uk/defaultpage121c0.aspx?pageID=144
Workshop Director: Dr Hamish Macleod, University of Edinburgh

Workshop 6 Issues of validity, reliability and fairness

www.enhancementthemes.ac.uk/defaultpage121c0.aspx?pageID=145
Workshop Director: Ms Pamela Flanagan, Royal Scottish Academy of Music and Drama

Workshop 7 Improving feedback to students (link between formative and summative assessment)

www.enhancementthemes.ac.uk/defaultpage121c0.aspx?pageID=146
Workshop Director: Dr Bob Matthew, University of Glasgow

Workshop 8 Assessing personal transferable skills

www.enhancementthemes.ac.uk/defaultpage121c0.aspx?pageID=147
Workshop Director: Dr Colin Mason, University of St Andrews

Assessment workshop series - No 1

Streamlining assessment - how to make assessment more efficient and more effective

Streamlining assessment - how to make assessment more efficient and more effective - An overview

Professor David A Ross, CELT, University of Abertay, Dundee and Workshop Director

Introduction

The series of Assessment workshops, which are part of the enhancement themes engagements in Scottish higher education, comes at a timely juncture in the sector. Primarily, this is because academic staff and students face enormous pressures and challenges in today's higher education environment to survive. There is also a general consensus among writers on student assessment that the process is both fundamentally important (to staff and students) and very challenging, for example:

'(assessment) is a serious and often tragic enterprise' (Ramsden, 2003)

'there is more bad practice and ignorance of significant issues in the area of assessment than in any other aspect of higher education' (Boud, 1995), and

'The spirit and style of student assessment defines the de facto curriculum' (Rowntree, 1987).

Perhaps one of the most penetrating quotes that sums up the importance of assessment and why we must ensure it is as efficient and effective as possible is the much quoted one used by Win Hornby in his keynote address to the workshop:

'Students can, with difficulty, escape from the effects of poor teaching, they cannot (by definition if they want to graduate) escape the effects of poor assessment' (Boud, 1995).

David Boud, in a seminal paper on student assessment (Boud, 2000) brings the importance of effective assessment to a new height by advocating the concept of 'sustainable assessment' which he defines as '…assessment that meets the needs of the present and prepares students to meet their own future learning needs'. In other words, he gives equal level of importance to assessment as a tool for developing learning alongside its function of measuring performance.

As staff then, we owe it to our students to explore in-depth key issues in assessment, in order to make their learning experiences as effective and stimulating as possible. But we have to do this in the face of an expanding sector (so-called 'massification'), increasing resource pressure, the need to be more efficient...and so on! So - where do we start? For me, the approach has to be the one advocated by Mantz Yorke (Yorke, 2001) in which he proposes that leadership needs to be taken by senior management in the sector in assisting teaching staff to develop enhanced student assessment practice - Yorke encourages institutions (the sector) to take this forward based on the construction of responses to a number of strategic questions:

Does the institutional policy or mission imply that its general approach to assessment should be changed? And, if so, in what way(s)?

Are there any general institutional weaknesses in assessment (such as might have emerged from subject reviews or from internal reflection on practices and procedures, which need to be tackled?

Since assessment is, by general consent, the least well-understood and enacted aspect of curricula, what developmental activity needs to be instigated?

In dealing with the preceding questions, is best use being made of existing expertise, both 'in-house' and from outside. And if not, why not?

What, if anything, needs to be done to make the institutional system that surrounds assessment function effectively and be compliant with external expectations?

How does the institution keep abreast of developments in assessment both nationally and internationally?

How does the institution learn from its diverse experiences regarding assessment, and develop?

In my view, the answers to these questions will assist in both developing action steps from the workshop series and within individual institutions, through the ongoing development of their teaching, learning and assessment strategies.

So - why did we pick the issue of streamlining assessment for the first workshop? Well, mainly because as an issue, it serves to introduce the general concept of enhanced assessment design and practice, which is after all, what the series is really all about. It also allows us to introduce some of the key issues and dilemmas such as:

- the conflict between streamlining assessment to make student and staff tasks easier (in face of massification) and diversifying assessment techniques and practices to cope with changing higher education environment and student learning needs (with one important group in this being students with disabilities)

- the concept of stripping back the intended nature of any assessment assignment to it's fundamental relationship with learning outcomes - this is allied to the constructive alignment models currently in vogue as proposed by writers such as John Biggs (Biggs, 2003) and Paul Ramsden (Ramsden, 2003)

- pressure to increase the use of computer-based assessments - some of this pressure is genuinely because there is a growing awareness of the potential usefulness of such techniques, through for example, innovative practices as demonstrated in today's case-studies and through literature examples and projects in Scotland such as the Student Enhanced Learning through Effective Feedback Project

- the current drive to make all aspects of teaching and assessment practice in the sector, more efficient and effective on both cost grounds and to give students a better learning experience (another dilemma!)

- the growing incidences of plagiarism in student assessment, particularly as the use of electronic sources of information and the use of communication and information technology as a medium for writing coursework becomes more prevalent

- the concept of developing more effective means of assessment to allow measurement of employability skills in the curriculum (Yorke, 2003).

Implicit in this is the need for us as staff to be aware of current practice and innovations in the sector and to be prepared to accept change (or at least to have

a dabble!) and challenge our own views. For example, as teachers, we need to make explicit connections between the skills we expect students to bring with them and traditional modes of assessment - if a particular skill or way of thinking underlies the proper solution to a problem, we should try to find ways to let the student assess this skill beforehand, rather than letting them stumble up against it when trying to solve a larger problem. It follows therefore, that as studentss needs and existing skills diversify, then just keeping up with this issue gets more challenging!

References

Biggs J (2003) *Teaching for Quality Learning at University*, second edition, Society for Research into Higher Education/Open University Press, Maidenhead

Boud D (1995) Assessment and learning: contradictory or complementary? in P Knight (ed) *Assessment for Learning in Higher Education*, Kogan Page, London

Boud D (2000) *Studies in Continuing Education*, 22, 2, 151-167

Ramsden P (2003) *Learning to teach in higher education*, second edition, Routledge, London and New York

Rowntree D (1977) *Assessing students: How shall we know them?* Kogan Page, London

Student Enhanced Learning through Effective Feedback project
www.heacademy.ac.uk/805.htm

Yorke M (2001) *Assessment - A Guide for Senior Management*
www.heacademy.ac.uk/resources.asp?process=full_record&section=generic&id=1

Yorke M and Knight P (2003) *The Undergraduate Curriculum and Employability*
www.heacademy.ac.uk/resources.asp?process=full_record&section=generic&id=248

Dogs, stars, Rolls Royces and old double-decker buses: efficiency and effectiveness in assessment

Win Hornby, Aberdeen Business School, The Robert Gordon University

Abstract

Although the environment within which assessment is now conducted has changed significantly in recent years, the way we assess students has not. As the number of students increases and the unit of resource is stretched, academic staff are finding that the pressure on traditional forms of assessments is mounting. Consequently, a number of unintended consequences arise as a result of the coping strategies that are frequently adopted by both staff and students. These may be summarised as:

- slow feedback to students
- little meaningful feedback
- almost no formative assessment by which students can learn from their mistakes
- learning outcomes assessed several times with little or no rationale for this
- no correlation between credit points and student workload on assignments
- no correlation between credit points and staff workload on assignments
- no mechanisms for coordinating assessment between modules resulting in students being assessed in the same way for almost all their modules
- bunching of assessments which results in students opting out of classes or assessments.

This paper outlines a portfolio analysis approach for evaluating assessment methods and outlines a number of possible strategies, discussed in the research literature, for streamlining assessment. These can be summarised as:

1 strategic reduction of summative assessment
2 front-end loading
3 in-class assessment
4 self and peer-assessment
5 group assessment
6 automated assessment.

Each strategy is examined and examples are given of how these might be implemented.

The paper concludes with a question for all of us:

The next time you sit down to mark 250 case studies, essays, laboratory reports or examination scripts ask your self one question. Can I do this differently in a way which makes my assessment more efficient and more effective?

Introduction[1]

'Students can with difficulty escape from the effects of poor teaching...they cannot (by definition if they wish to graduate) escape the effects of poor assessment' (Boud, 1995).

This paper will attempt to do four things. It will:

- look briefly at the changing environment within which assessment now takes place

- introduce a framework for evaluating efficiency and effectiveness of assesment

- refer to some empirical data based on a survey of assessment practice in my own university[2], and finally

- propose a number of strategies for streamlining assessment.

The changing environment

There are probably few things that academic staff do which are more important than assessing students. Nor as Boud (1995) indicates in the quotation above, is it something that students can easily avoid! In addition, it takes up a very significant amount of staff time. It has been calculated, based on a survey of the time spent by staff assessing students at one university, that if one takes all the assessments undertaken by students in one year at an average size university in Scotland, and asked one member of staff, working from 9.00am until 5.00pm every day for 365 days a year, to undertake all the assessment, it would take a total of 35 years to complete all the marking! (Hornby, 2003)

Nor is the task of marking becoming any easier. As we know from the literature, assessment has four main functions. These are:

- formative, to provide support for future learning

- summative, to provide information about performance at the end of a course

- certification, selecting by means of qualification, and

- evaluative, a means by which stakeholders can judge the effectiveness of the system as a whole.

Not all of these objectives are viewed as having equal importance by the increasing number of stakeholders in the process. Students, regarded by some as customers, are becoming more demanding; senior managers in universities are concerned about retention rates and will view assessment as a means of measuring a course's success in meeting its retention targets; external examiners and professional bodies are also key

---

[1] This paper is based upon two pieces of work, which were carried out during my secondment to the Centre for the Enhancement of Learning and Teaching at The Robert Gordon University. For further details see: Hornby W (2003) *Case Studies on Streamlining Assessment*, Centre for the Enhancement of Learning and Teaching, The Robert Gordon University, Aberdeen and Hornby W and Laing D (2003) *Efficiency and Effectiveness in Assessment*, Centre for the Enhancement of Learning and Teaching, The Robert Gordon University, Aberdeen, available at www.rgu.ac.uk/celt/quality/page.cfm?pge=7347#cases [Accessed 7 March 2004]

[2] The data is based on a survey from The Robert Gordon University of academic staff views on assessment. 168 staff from a total of 538 responded to a questionnaire on assessment between May and June 2003. An analysis of response rates by department, school and faculty revealed that the sample was representative of the University as a whole. For details see Hornby and Laing (2003) op cit.

players in this process, as well as employers, the media and other external bodies. The government and the quality assurance agencies also play an increasing role and have established guidelines on assessment, subject benchmark standards as well as qualification frameworks, all of which have an impact on the assessment process. The demands on staff have also significantly increased as a result of more student numbers. In spite of a significant increase in participation rates in higher education, (from about five per cent of the eligible population in the 1960s to the present rate of about 35 per cent with a national target of 50 per cent by 2010,) the unit of resource has significantly declined by 38 per cent and the average staff-student ratio across the sector of 17:1 in 2001, has increased by of over 40 per cent on a decade ago (MacLeod, 2001)[3].

Furthermore, the design and shape of a lot of university degree courses (the course architecture) has changed significantly in the past decade. Courses are now semesterised and unitised and credit-rated, and have programme specifications. Modules are designed around student learning outcomes. Therefore, staff are being put under increasing pressure to deliver an effective and efficient way of assessing students. That pressure will undoubtedly increase as other priorities crowd in on staff such as publishing RAE-rated papers, seeking alternative non-governmental funding sources or writing and developing new courses and new delivery methods. While some commentators may deplore these facts, this is nevertheless the reality for the great majority of academics in higher education today.

Assessment practices today

Given this changing environment, how then have our assessment practices changed? What do we do differently now from what we were doing say ten or even five years ago? According to Rust (2000):

> …although the prevailing orthodoxy in UK higher education is now to describe all courses in terms of learning outcomes, assessment systems have not changed. Under current arrangements, rather than students having to satisfactorily demonstrate each outcome (which is surely what should logically be the case…) marking still tends to be more subjective with the aggregation of positive and negative aspects of the work resulting in many cases in fairly meaningless marks being awarded with 40 per cent still being sufficient to pass.

In addition, a survey of assessment practices in UK universities in 1999 found that there was not much evidence of innovative assessment with 90 per cent of the assessment of a typical British degree consisting almost exclusively of examinations plus tutor marked reports or essays (Brown and Glasner, 1999). While evidence from a survey of assessment practices in my own university suggests that there has been a some shift away from this model in recent years (Hornby and Laing, 2003), the overall conclusion seems to be that by and large not much has changed.

---

[3] Even a top up fee of £3,000 will still not bring the unit of resource to the level it was in the 1990s according to evidence from UK universities quoted in *The Economist,* 6 March 2003.

Effectiveness and efficiency

How then can we encourage staff to become more effective in the way they assess students but at the same time be efficient? And what do we mean by efficiency and effectiveness in this context? In evaluating various assessment methods some key questions need to be addressed. These can be listed below under two broad sets of criteria:

a   educational effectiveness, and

b   management and resources efficiency.

Educational effectiveness

When evaluating assessment methods there are a numbers of questions that come under this heading.

- What are the strengths and weaknesses of each assessment method?

- To what extent are the methods used educationally valid?

- To what extent are the assessment methods used closely linked with desired skills and competences?

- Are the assessment methods 'constructively aligned' to the stated outcomes to use Biggs (1997) phrase?

- Does the assessment method match the task and outcomes?

- Do students, staff and external agencies such as external examiners professional bodies and quality assurance agencies understand the criteria employed in the assessment method and what they are designed to assess?

- Is there over-reliance on just one mode of assessment such as formal unseen examinations?

- Are students overloaded thus encouraging coping strategies which lead to what Entwistle (1981) has described as 'surface' as opposed to 'deep' learning?

- Are there wide variations in marker reliability between one assessment method and another?

- Are there variations in the demands being made on students between different assessment methods?

Effective assessment means having methods of assessment which address these questions.

Management and resource efficiency

Assessment also needs to be efficient. It needs to be cost effective as well as manageable. Thus, at a time when resource constraints within most universities are severe and when there is an expansion of students numbers, any assessment method which involves lots of staff time to undertake and to provide effective feedback to students is unlikely to be implemented successfully no matter how educationally sound it may be. In addition, it is argued that it is necessary for us to review existing practice since assessment methods designed for an era with a smaller and a more homogeneous student population are thus unlikely to be effective now (Gibbs, 1992).

The assessment methods used must also have integrity and fidelity and the costs of ensuring that assessment can achieve these must also be borne in mind. Transparency (both in terms of publishing the criteria and standards used in the assessment process and also in allowing students access to the methods and processes of assessment) is also an important consideration in evaluating assessment methods. Full disclosure of marks and grades is now no longer at the discretion of examiners. Students, as a result of legislative changes in the UK, now have a right of access to information about their assessments including comments made and marks/grades awarded. The costs of setting up systems to allow students to see all their assessments and to have access to comments and tutor evaluations must also be taken into account in evaluating assessment methods.

Assessments must also be reliable and there should be policies and procedures for ensuring that staff marking the same piece of work and employing the same criteria should award the same grades or marks. There is an obvious cost involved with this, for example, in setting up systems for 'blind' double marking. Our survey of staff attitudes indicates that while there is an acceptance of what is called 'confirmatory' double marking (where a second member of staff confirms the mark of the first market) there is some resistance to blind double marking (whereby the second marker does not know the grade or mark of the first marker.)[4] In addition, there also needs to be a safeguard for students against possible bias. Thus some universities have anonymous marking. Anonymous marking appears to be viewed more positively by staff than 'blind' double marking. In our survey of staff attitudes to anonymous marking, 78.6 per cent agreed with the statement that anonymous marking would remove any perceived bias in the assessment of students with 21.4 per cent of our sample disagreeing. Therefore the administrative costs (as well as the academic staff costs) of implementing such procedures must also be taken into account.

Furthermore, if there is a reliance on assessment methods that are conducted in a student's own time there must be systems in place to guarantee fidelity and ensure that students are not plagiarising the work of others and that impersonation is not a factor. The cost of setting up policies on plagiarism and the development or purchase of anti-plagiarism software should also be a factor in assessing the efficiency of assessment methods. Anti-plagiarism software is now commercially available and there is now an extensive literature on this topic[5]. Thus in summary, each assessment method needs to be evaluated against a set of efficiency criteria.

Combining these two sets of criteria produces a matrix of possibilities. Figure 1 below sets out possible evaluation criteria.

---

[4] In a sample of 163 staff, 59.5 per cent agreed with the statement that it was enough to confirm a colleague's mark rather than doing it blind, while 40.5 per cent disagreed with the statement.

[5] For further details see Carroll and Appleton (2001), Cole and Kiss (2000), Culwin and Lancaster (2001), Gajadhar (1998) and Van Bramer (1995).

Educational effectiveness

		C	A
	High	Old double decker bus methods	Star methods
Resource efficiency		B	D
	Low	Dogs or WOTS methods	Rolls Royce methods

Figure 1 The efficiency/effectiveness matrix
Source: Hornby (2003)

Borrowing some of the terminology of the Boston Consultancy Group (BCG) (Hedley, 1977, Hax and Majiluf, 1990), those assessment methods which meet both sets of criteria are in quadrant A and may be called star methods. These would be highly recommended. We presumably should be aspiring to have assessment methods that are in this quadrant.

Those assessment methods that are low on both sets of criteria are in quadrant B and could be called dogs (to use the BCG nomenclature, or WOTs - short for waste of time!) and should obviously be avoided. There are, however, two other possibilities which, again adapting the BCG labels, might be called old double-decker bus methods in quadrant C (they are efficient insofar as they can cater for large numbers, but they may not get you to where you should be going) and Rolls Royce methods in quadrant D (they get you to where you should be going but they are very expensive to run and maintain).

Therefore those in quadrants C and D indicate a conflict between educational effectiveness criteria and management/resource efficiency criteria. For example, unseen examinations might be regarded as an assessment method which falls into quadrant C. They are highly efficient but not necessarily very effective educationally[6]. Dissertations, on the other hand, might be regarded as highly effective in educational terms but require a lot of staff time supervising and assessing and may therefore score poorly on the efficiency dimension. They would therefore fall into quadrant D. From an analysis of some possible assessment methods (CELT et al, 2003), the following matrix has been produced:

---

[6] For example, for a controversial view of the value of educational qualifications and the skills that employers apparently value most see Education, Employers and Class Mobility, by Michelle Jackson, John Goldthorpe and Colin Mills, to be published in *Research in Social Stratification and Mobility* (Elsevier) quoted in *The Economist,* 18 January 2004 and available at www.economist.com/displaystory.cfm?story_id=2352932 [Accessed 7 March 2004].

Educational effectiveness

	Low	High

Figure 2 Survey of some assessment methods: summary findings
Source: Hornby (2003)

While it is obviously not possible to be absolutely definitive about the exact position of an assessment method's position on this portfolio analysis matrix, this framework does provide a useful tool for evaluating the efficiency and effectiveness of assessment methods.

Finally, it should be noted that the position of a given method of assessment on this matrix is in many cases capable of shifting or being shifted by making the assessment either more effective and/or more efficient in the way it is implemented. It is thus the aim of the quality enhancement initiative of which this workshop represents the start, firstly to map onto this matrix where we think our present assessment methods lie and then to shift these to a position close to or inside the quadrant A. We will return to this in a later section.

The consequences of over assessment

It is also clear from our survey of assessment methods (Hornby, 2003) and confirmed from experience in other institutions that staff are often guilty of over assessing students. In addition, there are also a number of well-documented unintended consequences of over assessment. These can be summarised as:

- poor feedback - as one student put it: When it comes to giving feedback some staff stop at nothing

- late feedback - as the old adage goes: feedback is like fish, it goes off after a week

- formative assessment is sacrificed

- students cut classes/tutorials

- students work strategically

- students look for 'short cuts' including plagiarising work and impersonating

- little meaningful learning takes place.

Institutions and departments which display any or all of these 'symptoms' are likely to have a problem with over assessment.

Strategies for streamlining assessment[7]

So how can we turn our assessment methods into 'star' methods (or make them even more 'star like')? There are six strategies that I want to look at briefly which may offer staff the chance to move their assessment methods into the star methods category. These are:

1   strategic reduction of summative assessment

2   front-end loading

3   in-class assessment

4   self and peer-assessment

5   group assessment

6   automated assessment and feedback.

1   Strategic reduction of summative assessment

Under this strategy, staff can ask themselves a number of questions in order to reduce the assessment load.

- Can I reduce the instances of assessment?

- Can I ensure that there is some relationship between assessment load and credit rating for my module?

- Can there be a policy of giving students exemptions from exams on basis of coursework performance?

- Can I assess learning outcomes once only?

- Can we combine assessments across modules?

- Should we abolish resit examinations and reassess differently?

- Are there mechanisms for reviewing and re-balancing types of assessment between modules?

- How best can we timetabling assessments?

2   Front-end loading

Under this strategy, staff would spend more time at the beginning of the assessment process doing the following:

- setting up coursework briefing sessions

---

[7] For detailed case studies of each of these strategies, readers should consult Hornby (2003) available at www.rgu.ac.uk/celt/quality/page.cfm?pge=7347#cases [Accessed 7 March 2004].

- discussing, explaining and unpacking the criteria by which students will be assessed
- getting students to engage with the criteria
- allow students input into deciding the criteria
- getting students to assess previous cohort's work to gauge the standards required.

Recent research from Rust et al (2003), reports that giving students the opportunity to review and assess previous examples of past students' work, to discuss how this was assessed, and to evaluate the criteria used can have significant, long-term benefits. His study reports that after just 90 minutes in a workshop working as student assessors, students subsequently performed significantly better than a control group who had not. What was particularly dramatic about this study was that the effects continued into the next academic session as well. The explanation offered for this phenomenon is that the students were more actively engaged with the criteria for assessment, in a way that simply passively giving them out the criteria does not achieve. Thus 90 minutes of front-end loading paid off both in terms of students learning and in streamlining the process of assessment.

3   In-class assessment

Getting students to undertake class assignments and getting them to assess one another's work in class is not new. It is, however, somewhat neglected as a simple technique which can be both effective and efficient. Having periodic short answer assignments or tests can give students quick feedback on their performance and learning. It can also act as a powerful motivation to learning if, as sometimes may be the case, performance in these assignments gains students an exemption from final examinations. Gow (reported in Hornby 2003) used this technique with first year engineering students at The Robert Gordon University and he reported:

- improved attendance rates because the assignments were based on work covered in tutorials
- improved motivation to learn
- higher retention rates as students got reassurance from the feedback on performance
- higher pass rates hence fewer resits to mark
- improved second year performance because underpinning first year knowledge was better.

4   Self and peer-assessment

There is some scepticism among staff about self and peer-assessment and our survey revealed that this technique was not much used in practice across the university. It is often alleged that students cannot be trusted to undertake assessment for themselves and that the method is unreliable. The research evidence on reliability is somewhat mixed (for examples see Boud and Falchikov, 1995; Fitzgerald, Gruppen, White and

Davis, 1997; Larres, Ballantine and Whittington, 2003). Combining the results from a number of these studies indicates:

- tutor/student correlations of their evaluations is low

- there is a high variability between tutor ratings and student ratings

- some students tend to over rate themselves.

However, more recent research points to self assessment properly introduced can be both highly effective and efficient. In a recent study, Rust, Price and O'Donovan (2003) report the results of self-assessment after students had been briefed and attended an assessment workshop on the criteria that were to be used to assess them. Of the 140 students who received the briefing, 39 per cent gave themselves exactly the same grade as their tutors while 82 per cent were within plus or minus one grade (on a five point scale) of the tutors' grade. Even with those students who had not participated in a workshop to discuss criteria (150 students) the comparable figures were 45 per cent and 80 per cent respectively. They also report that participants in the workshop were more inclined to underestimate their performance than non-participants. Therefore what this research demonstrates is that with careful briefing and some measure of independent verification by a tutor, there is no reason to believe that students would in fact inflate their own assessments.

Recently, at the Aberdeen Business School at The Robert Gordon University, 53 final students taking the Accounting and Finance degree and using the University's grade related criteria[8], were asked to self assess a piece of coursework for one of their options. The criteria were set out and explained and discussed both face-to-face and in an online discussion forum set up for the purpose. Of the 51 students who completed a self-assessment 97 per cent assessed themselves to within plus or minus one grade or classification. 53 per cent rated themselves with exactly the same classification as the tutors with a further 44 per cent having a one grade variance. In only eight per cent of the sample did students rate themselves higher than the tutor overall. There was a significant correlation between students rating and tutor ratings in all but one of the five criteria by which students were assessed. Overall the reported correlation between students and tutor ratings was much higher (r=0.61) than previously reported in the literature (Hornby, 2004).

In addition, students gained considerable benefit from the process and indicated that training on self-assessment should be an integral part of their studies. As one student put it in an email to his tutor:

> Thank you for the extensive feedback from you regarding my coursework. That type of information will be extremely useful to me in the future. The self-assessment form, although initially seemed annoying, gave me a good insight into my work and I had never reviewed my own work in the way I did until this paper. Being 'forced' to do this let me see that even marking it I could spot places where I fell down and could improve. (Shame I didn't self assess long before the deadline!)
>
> However, I would like to put in writing that this information would have been more useful had it been in use from right back in first year. I would suggest

---

[8] For details of this see www.rgu.ac.uk/academicaffairs/assessment/page.cfm?pge=2295 [Accessed 7 March 2004]. A grade 6 represents a top first class grade, a grade 5 represents a 2:1; grade 4 a 2:2 etc.

considering using that proforma feedback forms on all courseworks. I have never received feedback as extensive as yours and feel that getting this back each time will lead to a gradual improvement of grades throughout the years of the course. (I'm sure that this information in earlier years would have lead to improvements in my grades over the years as opposed to getting round about the same percentage every year.) Breaking down the grade into separate sections also helped understand more fully how our papers are marked.

Just wanted to get that in writing to you. I feel if you were to put this issue to the class the general consensus would be that although the self-assessment was a hassle at the time, their accuracy is testimony to their usefulness and that such feedback information such as what (sic) was given would be extremely beneficial starting much earlier than the last round of courseworks!'

Male student, overall grade 5

There is also evidence that some students got a considerable boost from being assessed by their tutor better than they rated themselves. As one student wrote:

Thank you! I can't believe I got that mark (I'm still shaking). I honestly did think this was one of my poorer pieces of coursework, but I'm very glad you didn't agree! Thanks also for such a detailed feedback, it's not often we get this and have found it very useful.

Female student, overall grade 6

5 Group assessment

Putting students into groups to produce work collaboratively is one strategy, which as well as providing very valuable educational opportunities can clearly reduce the marking burden on academic staff. It also has the added advantage that it encourages students to work collaboratively and to develop the necessary skills to make group working effective. These skills are generally speaking regarded by employers as important in the workplace. (See the reference in footnote 4 above.) There are, however, problems of what to do with the free-riders who contribute nothing to the group. In addition, tutors often have concerns that students may not rate each other fairly. Five options are mentioned in the literature (Rust, 2001). While no single option is without its drawbacks, each provides some answer to the free-rider problem. Below are the five options.

i Group mark

Each member of the team receives the same mark. This is a kind of rough justice and is probably the least convincing strategy. Rust (2001) quotes an example of a business course where a system of yellow/red cards was issued to those who are not attending or contributing. The yellow cards are instigated by team members and can be rescinded if there is an improvement in attendance/contribution by a given time period. If the yellow card is not rescinded a five or 10 per cent penalty is introduced. A second yellow card results in a red card and the students fails the module. He reports:

'Since the introduction of this scheme, a red card has never been used, no one has ever complained that the system is unfair and most yellow cards (and there have not been that many) have ultimately been rescinded' (Rust, 2001).

ii   Individual learning contracts

Each member of the group is assigned a role or a task and is assessed against the assigned role. The drawback with this option is that it can be difficult to discern who has actually done the work and involves a good deal of tutor time, thus defeating the objective of streamlining assessment.

iii   Peer-assessment/divided group mark

Marks are assigned by members of the team on the basis that they know best who has contributed the most to the performance of the group. Thus, in some cases, it can be done by formula with a group mark and variations around this mark +/- 10 per cent for example according to individual contribution. Alternatively, a group mark might be 60 per cent and with six members there would be (n - 1) x 60 or 300 marks available for distribution with a minimum of 0 and a maximum of 100. Each member allocates marks (excluding him/herself) and an averaging is carried out to reflect group views.

iv   Viva

Marks can be allocated on the basis of group presentations followed by a questions and answer session to ascertain who was making the most significant contributions.

v   Project exam

Students can expect to get questions in the examination, which specifically draw on their experience of their work in the project. Free riders will find it difficult if not impossible to answer this part of the examination.

Barnes (quoted in Hornby, 2003) undertook group assessment in the third year of the BA degree in Hospitality Management students using an evaluation instrument for both self and peer-assessment. This was administered electronically via QuestionMark Perception™ and assessed a number of different dimensions of individual and group work, such as attendance, ideas generation, contribution to group report, knowledge and skill acquisition, and effort. There was also an anonymous assessment of the work of others which was fed back to each student so that not only did students get a grade but they also got a score on each of the identified dimensions and a report on their performance. By using computer software a lot of the hassle of collating marks was removed.

6   Automated assessment and feedback

There are two areas where some automation of the assessment process can increase efficiency. Firstly, when assessing work, staff can use checklists and statement banks to automate the feedback process. For example, in the School of Applied Sciences at The Robert Gordon University staff use a series of checklists when marking laboratory reports. Statement banks are also used in giving students feedback in the Economics of Tax module. Using the University's criterion reference scheme, grade related statements on each of the dimensions were drawn up and cut and pasted into a feedback report. Each report was then personalised and emailed to the students in real-time. Thus the students received immediate feedback as the tutor was doing the marking ('quick and dirty' as opposed to 'clean and slow' to use the jargon). In other cases whole class feedback online can be given via the University's intranet which also speeds up the process of giving feedback (for details of this work see Hornby, 2004b).

Secondly, the assessment itself can be automated using computer automated assessment. Cooper (quoted in Hornby, 2003) uses this method with BSc (Hons) Sports Science students in the School of Health Sciences. Using a self assessment instrument designed with QuestionMark Perception™ linked to video clips of particular movements of the body, she enables students to identify key muscles and joints in particular sports. The tests required students to identify muscles producing the movements and the type of muscle work involved. As well as giving every student a consistent set of video examples (something that is almost impossible to do with live models), it enables students to use a stop/start/replay facility. The using multiple-choice questions with feedback on each response she was able to test students understanding of the topic. She also used an intranet discussion forum to ask questions about the various parts of the syllabus and set up self assessment assignments for the students. This is a technique which is both extremely effective as well as efficient. Instead of having to set up live demonstrations each time students wanted to test themselves, this technique allows students to test themselves when they were ready. It also gives students feedback for remedial action and hence is a form of formative assessment. This is clearly a technique which is transferable to areas such as science and engineering.

How will this save me time?

A number of the techniques outlined above can clearly save staff time. For example, a strategic review of assessment can yield quite significant savings. Taking a student's view of the assessment loading can be very instructive. Do students need to be assessed in exactly the same ways or is there scope for rationalisation across modules/subjects? Is there any way we can avoid bunching of assessments at the same time of the semester?

In a number of instances, it has to be admitted that there are some set up costs, for example, with computer automated assessment, and one should not underestimate the costs here. On the other hand, these costs can be viewed in the same way as economists view investment appraisal techniques. There will be an initial cost which will yield a return over the life of the assessment method giving a rate of return both in terms of releasing time and in being educationally more effective. The benefits of any change over a period should outweigh the costs (and if they don't then, assuming academic staff are rational and are out to maximise their efficiency and effectiveness, it would be irrational to adopt them).

So the next time you sit down to mark 250 case studies, essays, laboratory reports or examination scripts ask your self one question. Can I do this differently in a way which makes my assessment more efficient and more effective?

References

Biggs J (1997) *Enhancing teaching through constructive alignment, Higher Education*, 32, 347-64

Boud D (1995) Assessment and learning: contradictory or complementary? in Knight P (ed) *Assessment for Learning in Higher Education,* pp 35-38, Kogan Page, London

Boud D and Falchikov N (1995) What Does Research Tell us about Self Assessment? in Boud D (ed) *Enhancing Learning through Self Assessment*, Kogan Page, London

Brown G (2001) *Assessment: A Guide for Lecturers*, Assessment Series No 3, LTSN Generic Centre, London
www.heacademy.ac.uk/resources.asp?id=1&section=generic&process=filter_fields&type=all

Brown S and Glasner A (1999) *Assessing Student Learning in Higher Education*, Kogan Page, London

Carroll J and Appleton J (2001) *Plagiarism: a good practice guide,* Joint Information Systems Committee
www.jisc.ac.uk/index.cfm?name=programme_plagiarism [Accessed 7 March 2004]

CELT et al (2003) *Assessment Methods Report*, Centre for the Enhancement of Learning and Teaching in conjunction with the Department of Accounting and Finance, The Robert Gordon University, Aberdeen

Cole S and Kiss E (2000) What can we do about students cheating? *About Campus* May/June, pp 512

Culwin F and Lancaster T (2001) *Plagiarism, prevention, deterrence and detection*
www.heacademy.ac.uk/embedded_object.asp?id=21635&file [Accessed 7 March 2004]

Entwistle N (1981) *Styles of Learning and Teaching*, Wiley, New York

Fitzgerald T J, Gruppen L D, White C and Davis W K (1997) Medical Students' Self Assessment Abilities: Accuracy and Calibration, *Proceedings of the Meeting of the American Educational Research Association*, Chicago

Gajadhar J (1998) *Issues in Plagiarism for the New Millennium* (online)
http://ultibase.rmit.edu.au/Articles/dec98/gajad1.htm [Accessed 7 March 2004]

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

Hax A and Majiluf N (1990) The use of the growth-share matrix in strategic planning, in R D Dyson (ed) *Strategic Planning: Models and analytical techniques*, Wiley, New York

Hedley B (1977) Strategy and the Business Portfolio, *Long Range Planning*, Vol 1, No 1, 9-15

Hornby W (2003) *Case Studies on Streamlining Assessment*, Centre for the Enhancement of Learning and Teaching, The Robert Gordon University, Aberdeen
www.rgu.ac.uk/celt/learning/page.cfm?pge=7347#cases

Hornby W (2004a) Tutors, who needs them? Self Assessment using Grade Related Criteria in an Accounting Degree, *Journal of Accounting Education* (under review)

Hornby W (2004b) Enhancing Effectiveness and Efficiency in Student Feedback, in *Student Enhanced Learning through Effective Feedback*: *The SENLEF Project*, LTSN Generic Centre, London
www.heacademy.ac.uk/1346.htm [Accessed 7 March 2004]

Van Bramer S (1995) *What is Plagiarism?* (online)
http://science.widener.edu/svb/essay/plagiar.html

Streamlining assessment: making assessment more efficient and more effective

Professor Ray Land, Centre for Higher Education Development, Coventry University

Introduction

Few practitioners in UK higher education would fail to recognise the dysfunctional effects that a declining unit of resource set against increasing student numbers has had on assessment and other academic practices over the last two decades. There is, in many departments and courses, a clear and urgent need to reduce the amount of academic time that is currently invested in assessment and examination activities - time which detracts from other important academic pursuits.

The discussion that follows will consider various moves that have been adopted to ameliorate the assessment burden, either from a 'control' perspective whereby traditional methods are carried out with greater industrial efficiency, or from an 'independence' perspective which shifts a degree of responsibility for (usually formative) assessment to students themselves, on the grounds that, apart from economies of effort, this is in itself an educational benefit that is likely to enhance student performance (Gibbs and Jenkins, 1992). However, raw strategic reduction of assessment activity might well prove to have unintended consequences in terms of learning effectiveness. It would appear that though we can, without too much difficulty, identify various principles of good assessment, these same principles often contradict or undermine each other. We know, for example, that increased validity can sometimes lead to decreased reliability. What is efficient may not be effective, and vice versa. Any adopted blend of assessment practice entails both an opportunity cost and an implicit perspective on the nature and purpose of learning. It may be that as we widen and deepen participation in UK higher education we will need to 'repurpose' our assessment methods to suit these changed circumstances. The use of the 'patchwork text' approach will be discussed as one innovative example of such repurposing.

Finally, even where clear strategic priorities are identified, convincing colleagues to undertake what may be radical changes in their well-established practice can be a complex process, requiring differing degrees of risk and support. This session will advocate a 'satisficing' approach of informed expediency, which recognises the constraints and possibilities of working within specific disciplinary cultures and institutional contexts and will hope to open up discussion of the important issues arising from the tension between efficiency and effectiveness.

Principles of assessment

We should not underestimate the extent to which the identity of our students as learners is constructed through the operation of our assessment practices, both during, and often long after, the completion their programme of study. However, Paul Ramsden, the new Director of the Higher Education Academy, once referred to the student's experience of assessment in higher education as 'a serious and often tragic enterprise'. If we agree that assessment practices should be informed in a more coherent manner by educational principles, then a process of assessment, it might be argued, could be predicated on the following framework.

Assessment should be **beneficial** in its effect, inasmuch as it should not adversely affect or distort the processes of teaching or learning, and should not undermine the trust or relationships between teachers and learners. Nor should it weaken morale, discourage initiative or innovation, or lessen the commitment of either staff or students.

On the contrary, particularly when used in a formative sense, the purpose of any scheme of assessment should where possible, be to actively foster learning. Assessment is part of the process of learning. As Knight (2002) has indicated, timely feedback and sound formative assessment practices are likely to enhance learning more than almost any educational innovation.

The operation of the assessment process should be seen to be fair for all concerned. No individual or group, for whatever reason, should enjoy privileged status or suffer undue disadvantage in terms of the academic judgements that are made about their performance. Matters relating to illness, disability or other reasons for non-completion of work will need to be handled sensitively, equitably and consistently.

To ensure fairness, appropriate appeal procedures should always be in place to allow students a means of gaining the right of redress where this is deemed to be justified. Both staff and students should be fully cognisant with these procedures from the outset of the assessment process.

To further ensure fairness, the process of assessment should be varied and diverse in terms of the range and timing of the assessment methods used (eg formal written examination, coursework assignments, projects, practical work) and the benefit of multiple perspectives that might be gained from a variety of modes of assessment (eg assessment by diverse tutors, self-assessment, peer assessment, group assessment, work-based supervision, externals). Varied approaches to assessment also require appropriate means of recording achievement in order to profile diverse forms of knowledge, skills and attributes.

Related to the notion of fairness is that of reliability. Since assessment of academic or professional performance inevitably reflects assumptions about values, then one individual's judgements of academic achievement may necessarily differ in some degree from another's. Such difference needs to be acknowledged and appropriate measures put in place to ensure relative consistency. This means that judgements or scores awarded by the same assessor from year to year remain as consistent as possible, and that assessment of the same examinee by different assessors would be likely to be of the same standard.

However reliable an assessment system is, it will not be fair unless it is seen to be valid. It needs to be seen to be seeking evidence of what it claims to be seeking evidence of. Any judgements made should be based on that evidence and not other, non-relevant evidence. What this means in practice is that if an assessor is seeking to assess the degree of a student's understanding of, say, aspect x of state policy, then judgements of that student's achievement should not be distorted by their understanding of aspect y of state policy, nor, say, of the quality of the student's handwriting, unless this has been explicitly made clear in the question or in the assessment guidelines.

This leads on to the issue of transparency, whereby the purposes, procedures and criteria of the assessment process are open, clearly stated and understood by assessors, teachers and students. All need to understand, as far as is reasonably possible without compromising security or fairness, what the learning outcomes of a particular programme are; which of these are likely to be assessed (not, of course, the actual questions); on what basis decisions are to be made (marking criteria); the nature of any grading system and the nature of any appeals process.

Any scheme of assessment should be as representative as possible. An assessment specification which samples only three per cent of the work covered in a year, or of a student's performance, might be considered unfair, invalid, unreliable and ineffective.

An assessment scheme should be effective in that it achieves the purpose it is intended for. This purpose might be **summative**, in terms of providing an accurate judgement and record of a student's attainment in a subject at any given point, or it might be **formative** in terms of helping a student to learn from previous performance in order to improve subsequent performance, or it may be **diagnostic** in terms of ascertaining a student's particular strengths and weaknesses and their learning or developmental needs during a forthcoming programme. Whatever the purpose, if it is being fulfilled, then the scheme can be said to be effective.

Any assessment scheme will be ineffective if it cannot be adequately resourced or managed efficiently because it is overly bureaucratic, cumbersome or consumes inordinate amounts of either student or staff time in its operation. If a scheme is not practicable and cost effective, then clearly it cannot continue and so all other criteria fall by the wayside in due course.

Within learning environments, assessment strategies and instruments should be constructively aligned with other significant aspects of the learning environment, such as course design, learning outcomes, methods of teaching and learning, evaluation and feedback strategies.

Finally, a process of assessment must ensure the security of its operation in terms of the safe recording, transfer, storage and retrieval of information on student achievement. Fairness, effectiveness and the right to redress are all predicated on the assumption of security of operation and the prevention of any loss of information or fraudulent practice.

We need to bear in mind, however, that when applied in specific educational contexts, these principles of assessment, like any set of assessment principles, can contradict. For example, what is efficient may not be effective, and what is effective may not be efficient. What is valid may be so at the expense of reliability, and so on.

The nature of assessment

It might be worth pausing at this point to consider what constitutes the nature of assessment in higher education. Any assessment strategy seems to be characterised by four stages of activity. The first stage involves **sampling** evidence of performance. We can never assess the whole of a student's performance or achievement and a decision has to be made as to what an appropriate sample would comprise. The second stage is then to make **inferences** from that sample, and thirdly to **estimate** its worth. The final

stage involves **representing** that worth, usually symbolically through marks, letter grades, points, classifications or reified adjectives such as 'excellent' or 'satisfactory'.

Rudyard Kipling, in his *Just So Stories,* spoke of six 'serving-men' that proves useful in making sense of problematic undertakings.

> I keep six honest serving-men
> (They taught me all I knew)
> Their names are What and Why and When
> and How and Where and Who

These serving-men can prove instructive when designing an assessment strategy:

What kinds of knowledge, skills, abilities or values are we seeking evidence of?

Why are we seeking to gain evidence of achievement in the first place? For what purpose will this evidence be used?

When should assessment take place - at the end of the course? At periodic intervals throughout it? Continuously? At the start of the course?

How should we gather the evidence? What is the most appropriate strategy to use for our particular course?

Where should assessment take place - in the examination hall? During a group presentation? In the workplace? On a computer?

Who should undertake the assessment - the teacher? Anonymous examiners? A workplace/clinical supervisor? The student? The student's peers?

Purposes of assessment

In attempting to render assessment more efficient and effective we will need to clarify, in any given context, what it is for. What are the purposes of assessment and who will be the different stakeholders with an invested interest in receiving assessment information for any particular identified purpose? Our assessment strategy might serve any, or several, of the following purposes (or others not listed here, of course).

Diagnostic testing
Providing feedback to students
Monitoring progress
Motivating/regulating students (and staff)
Developing student confidence
Predicting future performance
Evaluation of teaching
Certification/licensing
Selection for subsequent courses
Information for employment
Public accountability
Institutional marketing
Social stratification

In terms of streamlining assessment we would, of course, have to ponder which of our identified purposes had been achieved, to what extent, and with what degree of efficiency - the latter, presumably, measured in terms of comparison of resource expenditure with some previous norm.

Effects of massification

The assessment enterprise has been complicated further by the massification of higher education in recent decades and, to some extent, by its accompanying managerialist strategy of modularisation. The combined effects of these tendencies have led to dysfunctional conditions within the system, which any serious attempt at streamlining will need to address. Over assessment has become an issue for both staff and students. This is compounded by a lack of advice to students on improvement and a lack of knowledge of progress. The climate of resource stringency and worsening staff/student ratio brought on by increased student numbers has led to slower rates of return of coursework. There is also increased redundancy within the system with the same learning outcomes or skill sets repeatedly assessed across different modules. The ratio of assessment workload per credit point is also frequently not well calibrated so that modules bearing similar credit ratings might generate widely differing assessment workloads for students and marking burdens for their teachers. Mass modular schemes can also witness a convergence of assessments, or 'bunching', leading to stress points at particular times of the academic year which even students with the finest honed time-management skills still find difficult to resolve. Without an efficient means of curriculum monitoring, there can develop a lack of variety of assessment methods, with a repetition of certain approaches across modules, such as poster presentations or web-based multiple-choice questioning.

Recent research (Naidoo, 2003) has indicated that, within a massified system with increasingly commodified curricula, consumer-oriented provision might be fostering a culture of passivity and non-criticality within fee-paying students seeking a trouble-free acquisition of knowledge. An awkward tension arises for tutors caught between the competing discourses of flexibility, access, meeting of student need (and demand) and widening participation on the one hand, and market economics and increased student numbers on the other. An effective system of assessment will need to be one which can maintain rigour in terms of fairness and reliability while widening access. A number of specific issues arise from widening participation which will need to be addressed if our assessment practices are to be effective.

For example, the globalisation of communications through the internet has complicated the issue of plagiarism through ease of access to and incorporation of available written materials into student work. An effective system of assessment will be one that can identify and implement effective measures for the prevention, detection and handling of student plagiarism (Carroll, 2003; Ashworth, 2003). Internationalisation has also drawn attention to different cultural understandings of assessment practice, with, for example, the need to explain to students from particular cultural backgrounds the employability rationale underlying the peer-assessment of group work or group presentations.

Widening participation also has highlighted the importance, within the more general area of academic literacies, of the acculturation of students from non-traditional

backgrounds into the practices, conventions and genres of academic life, which includes the conventions of assessment practice. The Centre for Academic Writing at Coventry University, for example, has been established to help address this kind of issue. Considered from the perspective of the 'new' higher education, attempting to meet the needs of a significantly widened intake, the salience and traditional dominance of particular instruments of assessment, such as the essay, become apparent. This indicates a need for a wider variety of assessment instruments, particularly in the early stages of a programme, allowing gradual familiarisation with the genres of assessment and styles of academic that are necessary. This emphasises the role of assessment in the pacing and speed of learning. The work of Winter (2003) and his colleagues in developing the 'patchwork text' approach to assessment and writing provides a useful illustrative example in this regard. Widened participation also has increased our understanding of the importance of formative assessment early in the first term with the accompanying need for sensitivity in feedback. Many first term assignments now carry marks that serve only to be indicative and do not count towards final module scores.

Choice of strategies - control or independence

Within a massified system, Gibbs and Jenkins (1992) argue that strategic options ultimately reduce to choice between what they term '**control' strategies** and '**independence' strategies**. In the former, traditional assessment methods are retained but carried out with increasingly greater (industrial) efficiency, much akin to the phenomenon of 'performativity' discussed by Lyotard (1986). Such an approach, these authors suggest, eventually runs up against limits determined by available resources. Within independence strategies, on the other hand, there is a shift of responsibility for (usually formative) assessment to students themselves, on the grounds that, apart from economies of effort, this is in itself an educational benefit that is likely to enhance student performance.

Examples of control strategies might include the following.

- Set fewer assignments.
- Set shorter assignments.
- Set fewer exam questions.
- Devise multiple-choice questions (MCQs).
- Use standardised feedback forms.
- Use computer-aided assessment (CAA), eg computerised objective tests, word-processed statement banks, optical mark readers.
- Set limits on assignment length and materials costs; penalise those who go beyond the limit.
- Conduct more in-class assessments.
- Mark in pencil on written assignments to make alterations without time or trouble.
- Sort scripts into piles according to approximate standard - this helps to maintain a consistent standard.
- Scrap grades and percentages and use pass/fail instead. Scrap degree classifications.

- Refuse to accept work that is difficult to read or late without reason. Keep to this policy.
- Use placement and clinical assessors - standardise criteria and provide training.

Examples of independence strategies would be the following.

- 'Front-ending' assessment.
- Devise group tasks and assess the group - group reach consensus on scores for individual members.
- Self-assessment - require students to mark their own work with comments, or submit a completed standardised feedback form with their work.
- Peer-assessment - ask students to use a marking scheme to mark each other's work.

(Adapted from Gibbs and Jenkins, 1992)

Within different institutional and disciplinary contexts, of course, adopting differing approaches might raise problematic change issues. Questions of the degree of both risk and support involved are likely to determine both the adoption and likely success of any initiative. Initiatives will be viewed, depending on the particularities of any given educational context as:

- high risk high support
- high risk low support
- low risk high support
- low risk low support.

Risk might relate to, say, the degree of endorsement or hostility from departmental colleagues, or the attitude of participating students. Support might refer to the technical help, additional equipment or other resource required for the propose change. Clearly staff are more likely to be receptive to what they perceive in their context to be low risk, low support initiatives. To conclude a number of low risk (though not always necessarily low support) initiatives are offered for consideration.

Efficient and effective assessment: five examples of low-risk adaptive initiatives

1   MCQs in introductory chemistry

In this example from a scientific context (cited in Rust, 2002) the course 'puts MCQs on the course's website each week for the students to do in their own time. There are no marks for these tests, and no record is kept of who has taken them, but the students do the questions because they know from the beginning of the course that the end of module exam will include a section of MCQs and that half of these questions will have been selected from those questions used in the weekly tests.'

2   Confidence measures in anatomy

In this medical example (adopted at University College, London) students undertake regular MCQ tests but instead of simply marking their chosen response on the test sheet they also have to indicate a 'confidence measure' (CF), in order to minimise the

element of guesswork common to many MCQs. If they indicate a high CF and get the correct answer they gain a high mark. If they indicate a high CF and get it wrong, they will be heavily penalised.

3   Adapting a geography course

The Assessment Strategies in Scottish Higher Education project provides a valuable repository of innovative assessment practices (Hounsell et al, 1996). The many case studies collected by the project can now be accessed online at www.heacademy.ac.uk/asshe/project.asp

One of these case studies, by David Sugden of the Department of Geography, University of Edinburgh, provides an interesting example of how assessment might be made more effective and efficient within an honours geography degree programme. The team converted all examination questions from 30/40 minute responses to hour-long essays, thus setting fewer questions, but removing the possibility of students getting away with only recall. The team found that the immediate response was that the quality of answers improved and students did more reading. The number of essays was also reduced throughout all modules, with each essay requiring more work on the part of the student, and contributing more to the final grade. Class examinations were removed from most modules and the number of exemptions given to first year students increased. Pass/fail examinations were used to a larger extent, and more were planned. Project work also increased and now accounts for up to 45 per cent of the module grade. A number of other adaptations were found to be effective. In the honours year, prior discussion between staff and students as to what makes a good project was followed up with input in tutorials. Several options within the degree programme now incorporate oral assessment and are peer assessed, with half of the grade being awarded by the students. In feedback, some of the discussions revolve around why peer and tutor marks differ. The Department encourages the submission of all work in a word-processed format (though concerns have been voiced about the increasingly poor writing styles on exam scripts). Finally, the dissertation deadline has been changed to later in the degree programme, and the greater experience gained from options and the like has resulted in higher quality dissertations, which have been winning national prizes.

4   Restructuring an education course

In a similar way, Gibbs and Jenkins (1992) report on the adaptations made to an education course, to shift the emphasis from a control strategy to an independence strategy, and in doing so render the programme not only more effective in encouraging the independent learning of students, but also considerably more efficient in terms of staff time, as the following table indicates.

Original course	Staff time	Staff time	New course
Lectures (18 weeks x two hours)	36 hours	36 hours	Workshops (18 weeks x two hours)
Seminars (18 weeks x nine groups of eight)	162 hours	36 hours	Autonomous seminars (18 weeks x two tutors touring six groups of 12)
Total teaching hours	**198 hours**	**72 hours**	
Essays (72 x two essays at 20 minutes)	48 hours	6 hours	MCQ tests (six x one hour administration)
Examination (72 x two questions at 10 minutes)	24 hours	18 hours	Seminar presentations (peer assessed, one hour/week collation)
		36 hours	Portfolio (72 at 30 minutes)
Total assessment hours	**72 hours**	**60 hours**	

5   Using 'Discourse' technology for formative assessment in history

Finally, the recent technological innovation of classroom communication systems (CCS) has been found to be a highly effective means of enabling diagnostic and formative assessment within the classroom. These systems, such as Discourse technology, allow students to respond anonymously within class, or for their response to be seen only by the tutor, so there is no public loss of face before the class, and the tutor is able to gain a much more accurate view of the understanding of a topic by the whole class, or by an individual student.

The following screenshots from a history class using Discourse technology indicate what the student would see, for example, a question concerning the effects of World War II the United States, and the student's own response.
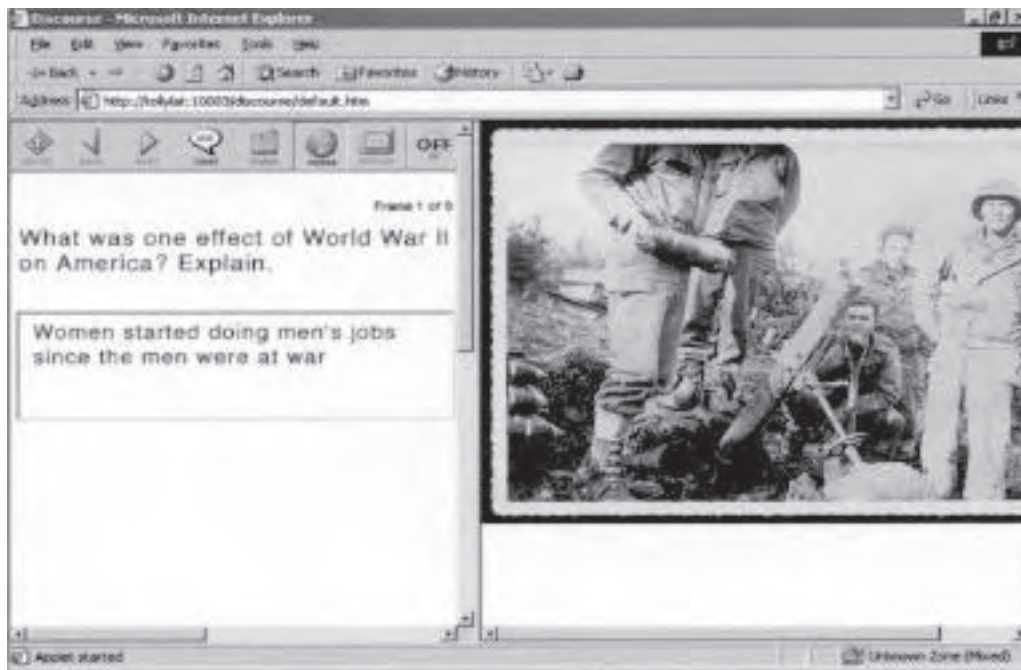
The tutor can set the system to see all student responses, as on the following image:
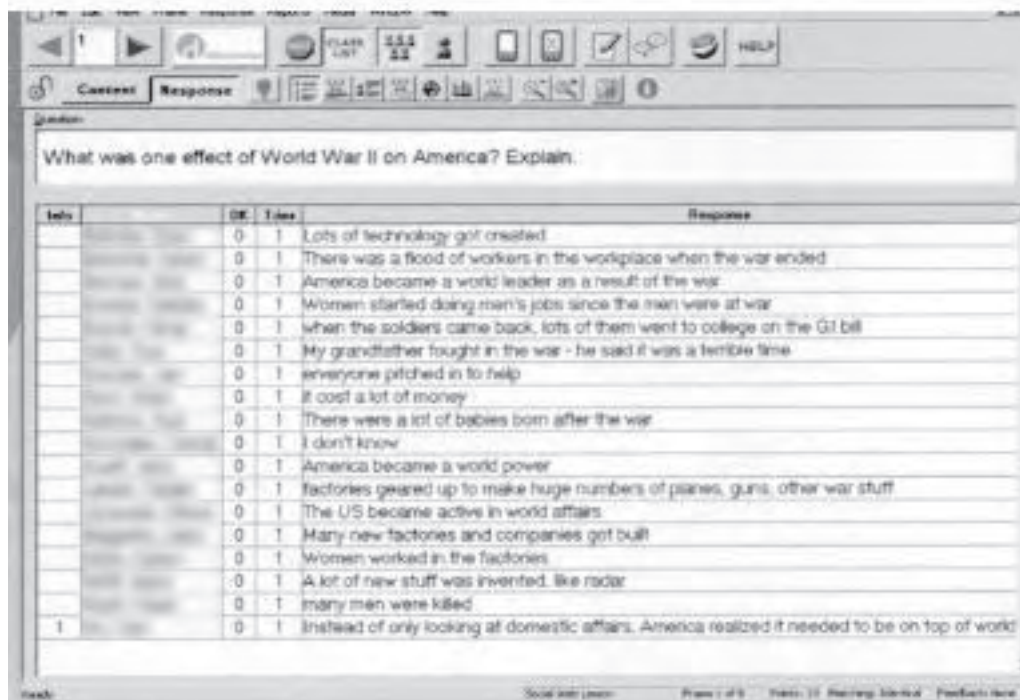
The tutor can choose to see student names next to their responses or leave the responses anonymous. The responses can be readily kept as a record and, depending on the judgement of the tutor, can then be used as an anonymous resource for the whole group, or, as below, an interesting response from an individual student can be shared with the whole group.
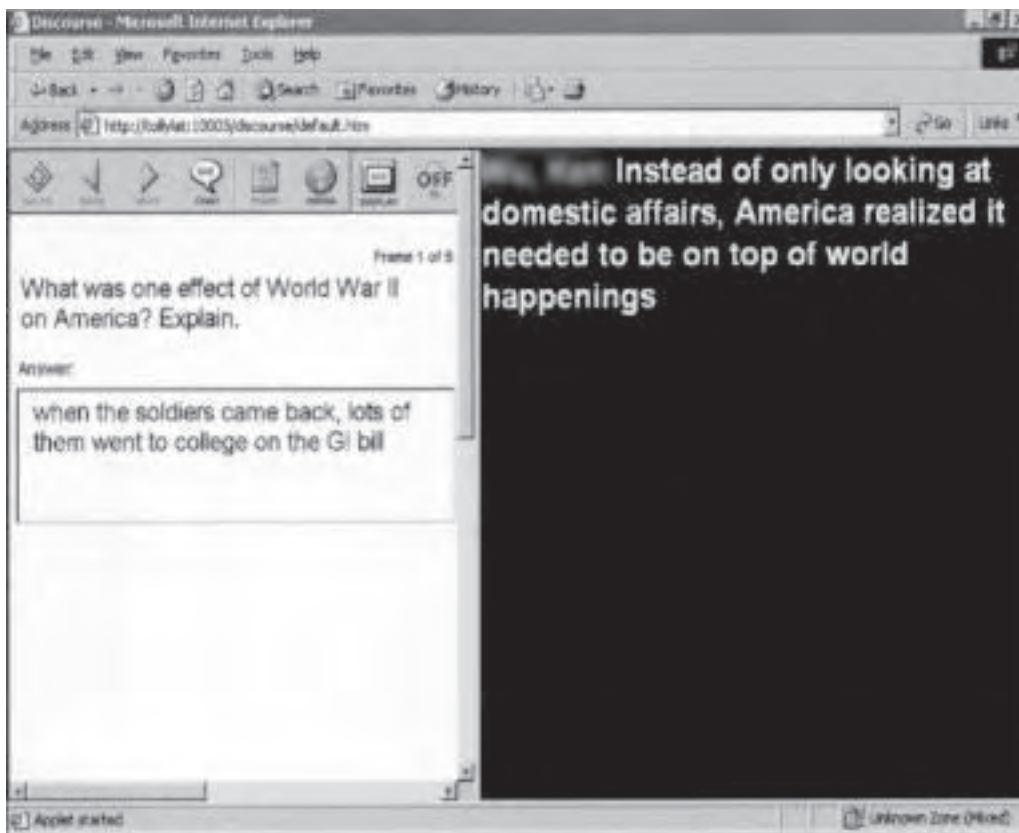
The effective uses of a CCS system seem to be its diagnostic/formative assessment potential, its quick-fire tests of understanding, the possibilities for peer judgement and promotion of discussion. In a discussion session with medical students after a clinical reasoning exercise, one medical student, responding anonymously, indicated that the intervention he would have made in a specific setting would have been to give an adrenalin injection - an intervention that would have proved fatal in the circumstances. The CCS system allowed the tutor to discuss and correct this response to the whole group without identifying the individual concerned, or indeed even indicating that this was the sole response to suggest such an intervention. In this example, the diagnostic assessment proved invaluable in identifying a serious professional misunderstanding and also allowing significant learning to take place without loss of face or humiliation. The tutor concerned remarked that it was unlikely that the student would have revealed this misunderstanding in a conventional face-to-face environment.

The system provides a useful research tool to create an instant teaching resource and also permits evaluation of the effectiveness of content and/or delivery. Student response seems to suggest that it makes lectures less boring! There is a degree of risk with this approach in that allowing students to respond anonymously can have unpredictable results! There is a degree of support required, obviously, in acquiring and installing the system though induction in its use for both staff and students is relatively simple. The differences between Discourse technology and previous generations of interactive software are that it is used between people that are linked (including being linked remotely), it is bi-directional, and web-based, and it doesn't need special equipment other than networked computers with IP addresses. Formative questions, and open-ended questions can be asked 'on the fly'.

References

Ashworth P D (2003) Why merely policing plagiarism is not the answer, in C Rust (ed) *Improving Student Learning: Ten Years On* (Proceedings of the 10th Improving Student Learning Symposium), pp 374-383, Oxford Centre for Staff and Learning Development, Oxford

Carroll J (2003) *A Handbook for Deterring Plagiarism in Higher Education*, Oxford Centre for Staff and Learning Development, Oxford

Gibbs G and Jenkins A (1992) *Teaching Large Classes in Higher Education*, RoutledgeFalmer, London

Hounsell D, McCulloch M, Scott M, Burley E, Day K, Falchikov N, Haywood J and Land R (1996) *The ASSHE Inventory: Changing Assessment Practices in Scottish Higher Education*, UCoSDA, Sheffield, with the University of Edinburgh and Napier University, Edinburgh
www.heacademy.ac.uk/asshe/project.asp

Knight P T (2002) *Being a Teacher in Higher Education*, SRHE, Open University Press, Buckingham

Lyotard J L (1986) *The Postmodern Condition: A Report on Knowledge*, Manchester University Press, Manchester

Naidoo R (2003) Corroding Relations: The Impact of Student Consumerism on Teaching and Learning in Higher Education, *SRHE Annual Conference, Research, Scholarship And Teaching: Changing Relationships?* 16-18 December, Royal Holloway University of London, London

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,* Vol 3, 2, 145-158

Winter R (2003) Contextualizing the Patchwork Text: addressing problems of coursework assessment in higher education, *Innovations In Education and Teaching International*, Vol 40, 2, pp112-122

Streamlining assessment: some options for assessment in class

Dr Sandy Hutchison, Centre for Learning and Teaching, University of Paisley

Abstract

Talking of streamlining assessment might simply imply doing fewer things and doing them faster, but an emphasis on efficiency and effectiveness should point to something more timely and productive. This paper considers some precepts for pedagogy and strategies for in-class assessment, with students actively participating, and designed to contribute to developmental learning. The original context for most of the examples provided was a pair of linked first year modules in communication and professional skills offered at the University of Paisley and designed as an introduction to the construction industry. Student-centred approaches to class activities and projects, and assessment methods associated with these, were also used in other engineering, media and science-related programmes as part of introductory modules to communication and information technology. The methods should have broader application to other subject areas too. The paper considers some approaches to providing feedback in class, facilitating and moderating peer and self-assessment. It also outlines criteria for assessment of transferable skills including written and oral presentation and group project work that involves role play and simulation. Further details on particular methods and activities will be available at www.paisley.ac.uk/clt

Introduction

Some precepts

Before I go into the details of curriculum design and assessment practices, I want to glance at three quotations from very different sources.

'The processes of assessment and curriculum design are one' (John Heywood, *Assessment in Higher Education*).

'The highest forms of originality are far more closely akin to the lowest biotic performance than the external circumstances would indicate' (Michael Polanyi, *Personal Knowledge*).

'Improvement makes straight roads, but the crooked roads without Improvement are roads of Genius' (William Blake, *The Marriage of Heaven and Hell*).

The first of these quotes is directly applicable to the matter at hand; the other two are only indirectly relevant - though I hope illuminative of broader teaching and learning issues worth bearing in mind.

Heywood encourages us to look at the whole system: if we are aiming to revise or refine what we are doing in assessment, then we have to consider carefully what and how and where and when we want students to be learning - and not simply chop or change or drop things because we are under pressure to do so. We have to consider the grounds of learning and clarify the context.

As to the second quote: this should give us all hope! What I take to be the main point of Polanyi's claim is that in fundamental terms - in terms of the working parts of living organisms, like physiology or biochemistry - there isn't a great deal of difference between the astonishing calculations of James Clerk Maxwell, or his successor Albert Einstein, and what it takes for a larder bug to navigate the floor boards in my kitchen when the lights are out. But what a difference in terms of the imagined reality, and scope of reference.

For my purposes here I just want to take this as a prompt to say: let's keep our expectations high for ourselves and our students - let's shift from dealing in deficit models of learning, and set our sights a bit higher up this continuum.

As to Blake's provocation from his *Proverbs of Hell*, it gives a reminder that improvement only takes us so far - we can make a kind of linear progression - but beyond improvement we may want to see things **transformed**. (Blake's use of the word 'without' here probably carries the sense of **beyond** as well as 'not connected with'.)

If that is so, we may have to depart from the familiar routes and methodologies and take a track or direction totally unfamiliar: relying on knowledge that is intuitive and implicit, not dependent on previous systems, and resisting easy quantification. Also it reinforces the fact that there are some outcomes from any learning experience - sometimes the most valuable - which are **unanticipated**. These need to be recognised; not discarded simply because we didn't plan for them.

So, looking again at these three texts, I would emphasise that streamlining assessment will proceed best if we keep checking the overall design of the setting for learning; and that before we start to prune and re-shape we give some thought to feeding the roots and also to the form the new growth may take. Blake also implies there has to be some imaginative spark and risk, or we're likely headed for slight advances and small potatoes.

Putting things into practice

With those texts as a backdrop, I want to consider pedagogy more directly, and particular case studies based on work with first year students at the University of Paisley that began in the early 90s. Student-centred learning was not a phrase with much currency then, but that (in retrospect) was what we were aiming for. I particularly acknowledge collaboration with Professor Bill Cranston, formerly head of the Department of Civil Engineering at the University of Paisley, and also the inspiration we took from Donald Woods at McMaster University in Canada, and Bob Matthew and Dave Hughes at the University of Bradford. Much of the time we were flying by the seat of our pants; much of the time we couldn't be sure if we were heading for the Maxwell end of the spectrum or down the other end with the larder bugs!

We did intend to shake up the prevailing system, especially in our methods of delivery and assessment schedules. What had gone before was becoming slightly stale and repetitive: students had begun to complain about lack of relevance, lack of real issues and engagement, and some were voting with their feet.

The main things that turned the tide were the key features of student-centred learning.

- **Active involvement** in all phases of learning, especially feedback and evaluation.
- **Responsibility** for shaping new and reorganising old concepts.
- **Ability to formulate good questions** rather than sit back and wait for the answers.

To help achieve this we recast both delivery and assessment methods. The previous reliance on lecture input to the whole class was cut back to a one hour lecture per week on current professional issues delivered by internal specialists and by invited external experts. The main class contact was based on seminar-workshops which ran in parallel with information technology (IT) and design laboratories, and were timetabled so that they could accommodate longer activities such as project work. Personal tutor meetings were built in to the required staff-student contact time and integrated with workshop and assessment activity. In line with the move away from deficit models, we were also keen to extend the scope of what our first year students might be asked to tackle, and how they might view their own professional competence.

Re-designing assessment

What about assessment? The inherited forms were library-based, written reports on predetermined topics, individual oral presentations and end of term exams. We did away with exams although for a couple of academic sessions we kept a class test as an option, mainly as a means of mopping up stragglers (those with borderline attendance or otherwise under-performing). The test had short essay questions based on the professional input lecture series. For students in good standing the requirement was waived.

Reports were retained, as were individual presentations, though the circumstances in which these were performed and assessed underwent change. In fact, though initially we **increased** the elements of assessment, we did some shaking down as we found out what worked and what didn't. How did that come about? And, how can we be considered to have **streamlined** assessment? We achieved this not simply by doing fewer things and doing them faster, but by making assessment timelier and more productive.

- We developed clear marking criteria, sometimes developed with student input explicitly linked to specified aims and outcomes.

- We increased opportunities for assessment and feedback in class.

- We introduced systems for peer and self-assessment.

- We made an effort to avoid bunching by mapping out assessed activities so as not to conflict with key assessments in other modules.

It is worth pointing out that in Scottish universities, particularly in the 19th century, as described in George Davie's *The Democratic Intellect*, the professors traditionally took a main part in the teaching of first year students, engaging directly with them, especially in tutorial sessions, which backed up their lectures, analogous to our approach.

These meetings 'in committee', as one contemporary professor described them, put less emphasis on questions of detail 'than on topics connected with what one might call the common sense of subjects, and to their relation to life in general'. What's more, we find: '*decisions as to the order of rank in class often lay not with the Professor but with the class*' (my italics). So peer assessment stretches back quite a way in Scottish higher education pedagogy.

What about the assessments we chose for the two linked modules? Here are the main elements, both formative (F) and summative (S), and sometimes optional, spread over two semesters. The first two activities didn't involve any formal assessment, but are included because, as I flagged up at the start, they helped establish a context for engagement and response which is gradual and cumulative, and part of the total assessment framework.

First semester

Ice-breaker - paired introductions to the class

M20 noise levels - simulation and role play of a council meeting

Halloween headlines - summary and brief individual presentation **(F or S)**

Preliminary report - plan and outline of technical report **(S)**

Town planning exercise or skyscraper project - group projects **(F or S)**

IT skills exercise - promotional folder for building supply firm **(S)**

Final report - completed individually researched report **(S)**

Second semester

Drawing exercise - individual drawings critiqued by external specialist **(F)**

Individual professional presentation - based on researched technical report **(S)**

Convincing the client exercise - group work simulation and role play **(F or S)**

Site visit report - report includes notes and sketches **(S)**

Mock public enquiry - group work role play/simulation **(F or S)**

Here are a couple of the examples in more detail.

Preliminary report (S)

Research topic on some aspect of the construction industry chosen by each student in consultation with lecturer. Preliminary report provides working title with clear aim and scope; contents page and brief indicative outline of subject matter under main headings; referenced summary of work already carried out; and work still to be completed. Tight word limit (350-500). No restriction on number of figures/diagrams. This is **not** a scaled-down mini-report, but an outline of intent, and an earnest of progress so far.

(**Assessment**: Summative. Peer assessed in class on established criteria, with moderated assessment and feedback from lecturer/facilitator. Designated coursework: submission required for overall assessment in work of module. See Appendix A for a case study in streamlining assessment linked to this assignment, and born of necessity.)

Convincing the client exercise (F or S)

Role play/simulation team exercise - requires at least a two or three hour slot - where groups are told they are substituting for the original team of consulting engineers, held up in a car accident. They have to present to a prospective client a description of a project, and how it was constructed, as an example of what their firm can produce. They make a quick choice from a selection of projects (provided as multiple A3 copies of current articles from the *Proceedings of the Institute of Civil Engineers*, the leading professional journal). They have roughly 45 minutes to prepare and 10 minutes to present. Tutor advice is available on a 'request for consultation' basis, as with a senior colleague. Students present to remainder of class and lecturers as a collective 'client'. This is a challenging exercise for first year students, but they usually make a good fist of it, and their progress/achievement to that point should be underlined for them.

(**Assessment**: Formative or summative - marks are assigned. Groups are assessed and given feedback by lecturers on quality of ideas/technical understanding; clarity and structure of argument and expression; and individual presentation skills and technique. Lecturers also moderate students' self and peer-assessment on teamwork and group support. The peer allocation includes marks for individual contributions, following negotiation within the teams. Bonus marks are given to individuals for good questions/interventions and responses. Bonuses are given too for the quality of listed points justifying team performance. These are based on team conferral and are submitted after presentations. Lecturers give formative technical feedback and feedback on presentations before close, and at subsequent class. Final lecturer-assigned marks are agreed and compiled in a moderation session following close of exercise.)

Discussion and conclusion

The main thrust of this paper on streamlining assessment has been to suggest advantages to assessment conducted predominantly in class, with students actively participating, and designed to contribute to developmental learning. The student-centred approaches, which were innovative in the field of engineering education at the time of their introduction, have subsequently been used in other engineering, media and science-related programmes, and they should have broader application to other subject areas.

In a number of respects, it is obvious that there is a lot of activity here - this isn't a simple formula for doing fewer things and doing them faster - and there needs to be checks on the implications for staff and resources. But a proper consideration of the balance and spread of assessment as an integral part of whole curriculum design, with an emphasis on timeliness and productivity, can have a beneficial impact on all participants.

Sharing responsibility with students in drawing up criteria for assessed work reinforces the link between learning outcomes and assessment - and helps paves the way for peer and self-assessment. Our experience was that student involvement in peer and self-assessment meant that they frequently perceived the process as fairer and this was associated with a reduced number of appeals. Peer-assessment also tended to raise student participation levels for those activities in which it was used.

From the lecturer's point of view, moderating a mark is less onerous than being the prime marker and can be a more efficient use of a lecturer's time. Taking a careful look at what is designated as formative and what is summative is also important, and noting how students respond. Professor Cranston and I felt that peer-assessment was effective for giving a broad range of feedback on preliminary or draft projects and appropriate, as in the case of the site visit report when students had shared the experience, which was the focus for assessment, and helped determine the criteria for grading.

Establishing the right tone, especially setting and maintaining high expectations for students and lecturers - moving away from deficit models - is crucial for enhancement.

Prepare the ground carefully. When you introduce something new, however much you are persuaded about its advantages, make some allowance for uncertainty and even resistance from those it will affect directly.

Keep things simple, especially timetabling, and especially with mixed cohorts or degree strands. Try to ensure that students do the same things and are assessed by the same methods, so invidious comparisons are avoided.

Keep in touch. Student-centred learning may increase responsibility for students, but it won't reduce your responsibility in key areas. You need to be alert and well advised about how things are going at each stage, and to anticipate and make changes quickly when required.

Be prepared to wait for the benefits. Positive responses may take time to filter through. Don't delay or neglect evaluation of what you have attempted: you need to know how things have been received in order to plan adjustments.

Take pleasure in it. Students still respond to goodwill and genuine interest in a subject, and the satisfaction when you create that response and develop that interest is worth a great deal. Manage all that and even Blake might think the word 'genius' was not misapplied!

References

Cranston W B and Hutchison A N (1993) Experiences with Peer and Self Assessment, *Proc Conf: Innovation and Change in Civil Engineering Education*, Belfast

Cranston W B and Hutchison A N (1997) Integrating Communication Skills into Civil Engineering Education, *Proc Istn Civ Engnrs Civ Engng*, 120, Feb, 36-37

Davie G (1964) *The Democratic Intellect: Scotland and her Universities in the Nineteenth Century*, second edition, The University Press, Edinburgh

Heywood J (1989) *Assessment in Higher Education*, second edition, Wiley, Chichester

Hutchison A N (1996) Developing Interpersonal and Discovery Skills Within a Modular Framework, *Pro: 3rd East-West Conf on Engineering Education*, Gydnia, Poland

Hutchison A N (1997) Developing Interpersonal and Discovery Skills Through Student-centred Learning. *Case Studies in Student-Centred Learning: A Handbook of Readings*, University of Paisley

Hutchison A N (2004) Streamlining Feedback, *Student Enhanced Learning Through Feedback (Universities Scotland/LTSN Project)* www.heacademy.ac.uk/assessment/ASSO420_SENLEF_StreamliningFeedback.doc

Matthew R G S and Hughes D C (1993) Engineering Education: Time for a Rethink? *Proc Conf: Innovation and Change in Civil Engineering Education*, Belfast

Polanyi M (1958, 1998) *Personal Knowledge: Towards a Post Critical Philosophy*, Routledge, London

Appendix A

Case study: providing feedback in class

This case study focuses to some extent on a successfully shared agenda, but it was a product in the first instance of dismay and overload on my part.

At the end of 2000, dealing with first year engineers in a communication class, I found myself with a big load of (about 60) technical reports to mark and hand back. I was swamped with other things, but I needed a quick turnaround, and since they were preliminary reports for a larger writing project there needed to be useful and detailed feedback. What to do?

Realising I didn't have time to give all the students my usual comments on their texts, nor to fill out the grading sheets with their detailed criteria, I stalled and improvised. I took the unmarked reports into class - a two hour session - divided the 40 or so present into groups of six or eight, set them an activity to check on topics for an upcoming group project, which would keep them busy in the library or computer laboratories for the duration of the class (except when they were required) and arranged a schedule of brief tutorial sessions starting there and then.

For roughly 15 minutes at a time I went through the reports of each group individually, giving verbal feedback on all of them to the complete group based on a quick perusal and sampling. I commented on impressions of content, language, structure, presentation, referencing. I was frank within the usual limits, asked questions, drew in the others even when discussing individual essays, invited responses and defences. After these exchanges, I held onto the scripts and promised that they would receive grades and further individually tailored written comments the following week, but they were to continue with the project in the meantime based on the information gathered there. The following week I checked on their suggested project topics and firmed up the composition of the groups.

The students took this all very well and the standards on the final reports in several respects were better than in previous years.

I feel that this makeshift pedagogy had worked out satisfactorily in a number of ways. It had been a more efficient use of my time - marking one of those reports would usually take between a quarter and half an hour - and the students seemed to me to have benefited not only from personal feedback, but also from feedback to others. In addition, since there was an element of, if not peer assessment, then at least peer scrutiny, they were inclined to pay close attention. I didn't have to repeat myself unnecessarily - writing 20 times a note about proper headings for illustrations, or the format for acknowledging sources, say - but instead could reinforce with emphasis.

The fact that they were **getting feedback before they received a grade** probably also contributed to their taking more note of what was said, counteracting a common tendency to check the mark and ignore the rest. I made a resolve for future practice on giving feedback for draft or preliminary projects: bunch them up, do it in class if possible and appropriate, and make sure the students have a chance to contribute.

Ensuring quality and efficiency with online assessments

Dr Richard Parsons, Centre for Learning and Teaching, University of Dundee

Introduction

The University of Dundee is making moderate usage of computer-aided assessments (CAAs) to support learning throughout the University. These assessments are delivered online and are provided to students in both formative and summative modes. Given the importance of assessments and the costs of providing secure and stable online systems, the University has invested in central resources to provide these facilities to all departments. This approach has been important for ensuring the quality and efficiency for the underlying systems, and it has also been useful for promoting good practice within the learning activities that utilise the systems.

Support for online assessment by the central systems can be divided into six parts (Figure 1) and each is discussed below. This represents a systematic approach to supporting the enterprise, but does not include the important activities of staff developing questions and students interacting with the assessments. Specifically the success of using online assessment depends crucially on having quality questions that either stimulate student learning (for formative deployment) or provide accurate determination of student competences (for summative deployment). Creating these questions is an academic activity and this largely lies with the academic staff who may view this as an additional burden and therefore restrict the successful use of the systems.

At the University of Dundee, we have had some excellent success with innovative staff who are using online assessment to produce demonstrable increases in student performance in assessments. Another approach we are evaluating is to provide students with the opportunity to collaborate and develop revision assessments by each submitting a number of questions to a pool, which are then collated and made available in a formative fashion to the students to support their revision. Writing questions for online assessment that address the higher orders of learning - skills, competences, synthesis and analysis is possible, but it requires careful thought and commitment on behalf of the question author. We aim to increase our development and usage of these questions to support 'deeper learning' through online assessment.
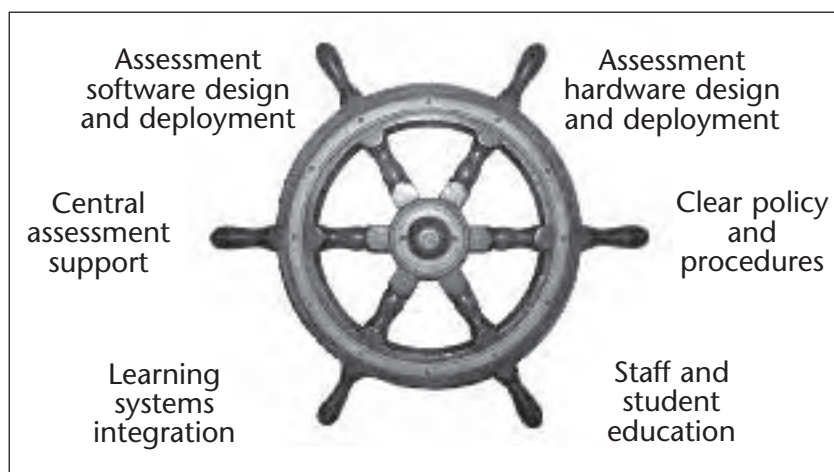

Figure 1 Components of a central online assessment system

Assessment software - design and deployment

The University of Dundee has a central virtual learning environment (VLE) that is available for all departments to use. This is presently a Blackboard version 6.0 system, and while the VLE provides some functionality for supporting online assessments, it also has some limitations when compared to specialist software tools that are devoted entirely to assessment. The assessment tools within the VLE are used by many staff to carry out small scale formative quizzes and they are also used for the end of module evaluations. For staff who wish to provide enhanced, flexible or summative assessments, the University has a full licence for Questionmark Perception™ (currently version 3.4). Frequently, staff become familiar with the utility of online assessment by using the tools within the Blackboard VLE and then subsequently progress to the more powerful and flexible environment that Questionmark Perception™ provides.

Authoring of questions on the Blackboard system is carried out using the internet, while we generally recommend that users of the Questionmark Perception™ software use the Microsoft Windows client. This system can operate on Macintosh computers using Virtual PC to create a Microsoft Windows environment. Academic authors receive full instruction in the use of Questionmark Perception™ software and online assessment through our staff education programmes (see below). Authors can create assessments and try out these assessments on their local computers, but uploading and configuration on the VLE and CAA servers is an activity carried out only by central staff in a gatekeeper role, again see below. It is important to note that various staff around the campus have widespread experience of CAA systems and some have been using Questionmark Perception™ for over nine years.

For the students, usage of the two systems is largely indistinguishable. Access to both Blackboard and Questionmark Perception™ assessments is provided from the VLE login, using a seamless transfer. Again the web interface provided by the Questionmark Perception™ software is more configurable and can be developed to provide specialist support for special assessments. The marks from the assessments taken in Blackboard or Questionmark Perception™ are automatically and privately available to the students within their Blackboard gradebook.

Assessment hardware - design and deployment

This is a specialist area and one that will only be briefly introduced here. The server systems that support the VLE and Questionmark Perception™ systems at the University of Dundee are typical of major information technology systems that are in place at many institutions. The University of Dundee has a policy of supporting Sun-Unix system platforms and Oracle databases for mission critical applications and these are the platforms that underpin the VLE and the Questionmark Perception™ databases. The Blackboard VLE is a dual server system with data storage managed through a storage area network using redundant array of independent disks five. Two Questionmark Perception™ application servers are deployed (one predominately for formative assessment delivery, and one for summative assessment delivery), and the formative server is coupled to a Unix Oracle database located on one of the VLE servers. The Questionmark Perception™ servers run on Microsoft Windows 2000 and use the Internet Information Service Recycle tool at 5am every morning to refresh

their memory usage. They have proved stable and reliable. Development of the production systems has now demanded the creation of test servers for both the VLE and Questionmark Perception™ systems, and these are utilised to test upgrades, new approaches and new installations, before subsequent deployment on the production systems. A diagram illustrating these server configurations is presented in Figure 2.
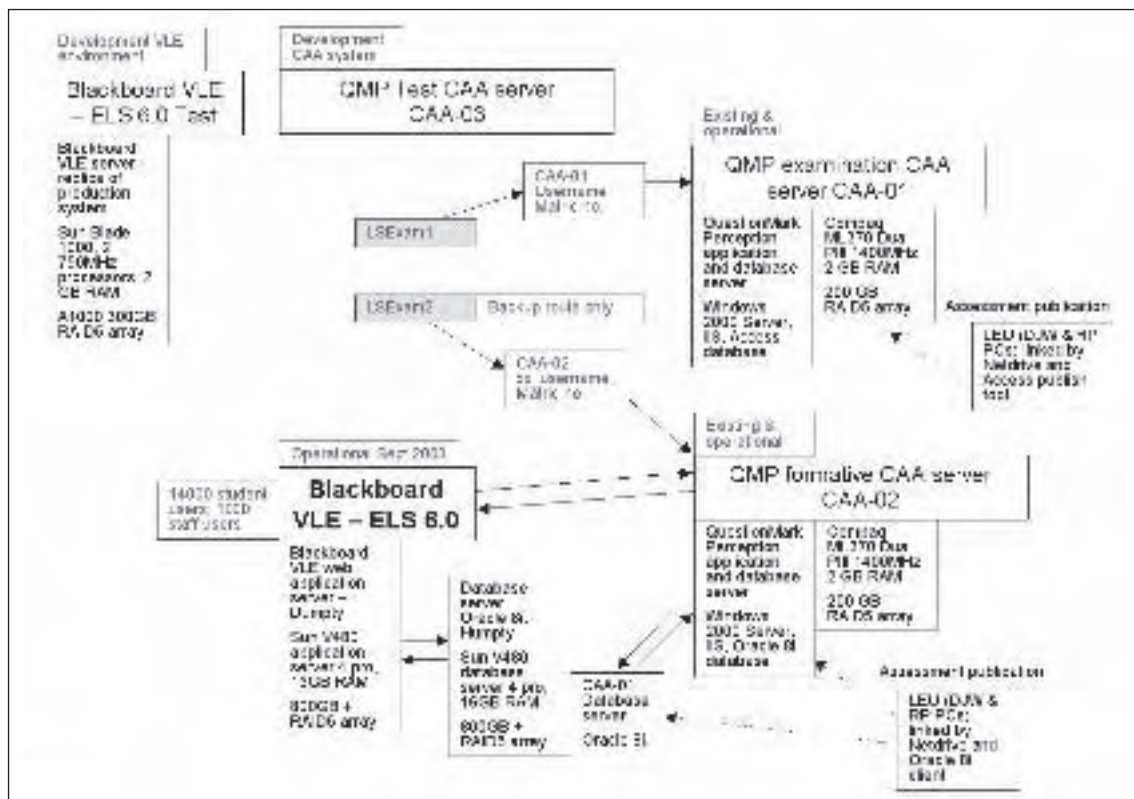
Figure 2 Central CAA server configuration at the University of Dundee (March 2004)

Clear policy and procedures

From the outset of developing a central assessment system, we created a draft policy and procedure document to ensure that a quality system was developed and understood by all. This document is designed to ensure that our systems and operations are compliant with the British Standard BS 7988:2002 - *Code of practice for the use of information technology (IT) in the delivery of assessments*. Our central principles are that assessments conducted with the CAA systems must be fair, reliable, confidential, accurate, secure, accessible and safe. The policy and procedures were originally introduced in draft format and they have now been passed officially at the University of Dundee Senate committee and now represent formal guidelines.

The policy and procedure documentation at the University of Dundee refers predominantly to summative assessments and students must have received formative assessments in a similar format to any summative assessments. While the details are largely relevant to the specific University of Dundee environment, the author is happy to pass on copies of these guidelines to assist others who may be developing their own.

Staff and student education

It is clearly crucial that both the students and staff are comfortable using the online technologies to receive their assessments and to support their learning and teaching.

Students entering the University of Dundee all receive an IT induction session of 90 minutes that includes a brief introduction to the Blackboard VLE and Questionmark Perception™. They complete a diagnostic assessment which is delivered by Questionmark Perception™. In addition, all students receive documentation to support their use of the VLE and CAA systems. In modules where the VLE and CAA systems are used, generally one of the lecturers will demonstrate how to access the material during a lecture and support is available from the IT Service's helpdesk. The vast majority of students are very comfortable and familiar with the IT environment and seem to have an expectation that the systems will be integrated and operate in a straightforward and reliable fashion.
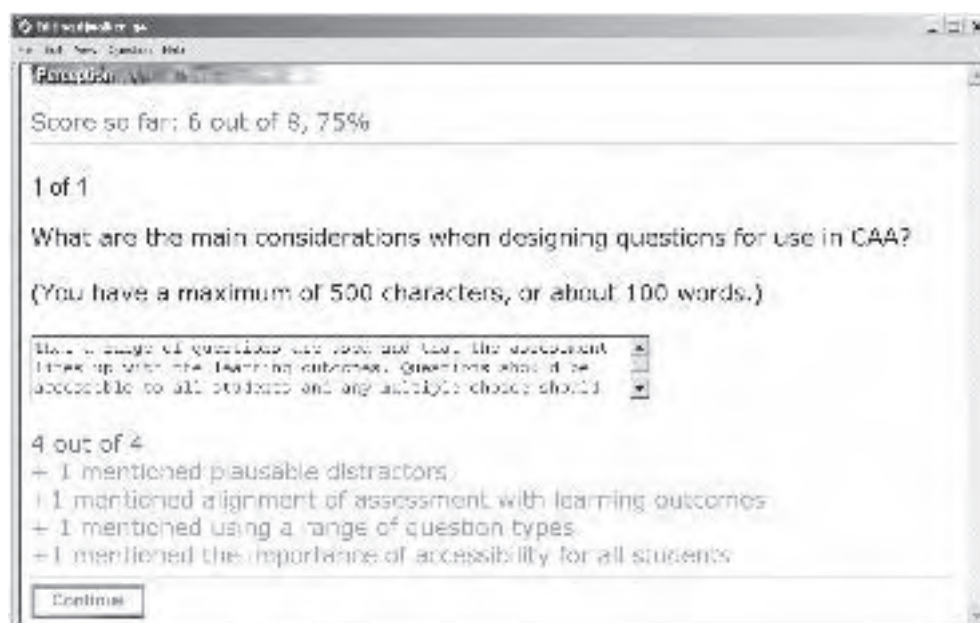
Figure 3 A text-matching question, automatically marked and provided with feedback

Staff at the University of Dundee have a more varied experience of IT and also have a more challenging task to complete as they have to construct the online material and author complex and specialist questions. Staff education is provided by the traditional face-to-face staff development activities, and these are sometimes focussed on a department where the staff have a particular interest in developing online assessment. The most successful staff education activity has been the development of an online course to support CAA and authoring within Questionmark Perception™. This course (CAA online) runs for five weeks, has been formally accredited at 10 Scottish Credit Accumulation and Transfer points by the Faculty of Education and Social Work and is run biannually. It is a demanding course where the two matters of pedagogy of online assessments and the technical authoring and delivery of assessments are examined. Typically those completing the course spend about five hours each week on the course material and are then well qualified to lead their students in online assessments. The course is open to staff from other institutions.

With a detailed knowledge of the benefits and limitations of online assessment, staff are free to develop innovative approaches for their teaching. An example of this is our online self and peer assessment system that has been utilised by staff in environmental sciences, medicine, education and accountancy. Using this system lecturers can set a text-based exercise for students to research and submit their answers online. The students are then presented with a set of marking criteria which they apply first to their own answer and then subsequently to two of their peers answers. The marking is completed anonymously and the lecturer can moderate and mark the work themselves if they wish. The results are emailed and texted (Figure 4) to the student author. Feedback from the students is very favourable as they find that they gain a much deeper understanding of the assessment process, particularly realising how criterion-based reference marking operates and how they can successfully predict the marking criteria. This system is available for lecturers to use within and beyond the University of Dundee at www.dundee.ac.uk/learning/ilt/

Figure 4 Online peer-assessment of text using software developed at the University of Dundee

Learning systems integration and Dundee specialisations

As VLEs and CAA systems are relatively new, integrating these systems is also a new activity, but one that pays dividends in that the resultant systems can be straightforward and consistent for users. The University of Dundee was one of the first institutions to install the Blackboard 6/Questionmark Perception™ bridge and the utility of this integration has proved worthwhile. Student use of the system is particularly straightforward and they are only challenged once for username/password when they log into the VLE, and all other username and access information is passed automatically between the systems.

Figure 5 Drag and drop question with visual cues and feedback designed to support learning

Some development work has been carried out to create tools to summarise assessment results in Questionmark Perception™ for subsequent direct upload into the Blackboard VLE. In this way, low stakes assessment marks can be based on any of the maximum, first, mean or final mark obtained in an assessment. This has proved important as it provides for flexibility in the system, allowing lecturers to design the deployment of their assessments as appropriate for the activity being completed. For example, some assessments where 20 questions randomly selected from a pool of 400 questions, are open for multiple attempts over a two week period and the maximum mark achieved is taken as the final score.

We have developed templates for the creation of questions in multiple response formats that allow for partial correct scoring and for numerical questions that include random variables within the question. We are developing an automated way of transferring questions with feedback from text to IMS QTI XML that is designed to allow questions to be directly incorporated into assessments. These tools are important to make the question authors task as technically straightforward as possible. They are also useful for allowing us to provide an efficient authoring service should students wish to develop their own revision assessments, or when authors provide questions for conversion to an online format.

Central support for online assessments

From the matters presented above, it is clear that online assessment systems require substantial commitment on behalf of those who wish to use them. At the University of Dundee, a decision was taken to provide these resources centrally, and to make the

software, servers, staff education and general support available to all faculties. The CAA operations are managed by staff in the Centre for Learning and Teaching, who have an academic teaching focus and are supported by IT Services who provide technical expertise. At present a single learning technologist provides specialist support to assessment authors, although this person is part of a group supporting the learning technology systems on campus. On the server and security side, the systems are supported by a team of people who support a large number of server-based systems. No single individual is responsible, but combined, they deal with issues of hardware, networking, operating systems, databases, storage area network, web servers, firewalls and security settings.

The central learning technology group operate in a gatekeeper fashion and are the only people with access rights to upload assessments to the CAA servers. This provides two purposes.

1 It frees the teaching staff of the need to learn about the upload and setup procedures for assessments.

2 It permits a quality assurance check, where general format, scoring and presentation can be double checked before publication to students.

This role of gatekeeper may be seen as restrictive by some, but it ensures the systems are stable and secure for everyone.

Attempting to detail the costs of deploying the VLE and particularly the CAA systems is of course challenging. Total cost accounting is not appropriate, as the academic staff, support staff, network, IT suites and many other infrastructure costs are required for other purposes, as well as for the CAA systems. Direct costs associated with the CAA systems are one staff member, software licences, hardware (lifetime of three years) and other directly attributable support costs. A fair estimate of these costs is about £60,000, representing £4 per student, or 50p per assessment delivered (about 120,000 assessments are delivered each year at present). Undoubtedly the use of CAA has a positive influence on student learning and they provide an efficient mechanism of completing summative assessments across a range of disciplines.

Summary

The example of the deployment of an online assessment system at the University of Dundee described above clearly illustrates that these relatively new learning technologies have a role to play in the efficient presentation of quality assessments. They are certainly not yet appropriate for carrying out all university assessments, but they can play an important role in providing and efficient, flexible and interesting component of the assessment of student learning.

SPIDER and SQuID - Online assessment in the sciences

Dr Ian Thompson, Department of Physiology & Pharmacology, Institute for Biomedical Sciences, University of Strathclyde

Abstract

Online assessment offers a powerful and flexible way of delivering tutor lead or self-learning for either on-campus or remote use. It offers many advantages over traditional assessment methods, such as instant feedback, automatic marking, question randomisation, monitoring and self-paced learning. SPIDER is a bespoke virtual learning environment (VLE) system, with over 4,500 staff and student users, which has a wide range of features and tools including a comprehensive and feature rich assessment environment. SPIDER's tools make it possible to create and deliver both formative and summative assessments, as well as monitor learner usage and progress, and create/export reports. It is now common practice for such assessments to be an essential tool in the delivery of the undergraduate curriculum. The SPIDER assessment system will be demonstrated and typical use cases described for undergraduate science courses at the University of Strathclyde, where online assessments are frequently used as both compulsory and optional components of class delivery.

Introduction - What is SPIDER?

The Strathclyde Personal Interactive Development & Educational Resource (SPIDER) VLE is an in-house development of a web portal that supports learning, teaching and administration. The system has been in place since 1998, at which point it supported a single class. It has since been developed and expanded from a departmental resource to an enterprise level system able to support thousands of users. SPIDER is currently in use by all 10 departments in the Faculty of Science at the University of Strathclyde, catering to over 4,500 users, and supplying content and information for over 300 classes. In addition, two postgraduate courses run by the Business School are hosted on the system (150 staff and students with around 30 classes). SPIDER has also been installed at the International Medical University, Kuala Lumpur, Malaysia as part of the twinned pharmacy programme.

Why in-house?

Most commercial VLE systems are based around the concept of a 'class', which belongs to an individual lecturer or designer. There is no representation of this class in a degree, course or department, nor a sense of the class belonging to more than just an individual. Students are then added to each of the classes they are registered on for the academic session, generating a flat representation of their degree (Figure 1a).

SPIDER differs from this approach as it attempts to model the existing teaching structure, style and philosophy used at the University of Strathclyde (Figure 1b). The concept of degrees are represented by the creation of clusters which group together common classes and information. Classes belong to a department, which are grouped into a cluster. Users belong to one (or more) of these clusters in the same way as they would be teaching on a degree (staff) or were part of a degree cohort (students). For example, a user may be in the 'pharmacy cluster' or the 'chemistry cluster', reflecting their primary teaching or learning area. The cluster determines the default settings that the user will view, but users are able to access content in any cluster if required. For example, users can add an elective class which 'belongs' to another cluster to their 'myClasses' list, in the same way that they would visit and be taught by different departments as part of their overall degree. The concept of clusters also allows for an area above the 'class object' where 'cluster-wide' news and information can be disseminated, and non-class information (eg year handbooks) stored, made accessible, and/or targeted at class, year or cohort level. This model offers much greater flexibility in terms of class and content management by staff users in the cluster, for example, staff can be allowed to manage all classes in their cluster, or subsets, or a single class, to reflect their status as degree/class coordinator/tutor, without affecting classes or content in other clusters. In addition, they are able to contribute news and information to the general cluster area.
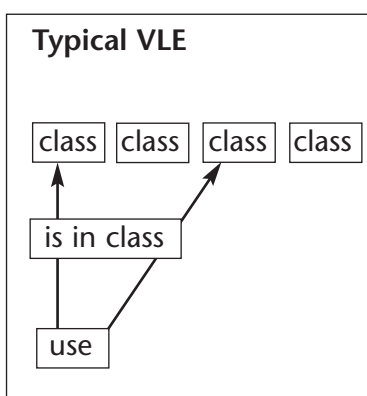
Figure 1a Typical VLE   **Figure 1b SPIDER Structure**
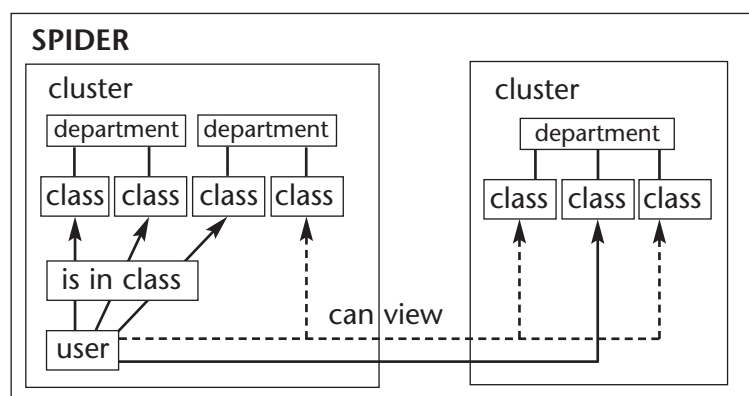
SPIDER also promotes an 'open learning' model, the majority of learning objects being visible to any user, regardless of the actual class and courses they are officially registered on. This information is stored separately and doesn't impact on the accessibility of learning materials. This not only allows students to access non-credit material, but previous years material (revision), as well as promoting staff peer review and good practice, by allowing staff to view each others classes.

The fact that the system is in house also allows for rapid customisation and development of new features and tools, rather than being reliant on the development cycle of a commercial product. It also allows simpler integration with other existing systems, and doesn't attract annual licence costs.

Technology

The SPIDER system is based on tried and tested technology comprising Linux as the operating system, the Apache web server and MySQL as the database platform. The programming development to integrate the systems is done with PHP, a powerful server-based scripting engine. The major advantage of this set up is that all the software is robust, widely used freeware, which is constantly evolving to meet the increasing needs of web systems. The system being in-house allows the feature set of the system to be customisable, and new modules are easily added to extend functionality.

What is SPIDER used for?

SPIDER, like any VLE, is primarily a tool to support and enhance teaching. It offers content management and creation, communication, peer or staff lead discussion/chat, targeted news and mailing list management, examination returns, student monitoring, and many other tools available in any modern VLE. SPIDER offers advanced assessment, monitoring and reporting tools, which are discussed below. In addition, having the source code available allows the feature set to easily be expanded. Recent developments include an electronic personal development portfolio, groups, file sharing, integration of departmental databases (safety, stock, assets, radiation), counsellee listings with an early warning system highlighting poor performance, a comprehensive poll and questionnaire system, laboratory attendance records, and a number of other modules. If anything else is needed, having the source code makes it possible to add it.

Assessment: SINQ or SWIMS

One of the major attractions of an online learning environment is the possibility of using assessment systems. SPIDER has its own assessment system which is made up of several components.

SINQ

The Shockwave InterNet Quiz viewer (SINQ, Figure 2) is a tool that allows learners to take assessments. It offers the learner free movement between questions and save and recovery at any point. For a particular assessment, the SINQ viewer can be run with or without feedback and/or marking.
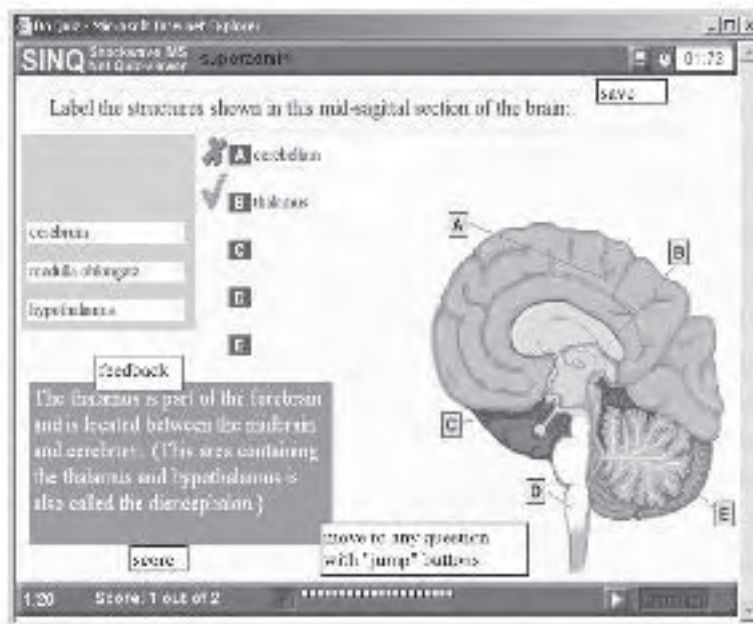
Figure 2 The SINQ viewer

SWIMS

The ShockWave IMS (SWIMS) question generator is a staff tool that is used to create question banks. Question types supported include multiple-choice questions (single answer, multiple true/false type), drag and drop/label or numeric questions that support randomisation and formulae. All parts of a question support HTML formatting, which allows varied formatting possibilities.
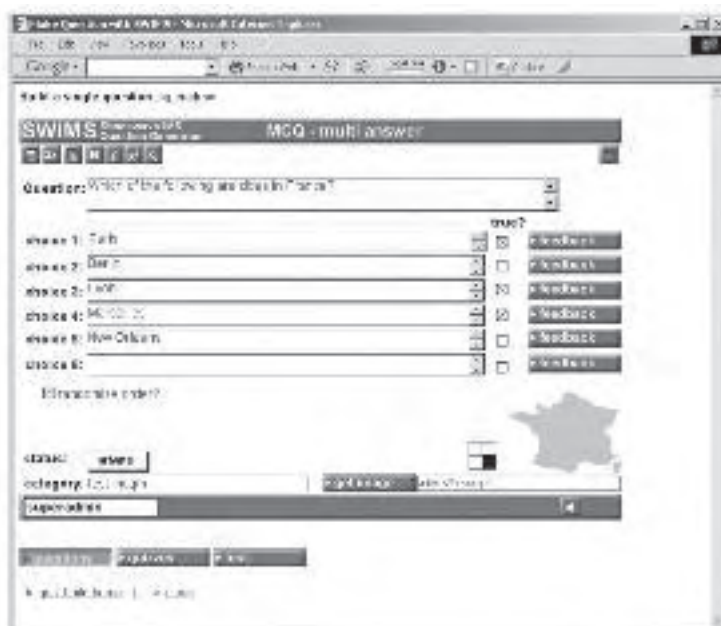
Figure 3 The SWIMS question generator

Once a user has created a variety of questions, assessments can then be built up from any number of questions, with various options such as order randomisation, sub-sets of questions, control over access (passwords, live/hidden) and number of attempts. Assessments can be assigned to one or more classes on SPIDER and run in a number of modes - formative or summative assessment - with or without multi-option feedback and marking.

Assessment monitoring - SQuID

Once assessment data has been accumulated, to be useful it needs to be viewed and reports generated. The Student Quiz Information Database (SQuID) is a fully integrated SPIDER module that is used for quiz/assessment monitoring and reporting. It records every response and attempt at any assessment and gives students and staff access to this data - students are limited to their own results, staff to assessments belonging to classes within their department/cluster.

Benefits to learners

The online assessments allow students to take part in self paced learning - they can start an assessment, save it then return to where they left off at a later date. They are able to access the resource at any time from anywhere they have access to the internet - 'any time, any place, anywhere' - Martini access. The system can be set to give instant marking, removing the wait for results, and offers feedback to reinforce learning. Multiple attempts allow for revision both in the current session, and due to SPIDER's open learning model, in later years of a students degree. If there is a large question bank associated with the assessment, different question sets can be presented each time the assessment is attempted, further increasing the usefulness and longevity of the assessment. The student also has access to their achievement history, allowing for self review and reflection, as well as acting as a general performance indicator.

Benefits to staff

Assessment through SPIDER offers many advantages to staff. Although there is obviously the initial outlay of effort in authoring the question banks and creating the assessments, time and effort is then saved through automatic marking and result accumulation. Reports can be generated detailing question by question responses for single students or entire classes (Figure 4), which can be exported into spreadsheets or other software for further analysis. This data can be used to identify subjects/topics, where students have difficulty or poorly designed questions, in more detail and far more conveniently than could be done with traditional assessments.

SPIDER also has tools that allow monitoring of usage and compliance, for example, SQuID allows staff to check who in a certain class has or hasn't attempted/completed any assessment. This information can be viewed for individuals or entire classes, from any academic session for which there is data. Questions and quizzes are re-usable and can be shared with others to gain maximum benefit from time spent on question authoring.

Figure 4 Example of response data for a single quiz - entire class would normally be shown on screen. Here responses for some of the top/bottom scores are shown.

Use cases

Two examples are show below of how online assessment is used in typical undergraduate science courses at the University of Strathclyde.

Use case 1 - 59102

This is a level 1 course, Physiology I, taught to undergraduate pharmacy students taking the MPharm degree. The typical cohort size is 120 students. The SPIDER assessment system has been used to create five formative quizzes, spread over semester one. Each quiz corresponds to one of the topics covered in the lecture part of the class. A deadline is given for each quiz, which falls after the completion of the lectures on that topic. The SQuID absentee monitoring tools are used to identify students who have failed to complete the quizzes, and an email query is sent to the student asking for a reason for non-completion. Continued non-completion of the quiz results in the student being 'NQ'd' and unable to sit the examination. Anecdotal evidence and student feedback suggest that the utilisation of these quizzes is beneficial to the student learning experience, students tending to spread their learning to coincide with the quizzes rather than cramming at exam time, as well as gaining some insight into their performance throughout the session.

Use case 2 - 09310

The Fundamental Immunology class is taught in the third year of joint honours BSc course. There are typically around 80 students in the class. The assessment of the class was changed in session 2000-01 from being purely a written examination to being 20 per cent essay, 15 per cent web exam, and 65 per cent written. In addition, five formative quizzes were made available to the students and their usage monitored. Initially, usage of the formative assessments was high, several attempts at the quizzes being commonplace. However, it has been noted that the number of attempts has dropped in recent years, perhaps as the novelty value of online assessment had worn off. Students coming in to the third year now often had experience of other quizzes on SPIDER. To try and counter this, the completion of the quizzes was made compulsory from session 2003-04. SPIDER was used to present and monitor both the formative and summative web assessments. The changes to the course made a dramatic improvement to student performance, the pass rate rising from 61 per cent in 1999-00 to consistently over 80 per cent since the changes to the class were made (89 per cent in 2000-01, 86 per cent in 2001-02, 85 per cent in 2002-03). Although this cannot be entirely attributed to the usage of online assessment, it will certainly have been a major contributing factor.

Success?

SPIDER is now a fundamental part of undergraduate learning in the sciences at the University of Strathclyde. The SQuID reporting and monitoring tools, along with the integration of assessment as core parts of class teaching, are part of the ongoing processes in improving teaching in the face of increasing student numbers and decreasing staff resource. Quality needs to be maintained, in both questions and feedback given, but once the initial outlay of effort has been completed, there is a reduced administration overhead compared to traditional, paper-based assessments resulting in improved use of staff time. Online assessment and learning thus offer a partial solution to the reduced staff contact hours and feedback students receive. This assessment system, along with the many advantages an online environment such as SPIDER offers (community, multimedia content, anytime access to learning, streamlined administration etc) can help to ameliorate the problems and demands put upon staff and students in the current academic climate, leading to improvement in the overall learning experience.

Links and contact details

SPIDER
http://spider.science.strath.ac.uk

SPIDER Development site - with guest access
http://spider-dev.pharmacy.strath.ac.uk

Ian Thompson
Department of Physiology & Pharmacology
Institute for Biomedical Sciences
John Arbuthnott Building
University of Strathclyde
27 Taylor Street
GLASGOW
G4 0NR
ian.thompson@strath.ac.uk
0141 548 2969

Streamlining assessment - how to make assessment more effective and more efficient - Post-workshop report

Professor David Ross, University of Abertay, Dundee and Workshop Director

Summary

The workshop was attended by 100 participants from all over Scotland. Overall feedback was that the day had been enjoyable, stimulating and challenging, with many participants saying they were going back to their institutions fully intending to enhance their own practice and influence the practice of colleagues - time will tell!

Keynote presentations

The first keynote speaker (Professor Ray Land, Coventry University) gave a wide-ranging and well-researched presentation on his view of today's environment for assessing students. Ray's natural enthusiasm, experience and command of the literature shone through and he managed to quote over a range from Rudyard Kipling to some of today's post-modernists.

Against a background of the effects of 'massification' (for example, over assessment, lack of advice on improvement and the slow return of coursework) Ray took us on a journey through some of the highs and lows of effective and efficient assessment and introduced participants to concepts such as control versus independence strategies in assessment.

He made some very salient and penetrating points regarding assessment. For instance, in trying to develop assessment practice which is both effective and efficient, it is worth remembering that 'What is efficient may not be effective and what is effective may not be efficient'. In other words, we must be on our guard against mixing these up rather than achieving an optimum balance which leads to enhancement of student learning.

The second keynote speaker (Win Hornby, The Robert Gordon University) took as his theme some of the key issues involved in trying to streamline assessment in a sector in which changes such as massification and reduction in resources have a complex and often negative effect. In a witty and professional way, Win gave participants six possible strategies for achieving streamlining, including automated techniques and 'front-loading', all of which have been shown to have some success. Several of these were thought to be particularly relevant to widening participation agendas in which the need to catch students 'at-risk' early in their studies is of paramount importance.

His analogy of assessment methodology ranging from Rolls Royce (low resource efficiency/high educational effectiveness) to 'stars' (high/high) via 'old double-decker buses' (high/low) and 'dogs' (low/low) was thought-provoking. It was viewed as an intriguing way of classifying assessment practices and tied in neatly with Ray Land's efficiency/effectiveness dilemma.

In conclusion, Win posed participants a very relevant question. 'How will this save me time?' This issue is one which comes up nearly every time a discussion on enhancing academic practice takes place. His paper should be read in this light - adopting these new techniques can save us time!

Case studies

Sandy Hutchison (University of Paisley) gave participants some interesting pointers on how to streamline assessment based on his own experiences in higher education. He started from the premise that streamlining might imply to some as doing fewer things and doing them faster but went on to show that there was much more than this to the issue. He also reminded participants of one of the fundamental current concepts in enhancing student learning - the need to promote student-centred learning through active involvement and engagement in the learning process, including assessment - a timely reminder! In Sandy's own experience, four improvements that can be made relatively easily for enhancing and streamlining assessment were:

- clear marking criteria, developed with student input

- increased opportunities for assessment and feedback in class

- introducing systems for peer and self-assessment

- reducing assessment 'bunching'.

Ian Thompson (University of Strathclyde) outlined the SPIDER and SQUID forms of online assessment in use at the University of Strathclyde. Both of these represent a growing trend in streamlining assessment and offer the benefits of instant feedback, automatic marking, question randomisation and self-paced learning. All of these are important benefits not only in assessment but in overall enhancement of student learning. A strong point in the SPIDER system is that it is based on in-house technology and tailored to articulate with existing (non-online) teaching methods and philosophy. This gives it extra credibility and acceptance with students and staff. Ian outlined the principal benefits to students as including self-paced learning, instant feedback and self-reflection on past assessments. The principal benefits to staff were outlined as saving in time and effort through automatic marking and result accumulation, better evaluation through detailed report generation (another important concept in quality enhancement) and closer monitoring of 'at-risk' students.

Richard Parsons (University of Dundee) also outlined online methods of assessment and argued strongly for these as examples of streamlined processes, using the catch phrase 'for efficiency and quality - go online'. He further stated that three key resources were used at the University of Dundee to support quality and efficiency in online assessment as part of an integrated managed learning environment.

- a campus-wide interactive virtual learning environment

- comprehensive policies for management and operation of computer aided assessment

- online self and peer assessment to enhance student understanding of the assessment process.

This is an important strategic point. Without institution-wide strategies and policies for streamlining assessment, then any amount of innovative approaches will do no good in the long term and only benefit narrow and isolated groups of students.

Breakout sessions

Participants were divided into six breakout groups, each one being posed with a question or issue relevant to the concept of streamlining assessment. The following points were raised in the groups.

- There is a trade-off between efficiency and student autonomy in assessment methods and this needs to be looked at more closely.

- A strong difference of opinion shone forth in one group as to the real usefulness of most online assessment methods, which seemed to split along arts versus science discipline lines. This group went on to agree that well-constructed and different styles of questioning would potentially help in arts-based online assessment.

- Possibly incentivise first year students more by rewarding good performance in initial in-course assessments through exemptions from exams.

- One group emphasised and endorsed the close relationship between assessment and curriculum design, particularly in a discipline context.

- There is a need for more attention being paid to training students in assessment methods during induction, with front-loading of resources into this crucial period.

- There was a plea from more than one group for more sharing of practice (good and bad) across the sector, both for enhancement purposes and to minimise wasting resources through reinvention of wheels.

Conclusions and strategic points for future development of assessment

This was a thoroughly good day, enjoyed by most and an excellent start to the workshop series. The main points, which should be considered as potentially worthy of follow up, are identified throughout the text of this paper but major ones are:

- remember the primary objective in streamlining should be enhancing student learning, not cost-cutting per se

- any planned enhancement of assessment practise should be closely articulated with current priorities of widening participation and lifelong learning

- care regarding efficiency and effectiveness

- encourage a change of culture in the sector through:

    i   more events of this type

    ii more sharing of practice across the sector - perhaps through compilation of an updated inventory of assessment practice, along the lines of the Assessment Strategies in Scottish Higher Education project of the early 1990s

- demonstrating 'big wins' early for staff and students (eg disseminating more widely staff time and effort saving through online assessments).

Assessment workshop series - No 2

Using assessment to motivate learning

Contents **Page**

Using assessment to motivate learning - An overview

Dr Andrew Eadie, Glasgow Caledonian University and Workshop Director

There can be little doubt that assessment motivates learning, at least for the majority of students whose aim is to successfully complete their programme of study. However, this kind of motivation, while understandable and even desirable in some senses, is not necessarily the kind of learning which is the desired outcome of higher education. The difference in the motivation we would wish to produce has been characterised by various authors as either intrinsic or extrinsic motivation (eg Morgan et al, 1980; Miller et al, 1998). Intrinsic motivation is described as when students are seeking intellectual stimulation from their studies and extrinsic motivation is when students are more concerned about their grades or marks and their future employment prospects. There is a long held perception (Paulsen, 1908; Elton, 1988) that extrinsic motivation interferes in a harmful way with intrinsic motivation. This view has seen examinations as a necessary but harmful feature of learning and recognises that extrinsic motivation is a legitimate feature of students' attitude to higher education. Recent work by Jacobs and Newstead (2000) on a cohort of 300 undergraduate students showed two distinct groups with the different types of motivation. The work also showed that motivation dipped markedly in the second year of a three-year programme but rose again in the final year. Murphy and Roopchand (2003) have shown, using a questionnaire to measure intrinsic motivation (Whitehead, 1984), that there is a strong link between self-esteem and intrinsic motivation and have also found higher levels of intrinsic motivation among female mature students.

This difference between the two types of motivation has often in the past been used to justify the gradual move away from end testing by formal examination to the many types of continuous assessment now in use (Beard and Senior, 1980). End testing provides an incentive for students to consider what they have learnt over a term or semester and can sometimes lead to students developing a holistic view of the subject which had not previously been apparent. However, the critics of this type of approach would have the view that this is not the most efficient or reliable way to produce intellectual stimulation. On the other hand, coursework, particularly if it is designed, delivered and fed back to students with continuous improvement of the student's performance in mind, develops skills and abilities which can lead incrementally to intellectual stimulation while at the same time overall producing the same evidence of attainment and standards that can be achieved by examination.

Elton (1996) has examined the balance between intrinsic and extrinsic motivation and concluded that the type of intrinsic motivation which the academic staff would wish the students to obtain might not always be possible. Elton does however point to a possible solution for the reconciliation of the two types of motivation which has been shown to be successful, namely, allowing students to negotiate their own learning objectives and to be assessed in terms of those objectives. Elton also states that changing assessment methods and abolishing the honours degree classification system could also have a positive effect.

Keller (1983) provides a detailed study of motivation and its relationship to instruction. This model can equally be applied to assessment as being a subset of the instruction process. Keller's model of motivational design includes four requirements to produce

learning which is meaningful and challenging. Firstly, it must be interesting, thus arousing the intellectual curiosity of the learner. Secondly, it must be perceived to be relevant, in that it must contribute to the learner's goals in undertaking the process. Thirdly, it must give the learner an expectation of success. Finally, it should produce satisfaction in the learner in having achieved their goals. Keller also describes a number of strategies which can lead to the four requirements including:

- guiding students into a process of inquiry
- team working
- using methods where the requirements for success are clear and under the student's control
- using appropriate and constructive feedback given at a point where it will be valuable to the student.

Boud (1995) gives an interesting set of criteria on which to critically judge assessment. A number of these criteria have a significant impact on motivation and raise the following questions.

- What is the contribution to learning made by the act of the assessment?
- To what extent does the totality of the assessment portray what is important for learning?
- What are the consequences of assessment on student learning?
- Does the assessment encourage learning?
- Are students able to shape their own assessments?
- Does the assessment lead to an outcome which properly indicates the accomplishments of the student?

Among the key reasons identified by Brown et al (1995) for undertaking assessment is to help motivate students. They specify the reasons for the choices of what and how is assessed should not include tradition, 'it is the way it has always been done'; inertia, a lack of desire to change; and control, 'students should do as they are told'.

Finally, they state that in order to assess well, the purposes of an assessment activity should be properly understood by both learners and assessors and the assessment should be designed accordingly.

Race (1995) has identified the key processes for successful learning, the following of which are directly linked to motivating students:

- learning by doing
- feedback
- digestion and reflection.

Farmer and Eastcott (1995) have discussed ways in which the three processes can be applied in practice.

- Using discussion with students about the nature of learning in relation to the subjects being studied and making assessment criteria explicit.

- Developing patterns of assessment which encourage productive learner activity and feedback; learning and feedback in groups; using portfolios to encourage learner activity and feedback.

- Using assessment strategies designed to encourage students to help each other learn.

These points are reinforced by the earlier work of Elton (1988) who suggested that intrinsic motivation can be increased by assessment strategies which:

- treat students as individuals

- expect students to show individuality, originality and creativity

- allow choices and preferences in their learning

- allow students to negotiate the means by which they are assessed.

A good example of the some of the above points is a study by Mortimer (1998) who used self and peer-assessment of reflective practice as a means to increase motivation. Mortimer found that intrinsic motivation could be increased by self and peer-assessment which:

- contained reflective writing assignments

- placed emphasis on both the process and progress of learning

- made sure the students understood the purpose of the assessment

- had opportunities for group work

- provided lots of opportunity for self, tutor and peer feedback and was linked to the student's personal development.

Leach et al (1996) had similar findings from a study relating mainly to self-assessment.

A key element of using assessment to motivate learning is the effective use of innovative methods. Race (1999) has examined the rationale behind innovation in assessment. He firstly examines the failings of traditional form of assessment, in particular unseen examinations and coursework. Race then goes on to suggest that each new form of assessment should be tested against five questions.

- Can the innovation increase learning?

- Can the innovation make the assessment more valid?

- Is the new assessment method more reliable?

- Does the new method have a negative effect on the workload of staff?

- Is the new method more efficient?

If the answer to each of the questions is yes then a better learning experience will be provided with the students achieving more valid qualifications. In another article by Race et al (2000), the view is also expressed that a key element in successful motivation is to ensure that the students are not overloaded by assessment.

McDowell and Sambell (1999) have examined the student experience of innovative assessment and found strong evidence of a link between innovation and motivation but did express a note of caution in stating the link could be broken if the assessment asked too much of the student. In designing innovative assessment, McDowell and Lambert gave the following guidelines:

- consider the student workload

- maintain motivation throughout the process by taking positive steps including feedback

- introduce new forms of assessment carefully

- establish clear guidelines

- ensure the students understand the assessment criteria.

Ecclestone (2001, 2002) has studied motivation in outcome based systems and concluded that the most effective method of motivation is by increasing the amount of formative assessment. The work also highlights the need for more fundamental research in the link between motivation and assessment.

Conclusion

There is clear evidence that assessment can motivate learning in the intrinsic sense of stimulating intellectual curiosity. This kind of intrinsic motivation is undoubtedly difficult using traditional end testing examinations. Assessment which motivates students is likely to be achieved by tasks which are some form of coursework and is probably more achievable when the method of assessment is innovative and has therefore not been encountered by the students previously. Care must however be taken not produce tasks which overload the student or do not assess the learning outcomes of the programme.

References

Beard R M and Senior I J (1980) *Motivating Students*, Routledge and Kegan Paul, London

Boud D (1995) Assessment and learning: contradictory or complementary? in Knight P (ed) *Assessment for Learning in Higher Education*, pp 35-38, Kogan Page, London

Brown S, Race P and Rust C (1995) Using and experiencing assessment, in Knight P (ed) *Assessment for Learning in Higher Education*, Kogan Page, London

Ecclestone K (2001) Creating compliance or commitment? Outcome-based assessment and post-16 learners' autonomy and motivation, *NASD Journal*, 45, 21-28

Ecclestone K (2002) Construction, motivation and learning in outcome-based assessment systems, *Prospero*, 8, 31-40

Elton L (1988) Student Motivation and Achievement, *Studies in Higher Education*, 13, 215-221

Elton L (1996) Strategies to Enhance Student Motivation: a conceptual analysis, *Studies in Higher Education*, 21, 57-68

Farmer B and Eastcott D (1995) Making Assessment a Positive Experience, in Knight P (ed) *Assessment for Learning in Higher Education*, Kogan Page, London

Jacobs P A and Newstead S E (1980) The Nature and Development of Student Motivation, *British Journal of Educational Psychology*, 79, 243-254

Keller J M (1983) Motivational Design of Instruction, in *Instructional Design Theories and Models*, Reigeluth C M (ed) Laurence Erlbaum, New Jersey

Leach L, Neutze G and Zepke N (1998) Motivation in Assessment, in *Motivating Students*, Brown S, Armstrong S and Thompson S (eds) Kogan Page, London

McDowell L and Sambell K (1999) The Experience of Innovative Assessment: Student Perspectives, in Brown S and Glasner A (eds) *Assessment Matters in Higher Education*, Society for Research in Higher Education, London

Miller A H, Imrie B W and Cox K (1998) *Student Assessment in Higher Education,* Kogan Page, London

Morgan A, Gibbs G, Taylor E (1980) *Students Approaches to Studying the Social Science and Technology Foundation Courses*, The Open University, Milton Keynes

Mortimer (1998) Motivating Learning through Facilitating Student Independence, in Brown S, Armstrong S and Thompson S (eds) *Motivating Students*, Kogan Page, London

Murphy H and Roopchand N (2003) Intrinsic Motivation and Self-esteem in Traditional and Mature Students at a Post 1992 University in the North East of England, *Education Studies,* 29, 243-259

Paulsen F (1908) *The German Universities and University Study,* Longmans, London (translation of the German edition, 1902)

Race P (1995) What Has Assessment Done for Us - and to Us? in Knight P (ed) *Assessment for Learning in Higher Education*, Kogan Page, London

Race P (1999) Why Assess Innovatively, in Brown S and Glasner A (eds) *Assessment Matters in Higher Education,* Society for Research in Higher Education, London

Race P, Collins J, Capon M, Ali A, Snell D, Parrish D and Salama A (2000) How to Increase Student's Motivation, *Educational Developments*, 1, 20-21

Whitehead J (1984) Motives for Higher Education, *Cambridge Journal of Education*, 14, 26-34

Designing assessment to enhance student learning

Professor Phil Race, University of Leeds

(Adapted from an article, Why fix assessment? I wrote in 2003 for Buckinghamshire Chilterns University College, and published in their internal publication *Seminar.*)

Setting the scene

A friend recently wrote to me:

Assessment is such an immovable and institutionalised process because it is the tip of the cultural iceberg on which elitist, class-ridden societies seek to maintain the status quo. You get to the top by winning at the existing game (which in the UK includes public school exams and the Oxbridge system). At the top universities you then read for exams, and the exams prove that you can write as well. The university may be happy for people like you to challenge its teaching and learning provision - up to a point - but not its assessment provision. A third-class degree from a top university carries more cultural clout, and opens more establishment doors, than a first-class degree from most of the other universities in the UK.

I will argue in this article that assessment is indeed broken in the UK in secondary, further and higher education nowadays, and needs fixing. 'If it ain't broke, don't fix it.' 'Don't throw out the baby with the bathwater.' 'You don't fatten pigs by weighing them.' That's almost enough of clichés. These were around long before assessment became as broken as it is today. My suspicion is that the assessment problem is not just a British one. But perhaps other cultures have at least solved some parts of the problem - most other countries **don't** try to run a degree classification system, for example.

Task

Without looking at the footnote[1] on this page, can you work out a single link connecting all the following statements about separate occurrences relating to assessment?

- The denial of the possibility of re-doing an assessment because of an arbitrary limit on the number of attempts allowed.

- The learners were not in a position to learn the things the assessment procedures assessed.

- The assessment discriminated unfairly against a group to which the learner belonged.

- The assessment was not valid for the purpose for which it was being used.

- The assessment had inadequate reliability.

- Assessment rules and regulations had not been properly observed.

- The pass mark had been set inadequately.

[1] **Response to task**
The obvious link is that all of these are examples of poor assessment practice, but in fact the real link is that these have all been grounds of successful litigation in the USA against colleges or universities. (These examples are quoted by Wakeford in Fry H, Ketteridge S and Marshall S (1999) *A Handbook for Teaching and Learning in Higher Education: Enhancing Academic Practice,* Kogan Page, London).

In this article, I'd like to take you through some tough thinking about assessment, and encourage you to play your part in working towards making it better. I'd like to challenge the status quo of assessment in education. In other publications (please see 'further reading' at the end of this article) I have sometimes tried to do this in the conventional scholarly manner, or tried to help staff and students make the best of a bad job, but now I think it is time to appeal to hearts and minds about the action which needs to be taken, and not just to air intellectual reservations.

But we all try ever so hard!

I would like to assert at the outset that the vast majority of assessors whom I know approach assessment with commendable professionalism and bring to bear upon it all of the integrity, patience and care that they can. They spend a long time adjusting the wording of assessment tasks and designing criteria with which to measure the evidence which students deliver to them. Moreover, the decisions they make on the basis of this evidence are made carefully and painstakingly. Their good intentions are unbounded. But - the final cliché - the way to hell is paved with such intentions. Perhaps because assessors tend to grow gradually into the assessment culture surrounding us, it is not surprising that they can be unaware of some of the prevailing problems that dominate the scene. At workshops I often liken many of the stalwart efforts which go into designing and implementing assessment as 'fine-tuning the engine of a vehicle which is actually off the road, facing in the wrong direction, and has no wheels left upon it!'.

How is assessment broken?

Assessment should be valid, reliable, transparent and authentic. Anyone who cares about the quality of the assessment they design for students will say how they strive to make it so. We are also **required** to make assessment valid, reliable, transparent and authentic by the Qualifications and Curriculum Authority in secondary and further education, and by the Quality Assurance Agency for Higher Education in higher education. Most institutional teaching and learning strategies embrace these three qualities in the aspirations of colleges and universities. But hang on, why have we all got 'teaching and learning' strategies in our institutions? Why have most institutions got 'teaching and learning' committees (or indeed 'learning and teaching' committees - small difference)? Why haven't we got 'teaching, learning and assessment' strategies, or indeed 'assessment, learning and teaching' committees, which would be the way round I would name them? Because assessment is the weakest link, I suggest. It's much easier (and safer) to fiddle around with the quality of teaching or learning than to tackle the big one: assessment. It's actually quite hard to **prove** that some teaching has been unsatisfactory, but only too easy to demonstrate when something has gone wrong with assessment.

'Come on, Phil', you may argue. 'We spend half of our lives on assessment. We have assessment boards, exam boards, external examiners approving our assessment instruments and practices and moderating our implementation of assessment. We've spent ages fine-tuning the assessment regulations. We've got years of experience at making assessment better. What more could we possibly be asked to do?'

'Assessment is the engine which drives student learning' (John Cowan). 'And our feedback is the oil which can lubricate this engine' (Phil Race). But sometimes we're too busy assessing to give really useful feedback. And students are too busy getting ready for their **next** assessment to take any notice of our feedback on their **previous** one. And when we come to the most important assessments (summative exams, and so on) feedback isn't even on the agenda all too often. And what do we measure in these important assessments? 'That which we can measure' - not always what we **should** be trying to measure. It's far easier to measure students' achievement of relatively routine objectives, and much harder to measure their achievement of really important objectives. This led me to write over 10 years ago 'if you **can** measure it, it probably isn't **it**'.

'So it's still broken,' I continue to argue. I'd better explain a bit more. Let's go back to 'valid, reliable, transparent and authentic' for a while. Let's just clear up the meanings of these four words.

Validity?

Valid assessment: this is about measuring that which we should be trying to measure. But still too often, we don't succeed in this intention. We measure what we can. We measure echoes of what we're trying to measure. We measure ghosts of the manifestation of the achievement of learning outcomes by students. Whenever we're just ending up measuring what they write about what they remember about what they once thought (or what we once said to them in our classes) we're measuring ghosts. Now if we were measuring what they could now **do** with what they'd **processed** from what they thought it would be better. 'But we do measure this?' Ask students, they know better than anyone else in the picture exactly what we end up measuring. For a start, let's remind ourselves that we're very hung up on measuring what students **write**. We don't say in our learning outcomes 'when you've studied this module you'll be able to write neatly, quickly and eloquently about it so as to demonstrate to us your understanding of it'. And what do we actually measure? We measure, to at least some extent the neatness, speed and eloquence of students' writing. What about those who aren't good at writing? Or to be more critical, what about those students who have at least some measure of **disability** when it comes to writing?

In the UK, the writing is on the wall for us regarding any tendency for our assessment instruments and processes to discriminate against students with disabilities. The *Special Educational Needs Discrimination Act* (SENDA) is likely to cause us to have to make far reaching changes to our assessment just to keep it within the law. SENDA came into force in September 2002, repealing the 'education exemption' which had previously applied to the *Disabilities Discrimination Act* 1995 in the UK. SENDA requires us to make 'reasonable adjustments' so that no student should be unfairly discriminated against by our education provision, not least the assessment-related aspects of this provision. SENDA also requires these reasonable adjustments to be made in an **anticipatory** manner, in other words not just dealing with instances of discrimination when it is found to have happened.

This is a tricky situation, as in one sense the purpose of assessment **is** to **discriminate** between students and to find which students have mastered the syllabus best, and

least, and so on. If we're honestly discriminating in terms of ability, that might be legal. But if we're discriminating in terms of disability it won't be legal. But aren't they the same thing? Where does ability stop and disability begin?

For a long time already, there have been those of us strongly arguing the case for diversifying assessment, so that the same students aren't discriminated against time and time again because they don't happen to be skilled at those forms of assessment which we over-use (such as, in some disciplines, tutor-marked time-constrained, unseen written examinations, tutor-marked coursework essays, and tutor-marked practical reports). We're entering an era where **inclusive** assessment will be much more firmly on the agenda than it has ever been to date.

We now know much more about the manifestations of dyslexia in assessment, and are just beginning to work out the effects of discalcula, disgraphia, dispraxia, and so on. Many of us are beginning to realise for the first time that even in that packed lecture theatre, we do indeed have students with disabilities, not just the occasional student in a wheelchair, but perhaps a quarter or a third of our students who are affected at some times in their learning by factors which we don't know about, and which many of them don't even know about themselves. So is it ever going to be possible to be satisfied with the levels of validity to which we aspire?

So we're not really in a position to be self-satisfied regarding the validity of even our most-used, and most practised assessment instruments and processes. But this isn't new, we've used them for ever it seems. That doesn't make them more valid. But we're experienced in using them? Admittedly, that makes us better able to make the best of a bad job with them. But should we not be making a better job with something else?

Reliability?

For many, this word is synonymous with 'fairness' and 'consistency'. This one is easier to put to the test. If several assessors mark the same piece of work and all agree (within reasonable error limits) about the grade or mark, we can claim we're being reliable. Not just moderation, of course. Reliability can only be tested by blind multiple marking. Double marking is about as far as we usually manage to get. And of course we agree often enough? No we don't, in many disciplines.

There are some honourable exceptions. 'Hard' subjects such as areas of maths and science lend themselves to better measures of agreement than 'softer' subjects such as literature, history, philosophy, psychology, you name it. By 'hard' and 'soft' I don't mean 'difficult' and 'easy' - far from it. 'But multiple marking just causes regression to the mean' can be the reply. 'And after all, the purpose of assessment is to sort students out - to discriminate between them - so it's no use everyone just ending up with a middle mark'. 'And besides, we spend quite long enough at the assessment grindstone; we just haven't room in our lives for more marking.'

So why is reliability so important anyhow? Not least, because assessing students' work is the single most important thing we ever do for them. Many staff in education regard themselves as **teachers**, with assessment as an additional chore (not to mention those who regard themselves as **researchers** with teaching and assessing as

additional chores). Perhaps if we were all to be called **assessors** rather than teachers it would help? And perhaps better, if we all regarded ourselves as researchers into assessment, alongside anything else we were researching into? 'Students can escape bad teaching, but they can't escape bad assessment', says David Boud. Our assessments can end up with students getting first class degrees, or thirds. This affects the rest of their lives. Now if our assessment were really fair (reliable), we could sleep easily about who got firsts or thirds. The students who worked hardest would get better degrees and the students who lazed around wouldn't. This indeed is often the case, but most of us can think of exceptions, where students got good degrees but didn't really deserve them, or students who seemed worthy of good degrees didn't come up with the goods, so we couldn't award them to them. So perhaps it's not just that our assessment isn't too reliable, it's our discrimination that's sometimes faulty too.

Transparency?

One way of putting 'transparency' is the extent to which students know where the goalposts are. The goalposts, we may argue are laid down by the intended learning outcomes, matched nicely to the assessment criteria which specify the standards to which these intended outcomes are to be demonstrated by students, and also specify the forms in which students will present evidence of their achievement of the outcomes. There's a nice sense of closure matching up assessment criteria to intended learning outcomes. It's almost a shame that there's yet another problem: some of the **real** learning outcomes go beyond the intended learning outcomes. Patrick Smith (Buckinghamshire Chilterns University College) argues that these are the **emergent** learning outcomes. Some of them are unanticipated learning outcomes. And it could be further extrapolated that there is some tendency for the 'you know it when you see it' extra qualities which get the best students the best degrees are firmly embedded in their achievement of **emergent** learning outcomes, and their evidencing of these outcomes within our assessment frameworks.

Leave aside this additional factor and go back to the links between intended outcomes and assessment criteria. How well do students themselves appreciate these links? How well, indeed, do assessors themselves consciously exercise their assessment-decision judgements to consolidate these links? Students often admit that one of their main problems is that they still don't really know where the goalposts lie, even despite our best efforts to spell out syllabus content in terms of intended learning outcomes in course handbooks, and to illustrate to students during our teaching the exact nature of the associated assessment criteria. In other words, students often find it hard to get their heads inside our assessment culture - the very culture which will determine their degree classifications.

The students who have least problems with this are often the ones who do well in assessment. Or is it that they do well in assessment **because** they have got their minds into our assessment culture? Is it that we're discriminating positively in the case of those students who manage this? Is this the ultimate assessment criterion? Is this the difference between a first and a third? And is this the **real** learning outcome, the achievement of which we're measuring? And if so, is this stated transparently in the course handbook?

So, we're not too hot on achieving transparency either. In fact, the arguments above can be taken as indicating that we rather often fail ourselves on all three - validity, reliability and transparency, when considered separately. What, then, is our probability of getting all three right at the same time? Indeed, is it even **possible** to get all three right at the same time?

Authenticity?

This one seems straightforward. It's about (on one level, at least) knowing that we're assessing the work of the candidate, not other people's work. In traditional time-constrained unseen written exams, we can be fairly sure that we are indeed assessing the work of each candidate, provided we ensure that unfair practices such as cheating or copying are prevented. But what about coursework? In the age of the internet, word processing and electronic communication, learners can download ready-made essays and incorporate elements from these into their own work. Some such practices can be detected electronically, but the most skilful plagiarists can remain one step ahead of us and make sufficient adjustments to the work they have found (or purchased) to prevent us seeing that it is not their own work.

Plagiarism is becoming one of the most significant problems which coursework assessors find themselves facing. Indeed, the difficulties associated with plagiarism are so severe that there is considerable pressure to retreat into the relative safety of traditional unseen written exams once again, and we are coming round full circle to resorting to assessment processes and instruments which can guarantee authenticity but at the expense of validity.

However, probably too much of the energy which is being put into tackling plagiarism is devoted to detecting the symptoms and punishing those found guilty of unfairly passing off other people's work as their own. After all, where are the moral and ethical borderlines? In many parts of the world, to quote back a teacher's words in an exam answer or coursework assignment is culturally accepted as 'honouring the teacher'. When students from these cultures, who happen to be continuing their studies in the UK, find themselves accused of plagiarism, they are surprised at our attitude.

Where are the borderlines between originality and authenticity? In a sense, true originality is extremely rare. In most disciplines, it is seldom possible to write anything without having already been influenced by what has been done before, what has been read, what has been heard and so on.

We need to be much more careful to explain exactly what is acceptable, and what is not. While some students may indeed deliberately engage in plagiarism, many others find themselves in trouble because they were not fully aware of how they are expected to treat other people's work. Sometimes they simply do not fully understand how they are expected to cite others' work in their own discussions or how to follow the appropriate referencing conventions.

Why is now the time to move towards fixing assessment?

OK, there's a problem, but we've just not got enough time to fix it? Why haven't we got time to fix it? Because we're so busy doing, to the best of our ability, and with

integrity and professionalism, the work which spins off from our existing patterns of assessment, so busy indeed that we haven't left ourselves time to face up to the weaknesses of what we're doing? Or because we simply dare not face up to the possibility that we may be making such a mess of such an important area of our work? It can help to pause and reflect about just how we got into this mess in the first place.

A couple of decades ago, the proportion of the 18-21 year old population of the UK participating in higher education was in single figures, now it's over 40 per cent, and the Government waxes lyrical about increasing it to 50 per cent. When there was only five per cent, it could be argued that the average ability of those students who participated in higher education was higher, and they were better able to fend for themselves in the various assessment formats they experienced. Indeed, they usually got into higher education in the first place because they'd already shown to some extent that they'd got at least a vestigial mastery of the assessment culture. Now, there are far more students who haven't yet made it in understanding our assessment culture, let alone gearing themselves up to demonstrate their achievement within it.

At the same time, when we were busy assessing just a few per cent of the population, we had time to try to do it well, using the time-honoured traditional assessment devices at our disposal. Trying to do the same for five or 10 times as many students is just not on. We can't do it. We can't do it well enough. We're assessing far too much to do it reliably, for a start.

And what about the students? Their lives are dominated by assessment. The intelligent response to this (thank goodness our students remain intelligent) is to become strategic. In other words, if there aren't any marks associated with some learning, strategic students will skip that bit of learning. If it counts, they'll do it. It's easy to go with the flow, and make everything important 'count' so that students will try to do all of it. But in the end this just leads to surface learning, quickly forgotten as the next instalment of assessment looms up. We're in danger of using assessment to **stop** learning instead of to start learning. It's no use us bemoaning the increased extent to which students have become strategic, when our assessment is the cause of this.

Who owns the problem of fixing assessment?

We can only ever really solve problems which we own. But the assessment problem is so widely owned. It's dangerously easy to feel there's just nothing that we can do about it. It's easy enough to identify scapegoats, including:

- professional bodies, in whose name we feel we need to stick to the status quo
- pre-university education systems, which cast the die and train pupils into particular expectations of learning and assessment
- institutional, faculty and departmental assessment regulations, which limit our room for manoeuvre
- teaching and learning strategies, which are so all-encompassing that we can't suspend belief and start afresh again
- heads of department or school, who are often seen (sometimes seen wrongly) to be content with the status quo

- external examiners who would have to be convinced when radical changes may need to be made

- students themselves who could or would complain about rapid changes to the level of the playing field or the position of the goalposts (even if the whole is enveloped in thick fog at present)

- the world outside academe, where there's a view about what a graduate should be, and so on

- journalists, broadcasters and editors who would give us a hard time if anything were to be found wrong in the way we did the job we are paid to do

- politicians and policy-makers who got to where they are by succeeding in the system of assessment we already have, and dare not admit that it might have been flawed

- parents, employers, taxpayers and others who foot the bill for education.

However, if we're perfectly frank about it, each assessment judgement is almost always initially made in the mind of one assessor in the first instance. True, it may well then be tempered by comparisons with judgements made in other people's minds, but to a large extent assessment remains dominated by single acts of decision-making in single minds, just as the evidence which is assessed is usually that arising from the product of a single mind at a given time within a given brief. Living on a crowded planet may be a collaborative game, but we tend to play the assessment game in predominantly singular circumstances, and competitive ones at that.

The fact of the matter is that to fix assessment will require individuals to change what they do, but that won't be enough to change the culture. Teams of individuals with a shared realisation of the problem will need to be the first step.

How can we fix assessment?

We need to work out a strategy. But any strategy has to be made up of a suitably-chosen array of tactics. Sometimes it's easier to start thinking of the tactics first. What could be a shopping list of tactics to play with for starters in this mission. They include:

- getting students into our assessment culture, by using peer-assessment and self-assessment more, so that they are better within our culture when **we** assess them

- reducing the quantity of assessment (say by a factor of three) so that we have time to do it well, and students have time for their learning not to be completely driven by assessment

- increasing the quality of assessment, so that it is fit for purpose, and more valid, more reliable and more transparent

- increasing the diversity of assessment instruments and processes, so that student casualties (where particular students are discriminated against repeatedly by the same old assessment formats) are not so frequent

- training (yes, training, not educating) our students to be better-able to play the game of working out where the goalposts are, and practising how to demonstrate their achievement of our intended learning outcomes.

'But I'm paying you to assess me'

It's all very well hoping to gain the benefits of involving students in self and peer-assessment, but the same UK Government initiative referred to earlier in this article has as a major proposed feature making students pay more for their higher education experience. 'Top-up' fees are proposed to increase from £1,000 per year to up to £3,000 per year in higher education, to be repaid by students once successful and in employment. It would not be surprising for them to turn back on us when we try to involve them in assessment, and tell us it's our job to do it.

So what can you do to fix assessment?

Turning tactics into a strategy is a big job, and beyond the scope of a short provocative article such as this. However, that big job won't even get started unless people are convinced that it needs to be done, and that was the purpose of this article. My aim was not on this occasion to write a scholarly article repeating what wise people have already written about in the literature (for years and years now). My intention was to employ challenging language to convince you that you've got a problem. What are you going to do about it?

Acknowledgements

Publications on assessment

I continue to work on both sides of the fence, helping students prepare to give assessment their best shot and helping staff to design assessment well. All of the publications below focus on assessment from one or other of these perspectives.

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

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

Race P (1992) *500 Tips for Students,* Blackwell, Oxford

Race P (1999) *Enhancing Student Learning*, SEDA Special 10, SEDA Publications, Birmingham

Race P (1999) *How to Get a Good Degree,* Open University Press, Buckingham

Race P (ed) (1999) *2000 Tips for Lecturers*, Kogan Page, London

Race P (2000) *How to Win as a Final Year Student*, Open University Press, Buckingham

Race P (2001) *The Lecturer's Toolkit,* second edition, Kogan Page, London

Race P (2001) *A Briefing on Self, Peers and Group Assessment,* LTSN Generic Centre, London
www.heacademy.ac.uk/resources.asp?process=full_record&section=generic&id=9

Race P (2001) *Assessment: A Guide fo Students,* LTSN Generic Centre, London
www.heacademy.ac.uk/resources.asp?process=full_record&section=generic&id=4

Race P and Brown S (2001) *The ILTA Guide,* Institute for Learning and Teaching in Higher Education, York, in association with EducationGuardian.co.uk

Additional references

Fry H, Ketteridge S and Marshall S (1999) *A Handbook for Teaching and Learning in Higher Education: Enhancing Academic* Practice, Kogan Page, London

Assessment driven learning

Dr Jean Cook, School of Computing and Mathematical Sciences, Glasgow Caledonian University

Abstract

Now, more than ever, assessment in education governs what students learn.

Assessment driven learning is a delivery method in which the assessment is an integral part of the learning process. In the form described here it was introduced at Glasgow Caledonian University under a Scottish Higher Education Funding Council (SHEFC) initiative to improve access to higher education and involved first year mathematics modules, but the method could be applied to other disciplines. The most significant aspect of this method of delivery is the continuous quality feedback provided to both staff and students. This improves student motivation, and, as a consequence, student performance.

Background

This paper describes a means whereby assessment can be made part of the learning process. Traditionally, assessment had three main functions: feedback for students; feedback for staff; and selection for future performance. All three of these outcomes are essential but in the first two, if only end of module assessments are used, the opportunity is lost for students to make good their deficiencies and for staff to improve their delivery at least for that group of students. It would seem sensible to add three further criteria for a good assessment. It should motivate students to engage in their learning, provide feedback early enough that their can remedy their deficiencies, and help them consolidate their learning. What will be described is a system of delivery termed assessment driven learning (ADL), in which the assessment is an integral part of the learning process and not the final verdict on the student. Students are given immediate feedback on their progress and staff get feedback when they can use it. ADL was used for modules which are quantitative in nature and computer-based assessments were used, but the technique of using assessment to drive the learning is not restricted to this scenario.

The initial implementation was launched at Glasgow Caledonian University under a SHEFC initiative to widen participation in higher education. For several years the pass rate in the mathematics module taken by computing students was unacceptably low. A significant number of these students came from postcode areas which had a low participation rate in higher education so it could be argued that it was appropriate to allocate funding under the initiative to these students. Prior to session 2000-01, many avenues had been pursued to improve the pass rate in this module. The syllabus had been pared down to essentials, at risk students who were identified early enough were directed to the pre-entry summer school, and a strict system of reporting attendance had been instituted. These measures did not significantly improve the pass rate. Staff were demoralised because most felt that the students had the ability to succeed but were not motivated enough to put in sufficient effort. The opportunity to change the teaching and learning methods from the locally accepted standard of three lectures and two tutorials a week, and a 30:70 split for coursework and final examination, was presented by the SHEFC initiative. The pilot scheme was trialled with two modules; the mathematics module for computing students and an algebra module taken mainly by mathematics students. Prior to 2000-01, both modules had some staff in common and were taught in the same way, with the coursework element taking the form of three closed book assessments and the completion of selected computer-based learning lessons. However, the algebra module had an acceptable pass rate. This may have been due to the superior mathematical maturity of the algebra students who, it seemed, were better able to recover and learn from a poor result in the first piece of coursework than were the computing students. The algebra module was included in the pilot study to see what effect ADL had on the pass rate.

The reasons for the poor pass rate which prompted the move to ADL are familiar to most staff in higher education and are certainly not new. *Enquiry into student progression*, by the University Grants Committee in 1968[1], noted that large numbers of undergraduates left university after the first year and, in 1980, the lack of

---

[1] University Grants Committee (1968), *Enquiry into Student Progress*, HMSO

motivation among first year students was explored in detail by Beard[2]. The problems faced by students and lecturers increased when the student grant was phased out. Most students now work to earn money to support themselves and, as a consequence, attendance at lectures and tutorials has fallen. The percentage of school leavers entering the sector has grown from six per cent to 50 per cent over the past 40 years. The resulting increased class sizes have meant that contact between lecturers and students decreased at a time when the background and skills of the student body was falling. Lectures to large groups now satisfy only a handful of these students; this may be more true in mathematics than in other subject areas. Cap this scenario with the pressure from within institutions to improve pass rates, and the scene is complete.

For the subject area of mathematics, there are other factors. Because syllabuses have been pared down to essentials, it is now not acceptable for students to aim at achieving a pass by getting mark of 40 per cent in the examination by concentrating on the easy topics of the syllabus. A quick calculation shows that if they are required to choose five questions from eight on the examination paper, a 40 per cent pass indicates that they may only be familiar with 25 per cent of the syllabus, slightly less for a three from five question paper. Mathematics is, in the main, a linearly ordered subject. If students are required to continue their study in mathematics, a good working knowledge of first year material is essential for success in later modules. For those students taking first year modules as a service subjects, a similar argument applies. The syllabus content, be it logic, matrices and probability for computing students or statistics for biologists, provides essential background for later study in the degree programme.

Operation

Assessment driven learning involved changing both delivery and assessment. In the first semester of the 2001-02 session, lecture hours were reduced from three to two, and the lecture material was arranged in learning plans which contained notes, references to computer aided learning (CAL) material and computer-delivered tutorial questions. Tutor groups were restricted to 20 or 24, the size of the computer laboratories, and every attempt was made to limit the number of tutors who were involved in a module. The timetables were reorganised so the lectures were delivered early in the week and all tutorial groups were scheduled to have one of their tutorials on Thursday afternoon or Friday. Computer-based assessments were held in this end-of-week tutorial session in 10 of the 12 weeks of the semester. Each assessment was worth eight marks and completion of the CAL material was worth 20 marks. All assessments had to be taken and all CAL material attempted. A carrot was dangled. Exemption from the final examination was granted for a mark of 70 per cent or above. In this case, the mark achieved was the final mark for the module. Students who did not gain an exemption took a final, worth 50 per cent of the final grade; the other 50 per cent was half the mark achieved on the assessments and CAL completion.

[2] Beard R M and Senior J (1980) *Motivating Students*, Routledge and Kegan Paul

CALMAT Courseware[3]

The software used is CALMAT, which was developed at Glasgow Caledonian University. It consists of learning material in the form of lessons, each with tutorials and assessments. There is also a bank of questions, all with random parameters which can de accessed singly or assembled into tests. The management system allows tutors to create learning plans so that they can customise the material to the needs of particular students. They can monitor both the progress of students on the learning material and their performance on tests. Students can buy the system on a CD and use it at home and then merge their records with those on the university system.

Results

At the end of the first semester in session 2001-02, staff and student satisfaction was sufficient to justify the conversion of a third module for delivery in semester two of that academic year. In the data displayed in the following tables, the modules are named A, B and C. A is the module offered to the computing students, B is the algebra module and C is the second semester module which was quickly converted when the success of the pilot scheme was realised. The results obtained for the two sessions 2001-02 and 2002-03 are shown in Table 1, along with the pass rates for the previous two years for comparison. The second column gives the number of students who enrolled in the module and the third gives the revised number, where those students who did not do any of the assessments have been excluded. The next three columns show the number of exemptions and the numbers passing the first and second diet examinations. The next column shows the percentage gaining an exemption and the last two columns show the cumulative percentages of students who had passed the modules after the first and second diet. In all cases, the pass rate after the first diet was at least as good as the second diet pass rate in the two years prior to the introduction of ADL.

One of the benefits of ADL is the quality and frequency of the feedback to both staff and students. At any time during the semester, the student has access to his score and full details of work still to be completed. At the end of each assessment, the score is displayed and the student can then access the solutions to the questions they failed to answer correctly. At this point, they are encouraged to discuss with their tutor any misunderstandings they may have. Weekly information on student progress was available and students experiencing difficulties were identified early in the semester so that tutors were able to target help where it was needed.

---

[3] CALMAT webpage www.calmat.gcal.ac.uk

Module	Session	Start roll	Revised roll	Exempt	Pass diet 1	Pass diet 2	% exempt	% pass diet 1	% pass diet 2
A	1999-2000	107	96		66	5		69	74
A	2000-01	148	125		74	14		59	71
A	2001-02	133	117	97	5	0	83	87	87
A	2002-03	143	128	109	0	4	85	85	88
B	1999-2000	50	47		37	4		79	87
B	2000-01	40	37		29	4		78	89
B	2001-02	59	57	42	11	0	74	93	93
B	2002-03	38	34	24	5	0	71	85	85
C	1999-2000		79		48	3		61	65
C	2000-01		78		54	10		69	82
C	2001-02	151	141	95	21	3	67	82	84
C	2002-03	108	100	91	2	1	91	93	94

Table 1 Comparison of pass rates for three modules with ADL implemented in 2001-02 and 2002-03

The effect of not having to resit a mathematics module would, no doubt, contribute to improving the progression rates for the programmes as a whole. It is hard to quantify just how many students would have failed without ADL, but it is worth noting that the fee income of every home student who progresses is £2,850 per year.

A saving which is easier to quantify is the time that is saved because of the reduced number of first and second diet papers which staff had to mark. Table 2 shows the estimates of the number of papers in the three modules in the first two years of implementation and in Table 3 there is an estimate of the staff time saved and cash equivalent, using a conservative estimate of a marking time of 30 minutes per paper.

Module	Number of students	Number of exemptions	Number passing after first diet (including exempted students)	Estimated number sitting second diet*	Actual number sitting second diet	Exam papers not marked*
A	117	94	101	49	16	94+33=127
B	57	42	53	15	4	42+11=53
C	141	95	116	50	29	95+19=114
Total in 2001-02		**231**	**271**	**114**	**51**	**295**
A	128	109	109	54	9	109+45=154
B	34	24	29	25	5	24+20=44
C	100	91	93	15	7	91+8=99
Total in 2002-03		**224**	**231**	**94**	**21**	**297**

*Assuming 2001-02 and 2002-03 numbers and 2000-01 first diet pass rates

Table 2 Number of exam papers which did not need marking in 2001-02 and 2002-03

Module	Total number of unmarked exam papers*	Estimated total staff time saved (hours) at 30 minutes per	Estimated £ saved at £30 script per hour
A	127	63.5	
B	53	26.5	
C	114	67	
Total in 2001-02		**147**	**£4,410**
A	154	77	
B	44	22	
C	99	49.5	
Total in 2002-03		**148.5**	**£4,455**

*Assuming 2001-02 and 2002-03 numbers and 2000-01 first diet pass rates

Table 3 Saved staff time due to CALMAT assisted ADL delivery in 2001-02 and 2002-03

Conclusions

At the end of each semester, students involved with ADL have completed a questionnaire. The results have been overwhelmingly positive. Many of the computing students, who were the main target, asked their parent department to consider using this method for other modules. As expected, they noted that the exemption provided the motivation. Interestingly, they felt that they did not need the lectures, however, the staff were not so sure. The staff were also enthusiastic, despite the fact that the delivery of ADL was hard work. They appreciated the reduced amount of marking and felt the usual lecturer satisfaction when their students did well.

In general, second diet pass rates were achieved at the first diet and the effect of not having to resit a mathematics module no doubt contributed to improving the progression rates of the programme as a whole. Students were enthusiastic, staff felt satisfaction to see the increased pass rates and the service departments were impressed.

References

Beard R M and Senior J (1980) *Motivating Students*, Routledge and Kegan Paul, London

University Grants Committee (1968) *Enquiry into Student Progress,* HMSO, London

CALMAT webpage www.calmat.gcal.ac.uk

Assessment strategies - a case in law

Peter F Scott, School of Law & Social Science, Glasgow Caledonian University

Context

As law becomes more complex, as resources dwindle and as the profession places increasing demands on academics to deliver on 'graduateness'[1] mechanisms must be found to cover the syllabus, embed skills and inspire the students.

The academic must also find time to keep up with developments in their specialist field, while juggling, of course, an increasing administrative workload.

The problems with law teaching are not unique to the study of law and the case study is, hopefully, relevant to study of what used to be described as the humanities, and, perhaps, beyond even if only to provoke a chain of thought.

Giving students autonomy in choice of assessment topic in a coursework only format cannot be 'one size fits all' and, for many lecturers, end testing remains the mechanism of choice as the clearest demonstration of academic ability. One must also concede that for some professional purposes end testing remains de rigeur, but the case study approach is an attempt to explore alternative mechanisms for assessment in the context of student centred learning.

So far as law is concerned, preparation is all and accuracy is everything. After over 20 years in practice, the presenter recalls no occasion when he did not appear with the case prepared and with fully annotated notes spread out on the agent's table. Feats of short term memory are, for the practitioner, mere party tricks. So why we do we assess students in a way which benefits the short term memory of the 'crammer'?

Modern educational techniques and their efficacy have been, and continue to be, explored in an attempt to maintain a quality educational provision within a Wider Access Framework, and contemporary thinking has been greatly influenced by the published research.[2] The message appears to be that law teaching must be practice-based and workplace relevant - an innovation, for such it is in law, to be admired if the graduates are then better equipped to do a job of work at the end of the day.

The inevitable conclusion must be that the methodology adopted must embed skills, must stimulate student interest, must relate to the real world of practice and, preferably, should give the under-resourced and underpaid academic something approaching the will to continue academic life.

---

[1] This peculiar word appears to have surfaced of late and further details can be obtained from the UK Centre for Legal Education website. It appears to mean the skills which academics are supposed to embed so as to satisfy the demands of professional employers.

[2] For example, see Brayne, Duncan and Grimes (1998) *Clinical Legal Education: Active Learning in Your Law School,* Blackstone Press; Burridge, Hinett, Paliwala and Varnava (2002) *Effective Learning and Teaching in Law,* Kogan Page; Duncan (1997) On Your Feet in the Industrial Tribunal - a Live Clinical Course for a Referral Profession, *Journal of Professional Legal Education*.

Resource issues

One possible approach is to take learning into the workplace and the presenter has also explored this elsewhere in a paper on Lecturing in a Live Environment presented at the sixth annual Learning in Law Initiative at the University of Warwick in 2004 suggesting that students can be given live case studies as project work.[3] That paper is, however, very specifically orientated towards legal practice. To anticipate the bottom line, students can be assessed on coursework alone and select the actual topic of assessment within strict guidelines - designed to eliminate the risk of plagiarism insofar as possible - but it might be useful to address the backdrop against which this innovation is proposed.

Today we have larger classes than could have been contemplated, say, 40 years ago[4] allied to a massive increase in the body of relevant substantive law.

Textbooks were familiar friends and principles were pretty well established. The growth in statute law can be measured with a tape measure. Consider, for example, the raft of significant legislation in the late 1940s and look at the thickness of, say, Current Law Statutes for 1946 as compared with the width of the tomes for 2003.

How is one to cover this vast body of law, even with the support of the giant computer databases now on-stream in every law school?

Given the complexity of modern law, the teaching of legal principle has never been more important, but to cover this within very real time constraints requires a little bit more than 'chalk and talk'.

Identification of coursework subject areas

The specific module which has been selected is only one of several but is offered on a 'flip flop' basis whereby every second year the module runs for both third and fourth year students.

The students are, in essence, environmental engineers who have had only the most basic exposure to legal principle and, yet, the substantive law is both complex and highly technical.

The first task, accordingly, is to identify the need to know which, on analysis, is no different to that required of a legal practitioner.

- Know what to look up.
- Know where to find it.
- Be able to understand it when you do find it.
- Be able to use what is found effectively.

If the law cannot be communicated, then the knowledge is a wasted asset, so at what level must one communicate? In broad terms, it is suggested that this will require to be communication:

---

[3] www.ukcle.ac.uk/lili/2004

[4] The very first intake of the new full-time LLB into the University of Edinburgh in 1961, as the author remembers only too well, was a mere 40 individuals with a massive preponderance of men over women and a very high proportion of students from fee paying schools.

- to an audience of lay-persons who need to be aware of the central issues
- to a client or employer who needs a comprehensive overview in context - a macro analysis
- to a formal hearing on very fine points of detailed law - a microanalysis.

The medium for communication may, of course, be oral, written or graphic, or a combination but, in the real world, will never be in the form of an essay. It may be in the form of a report or an article for a professional journal or a written submission to Court or the giving (or leading) of evidence. In short, the so-called personal transferable skills have to be applied.

In no case should the instrument lack intellectual rigour and even where a poster presentation is used the academic criteria must be spelled out.[5]

Instruments of assessment

If each student has to self select a topic then mechanisms such as end testing and multiple choice are self evidently not available.

Essay writing may be the main stand-by assessment of choice for many lecturers and, although one should not quarrel with a tried and tested mechanism which has, historically, served academe well, it is, bluntly, not a skill per se which is applied in the workplace. In any event, it is not an attractive option if the only possible title is in the nature of 'All I know about…', which would be the case with a diversity of topics. Alternatively, the lecturer could supply titles for the chosen project, but with a largish class this would be too tall an order for even the most imaginative and widely read scholar.

So what instrument should be adopted for each of the three levels?

Raising awareness

For raising awareness the instrument adopted in the module in point is a poster presentation.

In other modules, Microsoft PowerPoint presentations are used and are combined with peer group assessment but, in a large class, this is inordinately time consuming, valuable and interesting though this approach can be for the whole class. The scenario here is that the student prepares a poster, a mere hook upon which a presentation can be based. Images and graphics are actually secondary to the text, to which the student must be able to speak authoritatively. Ideally, the posters are hung in the examination room where they are examined by the assessment panel. The criteria are actually set out in the module handbook. Each student then has a slot

---

[5] A paper on poster presentations specifically was delivered to a recent BEST workshop in Glasgow and may be obtained from p.scott@gcal.ac.uk. Short and sweet but hopefully to the point.

when they have to speak to the subject matter. Time consuming? How long would it take to mark an individual project? In fact, the assessment is an allocation of marks to a sliding scale and is pretty straightforward.

Most modest sized classrooms can take around 80 plus posters especially if central display boards supplement wall space. Lecturers may suffer from information overload but the assessment does not all need to happen on the one day.

Actually, with strictly enforced time limits, the time devoted to assessment is arguably less than for the marking of conventional end tests and 'frabjous day' evenings and weekends are not lost to the chore of marking exam scripts.

However, this is more than mere assessment as not only are the posters there for all students to look at but they may even form part of a permanent display for future students, with a long term benefit for future teaching.

The overview

For the macro-analysis, this is presented in report format. The report must be concise but comprehensive and requires the ability to present in lay terms a clear exposition of the law which would satisfy a workplace need to know.

This takes the place of the essay so there is no downside from the perspective of assessment time.

The advantage, however, is that each project is individual and makes for better reading than the sixtieth essay on an identical topic.

Better still, the individual nature of the project within explicit guidelines means that the risk of plagiarism is greatly reduced. Here, the project has to cover the implications for land, air and water pollution and the research has to be fairly wide but almost any topic could be prescribed as having to stretch over several aspects and the poor student is then forced into referring to multiple sources for the material.

Best of all, the individual projects are likely to produce a data base for the lecturer who now has a whole class of active researchers, not a few of whom are likely to come across sources new to the lecturer.

In-depth research

The microanalysis has to be in the form of an article, written to publishable standard and which homes in one particular aspect where the boundaries of understanding can be pushed back. This tests the ability to research in depth.

In this module, it would have to be admitted that much of the material issued in class represents a composite of the best student work over the years. If attribution is given, the students become an ongoing learning resource for lecturer and students alike. The practice of a class test has become redundant. Using this mechanism, each piece

of work, accompanied by the sort of diagnostic feedback beloved of external examiners, informs the individual student of their progress and enables each student to work towards a steady improvement.

The research element will normally be the last to be submitted - and should attract the highest proportion of the marks - and if the occasional student wishes to 'free-wheel' on the basis of marks already in the bank then so be it. In fact, most students are highly competitive and the standard is usually pretty high.

All three mechanisms are workplace relevant skills and, to return to the concept of 'graduateness', establish that the student will be able to explain the impact of the law in broad terms to a client, to prepare a detailed brief and to research fine points of law. The presentation is also, of course, a standard practice in commerce/industry as well as in continuing professional development.

Criteria in self-selection

Each subject will be different.

The module dealt with here is on environmental law but the presenter has other modules[6] where a similar approach would work remarkably well. It does not adopt this methodology but only because changes to assessment would require validation and the module concerned has a limited shelf life.

Demonstration by example is perhaps the best way forward at this juncture.

Each student has to produce an abstract indicating the chosen topic. The topic is to be assessed at the three levels already indicated but must also incorporate international law, European law and UK law. The combination of the three topics must also be seen rigorously to look for cross contamination of other environmental mediums, ie the interplay among land, air and water.

It is for the student to decide which of the three legal jurisdictions (international, European and UK law) will be the focus for which instrument of assessment.

Let us take the not entirely hypothetical example where a student is interested in pollution of waterways, perhaps they enjoy fishing or live beside a river, but it matters not so long as this is the subject closest to the heart of the student. In workshops, this topic is discussed with subtle guidance from the facilitator (the lecturer) and all are agreed that pollution of international waterways raises specific trans-border problems.

In the event, it might be felt that international law could be the subject of a broad-brush presentation which dealt with the significance of river basin management[7] and related topics which would bring in the significance of land misuse.

The significance here is that the students as a group became actively involved in a discussion as to how this topic could be handled to best advantage and, usually, several students will be able to link this project with their own work with clear advantages for peer group research support.

---

[6] One module on Construction Law, for example, has elements of Delict/Tort, Employment Law and Health & Safety Law.
[7] *The Wetlands Convention Cm 6464 [1971]* etc.

Dealing with European law, one is spoiled for choice and the group might agree that perhaps a micro report on European policy would enable reference to air pollution and the impact of acid rain on waterways and conservation of species.[8]

The advantage for the facilitator is that this leaves the UK perspective as having to deal with the recent adoption in Scotland of the Framework Directive[9]. This is a very complex body of law with very detailed provisions to be brought in over an extended period and which will look at water use rather than water discharge control as the principal focus. In this context students are a learning resource and a detailed analysis of the provisions of the Act represents, effectively, the services of a researcher to analyse an issue which will become increasingly significant.

The bottom line is that the topic is covered at all levels - and the student will be able to demonstrate the key workplace skills - including grasp of legal principle, communication skills, analytical ability and significant research skills. The group discussion element, since it involves students in meaningful dialogue, is likely to promote deep learning among the group on the view that there is a very clear point to the exercise and the subject is actually interesting. Law was being introduced, and of which the students were previously unaware, in a practical context.

In a large class, the workshop replaces the seminar so there will always be a discussion forum for manageable class sizes and where, in addition, one can deal with specific points of law where the solution may depend on fine interpretation of detail. Very often the research will produce quite surprising results and sometimes the data can even produce proof positive that the received wisdom is actually quite wrong. Everyone knows, for example, that oil spill is a major source of marine pollution, but everyone is wrong as, in truth, well over 70 per cent of oil pollution in the North Sea comes from land run off.

Learning and assessment

It is suggested that assessment should not be viewed merely as a mechanism for testing what has been learned but should, actually, be integral to the learning process.

Students must learn how to investigate and by selecting a topic which tests the research capability, the student not only acquires that workplace relevant skill but can, in effect, be adding to the store of knowledge. The class, as a whole, are informed as is the lecturer. From the perspective of the individual, the knowledge thereby gained should be much more securely embedded than would be the case with rote learning for end testing.

In fine, the learning activity becomes student-centred and the actual mechanism for assessment can be part of the instruction process.

---

[8] *Water Framework Directive [200/60/EC], Urban Waste Water Treatment Directive [91/271/EEC], Nitrates Directive [91/676/EEC] etc.*

[9] *Water Environment & Water Services (Scotland) 2003.*

Results

Results have always been good with this approach.

It cannot be appropriate in all cases and may even be part of an assessment regime which includes some end testing where, say, a professional body requires such.

In looking at a module where the sole mechanism has been this approach over a 10 year period, the progression rates for the module have been a consistent 100 per cent after the second diet (averaging 90 per cent after the first diet and always because of delay in submission). This compares with progression for the year as a whole of 85 per cent on average for third year students and 100 per cent for fourth year students.

In other words, it is good for progression.

Marks tend to be consistent with marks for conventionally taught modules on the same programme at the upper end of the marking scale and significantly improved at the lower end of the marking scale, even if it is only the difference between a marginal pass and a clear fail.

A second module on an LLM programme has been an unsurprising 100 per cent progression. Over four years, marks tend to be slightly above the average for conventionally taught modules.

Where marked differences also occur is where students who have been able to self select a topic for assessment are compared with students who have never had that option. Honours dissertation marks tend to be significantly above the class average, hopefully making the point that improved research potential is carried forward to other subjects.

The bottom line is that progression rates are improved, weak students achieve slightly better grades and strong students will tend to fulfil their potential across the board to better effect.

Conclusions

The methodology whereby students are given a degree of autonomy seems to improve progression without significant impact on marks at the upper end of the ability scale but does improve marks at the lower end.

Progression rates are greatly assisted and not only does the quality of work attract favourable comment from external examiners but the exercise of personal transferable skills is significantly improved.

Plagiarism is the bane of coursework, of course, but a highly individualised project must surely reduce the possibility, especially if the module leader has the ability to direct the main thrust as part of a pattern of consultation and discussion within a workshop format.

The debate on the value of end testing is fairly well rehearsed, but can we continue with end testing in the face of increasing demands for integration of workplace relevant skills into law teaching?

Further, given the higher percentage of the population entering further education - currently 50 per cent of school leavers in Scotland - there is likely to be a broader spread of student ability and it becomes increasingly important to look for mechanisms whereby weaker students will be motivated and encouraged to give of their best.

Quality demands that we look at innovation and here, at least, is an innovation, with a 10 year history, and one which seems to work!

References

Brayne, Duncan and Grimes (1998) *Clinical Legal Education: Active Learning in your Law School*, Blackstone Press

Burridge, Hinett, Paluirala and Varnava (2002) *Effective Learning and Teaching in Law,* Kogan Page

Duncan (1997) On Your in the Industrial Tribunal - a Live Clinical Course for a Referral Profession, *Journal of Professional Legal Education*

Motivating students through group project and open-notes examination

Ian Smith, School of Computing, Napier University

Abstract

The case study is based on learning, teaching and assessment (LTA) practice developed over the last few years at both undergraduate and postgraduate levels. Reference will be made to empirical data gleaned from examination boards and student feedback from the module review process.

The paper will discuss a level 7 module, Object-Oriented Multimedia Design & Development and a level 4 module, Advanced Digital Media. The LTA approach integrates assessment and feedback into the overall learning experience. Formative feedback on the group project assessment is provided on a weekly basis, with the summative feedback incorporated into the examination preparation. The open-notes/open-book examination is based on topics of directed study introduced through the lecture series and supported through a virtual learning environment.

The LTA approach has been demonstrated to have a positive impact on student engagement and individual achievement.

Background and context

In 2000, it was noted that a high percentage of the students studying the MSc in Multimedia and Interactive Systems programme offered by the School of Computing at Napier University were having difficulty with software engineering modules.

The programme is unique within the School of Computing in that it attracts a majority of students whose first degree is BA, rather than BSc or BEng. Our experience within the School of Computing, and that of other institutions[1], is that in general students from liberal arts backgrounds have difficulty in learning software engineering. This was a concern with respect to engagement and student achievement.

In June 2001, as part of a major revision of the MSc Multimedia & Interactive Systems programme a new module Object-Oriented Multimedia Design & Development (OOMDD) was conceived to address the issues.

Object-Oriented Multimedia Design and Development

The module endeavours to provide an enhanced learning experience to encourage students to take responsibility for their own learning; engage in active learning and student interaction; and develop as autonomous learners[2].

Rationale

The module has been designed to provide an enhanced learning experience that develops knowledge, skills and attitudes with respect to the development of interactive multimedia applications through a culture that values enquiry, investigation, research and reflection.

The learning and assessment engenders key employability skills.

- A project coursework provides students with the opportunity to design and implement an interactive multimedia application to their own specification that develops employability skills: creative problem solving, design, planning, organisation, and time management.

- A critical report develops employability skills: communication, organisational, and presentation.

- Group work develops employability skills: teamwork, leadership, negotiation, communication and management.

Delivery

The module utilises a traditional lecture/tutorial/practical format supplemented with web-based resources. In addition students are expected to do further reading, both directed and independent. The required text Using UML: Software Engineering with Objects and Components[3] is available for reference during the open-book exam at the end of the module.

---

[1] Anderson P, Bennedsen J, Brandorff S, Caspersen M E and Mosegaard J (2003) Teaching Programming to Liberal Arts Students - a Narrative Media Approach, *IT;CSE 2003 conference (Thessaloniki, Greece, University of Macedonia) ACM Special Interest Group in Computer Science Education (SIGCSE).*

[2] Little D (1991) Learner Autonomy, 1: *Definitions, Issues and Problems*, Authentik, Dublin

[3] Stevens P (2001) *Using UML: Software Engineering with Objects and Components*, Addison-Wesley

Learning support and feedback

Each week the students have the opportunity of attending a supervised one-hour tutorial session where the module tutor and experienced demonstrators are in attendance to provide technical support and answer queries.

Group work

The students form themselves into groups of three in week 1. The module tutor meets each student group for 15 minutes every two weeks to provide formative feedback on the progress of their projects. The meetings are scheduled to meet the needs of the students and a record is kept with counselling given to any group that fails to attend. Small group activity brings the students together, encouraging peer support and engendering engagement. On completion of their projects, feedback on the coursework is provided via a feedback sheet and discussed during individual group conferences timetabled to facilitate exam preparation.

Learning resources

All of the required teaching materials can be accessed from the virtual learning environment. This also includes links to websites of interest, electronic copies of the module organiser, handouts, coursework specifications and additional materials.

Additional resources are available from the Multimedia & Interactive Systems Resource Centre developed by the author, with specific pages dealing with object-oriented analysis and design[4].

Assessment

Biggs[5] proposed that the assessment must align with the learning objectives of a course to encourage deeper learning. The module descriptor establishes a direct mapping of the module learning outcomes to the assessment strategy.

The assessment for the module comprises both a coursework and an examination. The coursework assessment is developed as a group project and as a consequence produces homogeneous results; therefore an examination is included to allow the student to demonstrate individual learning.

Coursework

As a member of a group, the students are tasked to develop a game or an interactive simulation to their own design and specification suitable for implementation on the internet. The students are supplied with materials on game design. Each group is required to produce a detailed object-oriented design model from a requirements specification, using object-oriented analysis and design methods; develop a prototype from the object-oriented design model, using Macromedia Flash MX; and provide a critical evaluation of the object-oriented design incorporating reference to additional reading, and alternative methods and solutions.

---

[4] Multimedia & Interactive Systems Resource Centre www.dcs.napier.ac.uk/~mm/ooa.htm

[5] Biggs J (1999) *Teaching for Quality Learning at University*, Open University Press

The students determining and managing their own projects encourage ownership of the assessment and learning process.

Group projects enhance employability skills[6][7] such as teamwork, leadership, creative problem solving, design, communication and management. Deep learning is encouraged, as the student is required to engage in active learning and student interaction in addition to taking responsibility for their own learning and assessment. However, group work can allow students to be awarded marks based on the work of others if appropriate measures are not introduced. Therefore, procedures were introduced to allow individual submission in the event of a complete breakdown of a group.

Essay-based coursework further develops employability skills[6][7] such as communication, organisational and presentation. Deep learning is encouraged, as the student is required to demonstrate a range of higher cognitive abilities in developing the critical report that in turn encourages the student to become a reflective practitioner.

Open-book examination

As the coursework assessment is a group project it produces homogeneous results; to balance this, an examination is included to allow the student to demonstrate individual learning.

Supervised assessments[6][7] are a secure method that minimise cheating and open book examinations assess critical thinking as opposed to memory. Deep Learning is encouraged, as the student is required to demonstrate a range of higher cognitive abilities. Many students perceive examinations as stressful, therefore appropriate preparation is provided via a two hour exam briefing and access to previous papers with worked solutions. A number of students ran out of time on the first delivery this was rectified by extending the exam from two to three hours to ensure that adequate time is available.

Discussion

The OOMDDM has been delivered for two academic years and has been given positive criticism by both students, by way of the module feedback questionnaires, and external examiners at module boards.

In academic year 2001-02, semester 2 with 77 students, the overall pass rate was 83 per cent; the mean was 58; the standard deviation was 9.7; the Dbar[8] was -0.8; and six students (8%) gained distinctions (=75%). Analysis indicated that individual student achievement correlated with their performance in the other modules studied during the same period. The LTA approach in relation to the module outcomes proved to be effective. However, the lecture sequence was changed to enhance learning and additional teaching and support materials were developed.

In academic year 2002-03, semester 2 with 45 students, the overall pass rate was 78 per cent; the mean was 59; the standard deviation was 14.8; the Dbar was -0.5; and seven students (15%) gained distinctions (=75%). Initial analysis indicated a positive correlation with individual student performance on other modules studied during the same period.

---

[6] Freeman R and Lewis R (1998) *Planning and Implementing Assessment*, Kogan Page

[7] Brown S and Knight P (1998) *Assessing Learners in Higher Education*, Kogan Page

[8] A method of comparing individual student results across modules to establish whether consistency occurs, 0 is ideal and within $\pm$ 4 is considered good.

The LTA strategy for the module has proven to be successful but appeared to be no more than that of other modules studied by the students. Analysis of the exam results with another exam sat by the students during the same period indicated a positive correlation (r = 0.7) with no significant difference in the means.

A comparison of the performance of students with that of another software engineering module again indicated a positive correlation (r=0.5) with no significant difference in the means. Though the approach is innovative and promotes student engagement, no indication of improved student performance seemed to be apparent.

However, further scrutiny of overall student achievement highlights significant differences between the modules taken by the 41 full-time students during the academic year 2002-03 as shown in Table 1.

Though individual student performance was comparable, the number of students submitting assessments varied dramatically across the modules. The students were more likely to submit assessments for the multimedia and interactive systems design (MISD) modules (95 per cent) than the software engineering (SE) modules (73 per cent). With respect to the percentage of students that passed the assessment once submitted, there is little difference across the teaching groups (MISD: 83 per cent, SE: 79.5 per cent). Consequently, students were more likely to successfully complete MISD modules (79 per cent) rather than SE modules (57 per cent) at the first attempt. These findings are consistent with the postulation that students from liberal arts backgrounds continue to have difficulty with software engineering and suggests that students only submitted if they anticipated passing. The validation of this theory could be reinforced through a similar study of students with a science or engineering background.

Module	Semester	Teaching group	Assessment submissions	Total passed	% submissions that passed	% of total passes
SD1	1	Software engineering	30 (73.2%)	22	73.3%	53.7%
DMIA	1	Multimedia and interactive systems design	38 (92.7%)	36	94.7%	87.8%
PMDM	1	Multimedia and interactive systems design	40 (97.6%)	37	92.5%	90.2%
ISD	1	Multimedia and interactive systems design	41 (100%)	34	82.9%	82.9%
SD2	2	Software engineering	19 (46.4%)	17	89.4%	41.5%
OOMDD	2	Software engineering	41 (100%)	31	75.6%	75.6%
GP	2	Project	40 (97.6%)	36	90.0%	87.8%
AISD	2	Multimedia and interactive systems design	37 (90.2%)	23	62.2%	56.1%

Table 1 Achievement of 41 full-time students on MSc Multimedia and Interactive Systems

The OOMDD module, with a 100 per cent submission contrasts favourably with the other software engineering modules, SD1 (73.2 per cent) and SD2 (46.4 per cent). OOMDD and SD2 are delivered concurrently thus allowing direct comparison. The performance of those students who submitted the assessments for both modules shows a positive correlation ($r = 0.5$) without any significant difference in means. Due to the lower level of submissions in SD2 only 41.5 per cent of students passed, compared to a pass rate of 75.6 per cent in OOMDD (Figure 1).

The decision whether or not to submit was taken by both weak and strong students (Figure 1), and was independent of subject matter, teaching materials and lecturers. The only identifiable difference is the approach to LTA, specifically the enhanced learning experience.
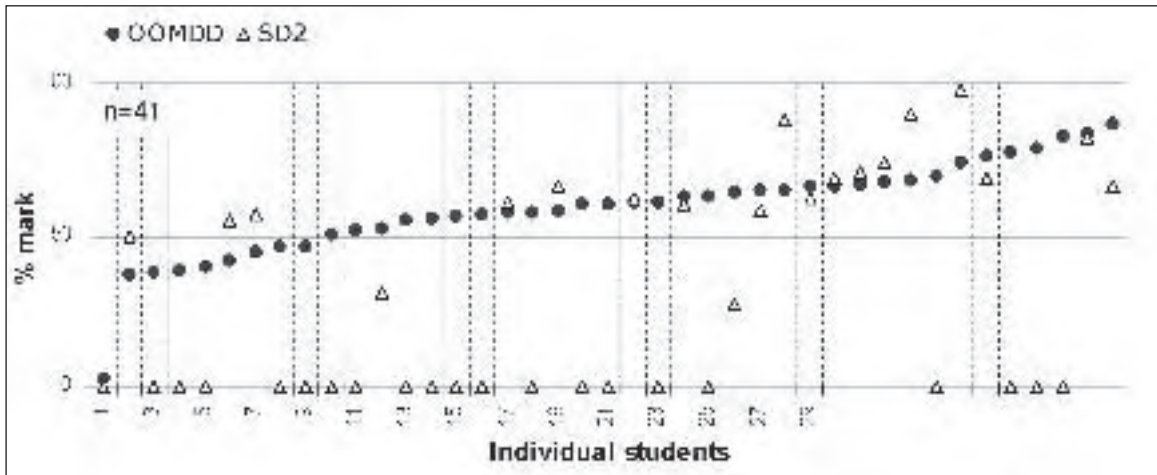
Figure 1 Comparison of individual student achievement in second semester software engineering modules (academic year 2002-03)

There is no evidence that individual student performance was improved overall. However, the cohort under study demonstrated a higher pass rate when compared to similar modules. This can be attributed to a caring and supportive environment, developed to nurture and encourage achievement.

Advanced Digital Media (ADM)

In June 2003, the LTA approach was integrated into a new level 4 module, ADM. As previously implemented with OOMDD, the module utilises a traditional lecture/ tutorial/practical format supplemented with web-based resources. In addition, students are expected to do further reading, both directed and independent.

Assessment

The assessment for the module comprises both a coursework and an examination. The coursework assessment is developed as a group project and as a consequence produces homogeneous results; therefore an examination is included to allow the student to demonstrate individual learning as individual grades from the module are used to determine final degree classification.

Coursework

The coursework is considered by both external examiner and students as challenging and is soundly underpinned by prior learning. The students are tasked to develop an interactive DVD for the movie *Living Dead in the School of Computing* and an associated website featuring the flash-based game *Zombie Resurrection: The Curse of the Computing Centre*. Each group is required to storyboard and shoot a five minute movie; determine an appropriate requirements specification for both DVD and game; produce a detailed design model from the requirements specification, using a range of analysis and design methods; develop a prototype DVD from the design model; develop a prototype game from the design model; and provide a critical evaluation of the design and development process.

Open-notes examination

The open-notes examination, comprising three compulsory questions, is solely based on six topics of directed study that are introduced through the lecture series and supported through a virtual learning environment. The move towards open-notes was taken to encourage the students to better prepare for the examination by utilising techniques such as essay plans and mind maps. Analysis of the exam scripts confirmed that the students who developed and adhered to essay plans achieved the higher grades.

Discussion

The ADM module was introduced for the first time in October 2003 to the BEng Multimedia Systems and BSc Multimedia Technology programmes, and was very highly ranked across all criteria in the student feedback survey (Table 2). It was perceived to be an enhanced learning experience and second only to the honours project for personal contribution.

Module ranking (n=45)				
	Advanced Digital Media	Media Computing	Interactive Application Development*	Computing & Society**
Content	6	21	32	31
Administration	5	33	28	13
Learning	6	43	24	25
Feedback	4	1	33	38
Student contribution	2	27	32	25

*BEng only **BSc only
Source: Student Feedback Survey, All School of Computing Modules, Semester 1 2003-04

Table 2 Ranking of modules studied by subject group in first semester

In academic year 2003-04, semester 1 with 27 students, the overall pass rate was 93 per cent; the mean was 57; the standard deviation was 10.5; the Dbar was 3.9; and seven students (26%) gained merits (=65%). Analysis indicated that individual student achievement correlated with their performance in the other modules studied during the same period.

A comparison of student performance in similar examination assessments during the first semester of ADM indicated a positive correlation ($r = 0.59$) with no significant difference in the means.

A comparison of student performance in the similar Media Computing module (examination 50 per cent, coursework 50 per cent) also indicated a positive correlation ($r=0.5$).

However, further analysis determined a difference in cohort achievement of nearly five percentage points (Table 3). Both modules had very similar results with comparable numbers achieving merit passes. The key difference was a lower number of fails in ADM (one student) in comparison to Media Computing (three students).

	ADM	Media Computing
Mean	57.3	52.6
Fails (overall<40)	1	3 (11%)
Qualified fail (failed one component)	1* (3.7%)	1
Pass (overall $\geq$ 40)	18	17
Merit (overall $\geq$ 40)	7	6

* Student had an appeal for mitigating circumstances upheld at the Board of Examiners.

Table 3 Comparison of cohort performance in similar first semester modules (academic year 2003-04)

The cohort under study comprised 27 students including significant numbers of other nationalities (26 per cent: two Chinese, five French) and special learning needs students (11 per cent: two Dyslexia, one Asperger's syndrome). All seven other nationals passed ADM but three (one Chinese, two French) failed Media Computing. The individual special learning needs had similar performances on both modules. The student who failed ADM gained a marginal pass in Media Computing.

There is insufficient data to make any conclusive judgement but futures studies should further investigate the relationship between the LTA approach and the learning experience of students with English as a foreign language.

Conclusions

This paper has demonstrated that an enhanced learning experience can encourage students to take responsibility for their own learning and develop as autonomous learners. Student engagement was encouraged through active learning and collaborative learning.

Regular feedback allows students to measure their progress, and the tutor to identify gaps in their knowledge. Student disengagement can be readily identified and prompt remedial action can be taken, ensuring that the focus of the module is relevant to the students engenders engagement. The process of collaborative learning and peer support encourages the students to assist and take care of each other.

This paper has described computing modules but the approach would be appropriate for any discipline, at both undergraduate and postgraduate level, where students are experiencing difficulties with subject matter and are disengaging from the learning process.

Future research will investigate the relationship between the LTA approach and the learning experience of students with language and cultural difficulties.

References

Anderson P, Bennedsen J, Brandorff S, Caspersen M E and Mosegaard J (2003) Teaching Programming to Liberal Arts Students - a Narrative Media Approach, *ITiCSE 2003 conference (Thessaloniki, Greece, University of Macedonia) ACM Special Interest Group in Computer Science Education* (*SIGCSE*)

Biggs J (1999) *Teaching for Quality Learning at University*, Open University Press, Buckingham

Brown S and Knight P (1998) *Assessing Learners in Higher Education,* Kogan Page, London

Freeman R and Lewis R (1988) *Planning and Implementing Assessment,* Kogan Page, London

Little D (1991) Learner Autonomy*, 1: Definitions, Issues and Problems*, Authentik, Dublin

Stevens P (2001) *Using UML: Software Engineering with Objects and Components,* Addison-Wesley

Yair G (2000) Reforming motivation: How the structure of instruction affects students' learning experiences, *British Educational Journal*, 26, 2, 191-210

Multimedia & Interactive Systems Resource Centre
www.dcs.napier.ac.uk/~mm/ooa.htm

Using assessment to motivate learning - Post-workshop report

Dr Andrew Eadie, Glasgow Caledonian University and Workshop Director

The one-day workshop 'Using assessment to motivate learning' took place at the Hilton Glasgow Grosvenor Hotel on 5 February 2004. The workshop consisted of keynote addresses from Professor Craig McInnis, University of Melbourne and Professor Phil Race, University of Leeds; case studies from Dr Jean Cook, Glasgow Caledonian University, Ian Smith, Napier University and Peter Scott, Glasgow Caledonian University; two breakout sessions; and a plenary session to give an opportunity for the delegates to participate.

The talks and case studies

Professor McInnis gave an interesting and thought provoking talk on 'Repositioning assessment to enhance learning outcomes'. The talk contained much practical advice on achieving quality in student assessment particularly in the context of large class sizes and the need to provide high quality feedback to students. He finished by providing a checklist of sixteen indicators of quality in student assessment. His talk also contained valuable pointers to online resources.

Professor Race gave a very entertaining and engaging talk entitled 'Designing assessment to enhance student learning'. He began by asking the audience to share their personal worst assessment nightmare with their neighbours and then with the rest of the audience. He then examined why we assess, what is wrong with assessment currently and finally, what should be done to correct the problems.

Peter Scott gave a humorous talk on his work teaching law to environment students entitled 'Contextualised assessment strategies'. His work showed that by contextualising the assessment, making the assessment relevant to the students and also making the assessment part of the learning experience a module on fairly complex legal issues could by successfully delivered to non law students.

Dr Jean Cook showed in her talk, Assessment driven learning, how assessment can be used to motivate students undertaking mathematics. Assessment driven learning was described as a process where lecture time is reduced, tutorial time increased and computer-based assessment with almost immediate feedback is part of the teaching process. Part of the student motivation in this process is because successful completion of the computer-based assessment results in exemption from end of module examinations. This process has result in greatly increased pass rates for the module concerned.

Ian Smith gave an enthusiastic talk on 'Motivating students through group project and open-notes examinations'. His starting point was that all aspects of learning, teaching and assessment are interrelated and should not be considered in isolation. His aim was to enthuse, engage with and support the students. His results showed that the combination of group project work and open-notes examinations was well liked by the students. However, it was too early to say if there had been a significant effect on the student's results.

Breakout groups[1]

Issues identified

(main issues highlighted in bold)

- The need for rapidly delivered feedback on formative assessments.
- Gradation of assessment methods is required when you consider the development of a degree and the students studying it ie methods/approaches should change as course progresses.
- Assessing how and why, not just what?
- Time is required by staff and students to do this properly.
- How do we get students interested in/motivated for computer assisted learning (CAL)?
- There should be further focus on 'diagnostic' assessment, particularly as a tool for feedback (at a time when time for feedback is being reduced).
- Be more analytical of existing assessment types ie determine what we have in common and develop existing practice.
- Be explicit in indicating expectations to students and using assessment to shape their learning.
- How to cover the class where there is a wide spectrum of student abilities? The assessment often does not reflect this scope (unless it is failure rate and quite significant)?
- Exploit CAL and computer assisted assessment to lighten assessment burden.

Ensuring students read and take notice of feedback

- Provide qualitative commentary on assignments without adding a grade or mark and ensure that students don't get a mark until they have read and understood the feedback.
- Incorporate a reward in the next assignment for positively addressing the issues (including a reflective analysis) raised by the feedback on a previous assignment (work that ignores this may receive a lower grade or mark for not taking this criterion into account).
- Develop progressively weighted coursework assignments (eg 5 per cent, 10 per cent, 15 per cent instead of 10 per cent, 10 per cent, 10 per cent or some other combination) so that there are higher rewards for improving based upon feedback given on earlier pieces of work.
- Involve students in giving feedback to others on their work.

Creating more opportunities for using formative assessment (not hybrid with summative) to demonstrate their learning (not for marks)

- Provide only qualitative comments even if these become linked in some way to standardised 'grades' of work. Avoid grades and marks altogether.
- Use class responses to questions during teaching to identify areas of major (or critical) misunderstanding. This may be automated using electronic 'voting

---

[1] Thanks are due to colleagues who led the discussions at the five breakout groups and to those colleagues who acted as scribes.

systems' or free space for writing qualitative answers which are then visible to the whole class for comment (by them as well as the teacher). The system adopted should be designed for student anonymity (though teacher may be aware of who is answering questions in particular ways).

- Asking a student to summarise the session/lesson with others asking questions.

Fears about using group-based assessment where peer assessment is involved

- Make all ground rules known to students about their responsibilities and how the reward system will work, including penalties for not pulling weight, contributing, absence (and how this will be dealt with) etc.

- Ensure authentic opportunities for students to record their contributions to group work, eg have scheduled meetings for review of progress where minutes, attendance etc are kept; these together with logs and diaries are all submitted with the final group assignment.

- Provide training in working in groups as well as involving students in creating and implementing the peer assessment procedure for their assignment or project.

- Specifically award grades or marks to the group process as well as to the work (or products) of the group for assignments.

- Provide a formative-only assignment involving groups and peer assessment before embarking on a summative assignment.

- Start as early as possible in first year developing group-based work and peer involvement in assessment (gently and not too high risk at first!).

- Draw attention to use of peer assessment in other real world activities, eg peer review (publications, books research grants, teaching review and quality assessment) in academia; 360 degree feedback, appraisal etc in industry and commerce as well as academia.

Reducing the impact of summative assignments on learning and attendance (timing)

- Closer coordination of module programmes within a department where possible.

- Staggered dates for assignments across clusters of modules.

- Free time at key points in the semester to permit students to take time off to study, write or perform for their assignments without jeopardising their other studies (mid-semester breaks, class tests etc?).

Providing diversity in assessment (eg avoiding traditional examinations)

- Provide options for students to choose from to undertake assignments to demonstrate their own achievement of the specified learning outcomes (avoid students repeating playing to their strengths, however).

- Provide seen exams, eg a question (on the exam paper) is provided up to two weeks in advance and students can thoroughly research it in advance, but must write their answer under exam conditions. Often such assignments are problem-based and it is important to provide appropriate attainable resources, references etc.

- Encourage all staff to 'dip' into the Generic Centre handbooks on assessment as well as the Assessment Strategies in Scottish Higher Education project for inspiration and tried and trusted methods in use elsewhere - in addition to considering the case studies presented in this series of workshops.

Assessment of 'skills' and surrogate measures

A written report (often) is used to assess laboratory skills in sciences but may overestimate the skills (reflected by a good score in the written report) or may reflect a poor estimate because the skills are highly developed but a student may lack report-writing skills.

- Develop appropriate, relevant assessment (ie high validity and authenticity) adequately supported by additional tutors if necessary, eg checklists and careful observation for laboratory skills, debating skills for lawyers, presentations skills in all subjects.

- Don't measure the same sets of skills with all assignments (inadvertently or otherwise).

- Provide training in developing particular assessment format skills - whether these are exam writing technique, report writing, laboratory skills, eg operating microscopes etc - Provide training, eg video, animations etc to assist students develop skills.

- Provide specific tutoring in academic writing skills - preferably in context, though this may be supplemented by central university generic provision in skills for academic effectiveness (study skills, personal skills etc).

- How can we diversify the means of assessment and issues of learning cultures because undergraduate students expect to be 'fed and supported' rather than engage in thinking?

- How can we reserve grades to award students who do more than we can - we do it in professional life?

- How can we change our assessment when we are trained by professional bodies?

- How can we be creative in assessment in a rapidly changing professional field like those involved in the NHS?

- How can we take advantage of what further education has to offer without making them more conservative when they join higher education (we have the same funding council!)?

- How can we encourage flexibility in assessment within structures and cultures of institutions and what structures might we find helpful? eg setting a maximum of 50 per cent by exams?

- What can we do to motivate lecturers to adopt different strategies in assessment?

- How can we encourage more feedback to students as this is what enhances learning? - It also informs the teaching.

- How do we cope with the 'conservatisation' of assessment to offset student appeals?

- How much support do academics require to engage in online assessment? And how might you make a case for applying for extra funding?

- Enormous benefits for departments of having a uniform policy on assessment so that students can have a lot of practice at different assessment types - and they need training in **how** to do this assessment.

- Summative assessment takes place in a module rather than a course, so it can be difficult to design coherent assessment strategies - this may be dealt with through the progress file approach with self-assessment so students write about the building of skills like group work slides.

- Dependence on reliability rather than validity, with summative assessment being increasingly high stakes summative assessment, including A levels.

- There needs to be a variety of assessments - in some places this means re-introducing exams.

- There are tensions between being predictable and being fixed by structures.

- Assessment to motivate learning.

- Formative approaches.

- Critique - small groups - especially in fine art/arts.

- Discipline-based - gives students an opportunity to interact with tutor.

- Fear - motivates students to 'work'.

- 'Uncertainty' group work - peer pressure to perform using group work important but assessing difficult.

- Teaching 'communication' for presentations.

- Assessing content not presentation but provide feedback on presentation.

- Presentations get better grades - is this an issue in regards quality? Or perceived less value to other options - such as exams? Is there a 'misgiving' about reliability of alternatives to traditional models of assessment?

- Do your students understand weighting?

- Anxiety a real issue.

- Peer assessment.

- Less threatening - not always contribute to grade.

- Formative peer input - grade presence devalues formative feedback.

- Grading scheme - qualitative schemes?

- Presentations - marking seen by all - transparency and how/what is expected.

- Peer assessment for grades - no one in-group has tried it yet. One person will be trying soon. Napier University 100 per cent per assessment - quality check - done by three students on a panel - **need for careful monitoring**.

- Uniformity - word counts?

  i   clearer explanations to students about how assessment contributes to learning and how you are teaching

  ii course teams should look at assessment timetables/loading/workload/volume

  iii meta level of student understanding

a   perhaps an induction into requirements

b   students? Where is/are study skills included?

c   clarify with students

d   guinea pigs being upfront about innovation and requirements

iv   move away from academic to applied/vocational ??so?? reports softening or assessment?

v   Motivation/enthusiasm

vi   authentic assessment - reflective cycles can contribute to 'how students know'. Issue associated with 'in situ' vs 'academic' context.

Motivation

- Importance of design of assessment into totality of teaching and learning.
- Peer formative assessment (free from need to give mark) can be effective way of providing formative feedback.
- Peer and self-assessment feeding into the final summative assessment made by the lecturer.
- Reflection - ubiquitous but very often misapplied - need to find 'good assessment tasks.
- The need for student involvement in self and peer-assessment from as early stage as possible so that they developing a real understanding of what independent learning outcomes mean and also that the criteria used for assessment are understood.
- The need for formative assessment to help students learn, rather than just summative assessment to make decisions.
- An observation that much of discussion about assessment in higher education is echoed in what is currently being discussed in the primary and secondary sector.
- The need to remember that some students like lots of assessment - it spreads risk and also tells them how they are doing.
- Students like feedback on traditional examinations.

Questions emerging from the breakout groups

- For a 'traditional' department with primarily exam-based assessments, what are the best first steps, which could be taken to move towards processes which promote deeper learning and give good feedback?
- Can we tailor individual assessment to individual learning styles?
- How can we get the colleagues who aren't here to engage with the issues we have discussed?
- How do we free up the structures to allow us to assess effectively, reflectively and validly within different disciplines?
- How do we get an efficient and effective assessment system that drives quality student learning?
- Are these methods of assessment that should never be used?

- Is gender an issue in assessment and if yes what can we do to minimise any adverse effects?

Many of the participants were concerned that their voices, initiatives and desires for change at every level are not recognised or supported in their institutions. The thorny issue of priorities, especially the Research Assessment Exercise and culture of research rewards, mitigated against staff committing themselves to change in teaching whether or not it was recognised as benefiting students, their learning or improving fairness, validity and reliability in assessment!

Conclusions and issues for the future

The following common themes emerged from the breakout groups as being important in motivating students.

1   The quantity and quality of feedback on assessment.

2   The time commitment required for both staff and students.

3   The use of group work.

4   The impact of large classes.

5   The increased use of information technology in assessment.

6   The increased use of formative assessment.

7   The relevance of assessment methods.

8   Coherent assessment strategies for programmes.

9   Innovation in assessment.

10 Ensuring students are aware and understand assessment criteria.

The questions arising from the breakout groups point in conjunction with the issues above point to areas for enhancement by institutions and the Scottish higher education sector.

Assessment workshop series - No 3

Constructive alignment of learning outcomes to assessment methods

Contents   **Page**

Constructive alignment of learning outcomes to assessment methods - An overview

Professor Mike Osborne, Institute of Education, University of Stirling and Workshop Director

The importance of aligning teaching methods and assessment tasks is stressed in many publications pertaining to curriculum development in higher education. In essence the notion of constructive alignment derives from **constructivist** theory, which suggests that learners actively construct their own knowledge and understanding. Within the constructivist paradigm considerable stress is given to meaning, reflection and context, and teaching is about the provision of a context that allows the facilitation of desirable learning outcomes.

For Biggs (1995, 1996), in a constructively aligned teaching environment all stages of the process of teaching are closely linked. Thereby as Dart and Boulton-Lewis (1998) suggest 'course objectives, the teaching context, and the assessment tasks should all address the same student learning-related cognitive activities'.

A summary of much of Biggs' thinking in relation to constructive alignment is readily available through online sources, for example at the Higher Education Academy's website where a paper 'Aligning the Curriculum to promote good learning' can be found (Biggs, 2002). Here he speaks of there being four major steps in curriculum design and the delivery of programmes: 'defining the intended outcomes (the objectives); choosing teaching/learning activities likely to lead to attaining the objectives; assessing students' learning outcomes to see how well they match what was intended; and arriving at a final grade'.

Within his structure great importance is given to choice of verbs that become 'markers' in the system and which guide teaching/learning activities and assessment. A hierarchy of verbs on four levels that represent increasingly deeper levels of understanding aligns teaching/learning activities and assessment to objectives (Biggs, 2003). For those familiar with Bloom's (1965) *Taxonomy of Educational Objectives*[1], Biggs SOLO taxonomy will not be a surprise. Thus at the lowest unistructural level, we find verbs such as 'identify'. We move through multi-structural (eg 'enumerate') to relational (eg 'explain') to extended abstract (eg 'theorise'), and in the aligned system, the teaching and learning activities 'help activate the required verbs'.

---

[1] Those less familiar with Bloom's taxonomy will find references to his work in a multiplicity of texts on educational assessment (see for example Curzon, 1985, Chapter 9). His work attempts to determine the content validity of tests of achievement, and relates educational goals to measurement.

Teaching/learning activities	Curriculum objectives expressed as verbs Students have to enact	Assessment tasks
Designed to generate elicit desired verbs May be: Teacher-controlled Peer-controlled Self-controlled As best suits context	**A** The very best understanding that could be reasonably expected: might contain verbs such as hypothesize, apply to 'far' domains etc. **B** Highly satisfactory understanding: might contain verbs such as explain, solve, analyse, compare etc, **C** Quite satisfactory learning, with understanding at a declarative level: verbs such as elaborate, classify, cover topics a to n. **D** Understanding at a level that would warrant a pass: low level verbs, also inadequate but salvageable higher level attempts.	Evaluate how well the target verbs are deployed in context. The highest level verb to be clearly manifested becomes the final grade (A, B, C etc)

Figure 1 Aligning teaching/learning activities and assessment tasks to the curriculum (Biggs, 2002)

Biggs here is speaking of constructive alignment in the context of the development of a particular unit of study. David Boud takes the issue forward into the period beyond university. In his presentation at the workshop he began with the question: how can we connect higher education courses with the learning in which students need to engage after graduation throughout their lives? In a summary of the domain, which he covered in his introductory notes he remarked:

> A vital role is to prepare students for a future that is unknown to us and to them. The unknown future creates great problems for learning and assessment now and will place demands on students for new knowledge and skills beyond anything they learn in their courses.

What can we do to equip students for this? The challenges are substantial. Among many things, we will need to shift our focus to consider the ways in which current assessment practices either assist or inhibit students in developing skills for lifelong learning. We need to align assessment not only with short-term learning outcomes, but also with longer-term aspirations. Unfortunately, many traditional assessment practices inadvertently de-skill students in various ways. They focus attention on the immediate task of passing examinations or completing tasks and distract students from the more vital task of learning how to assess themselves.

Lifelong learning requires students to be lifelong assessors and to develop assessment skills they can deploy in the very many situations they face after leaving university. The presentation will explore what might be involved in promoting this. It will introduce the idea of establishing sustainable assessment as a central feature of all courses. Sustainable assessment practices are those that meet immediate assessment needs without compromising the ability of students to meet future learning needs. Key characteristics of sustainable assessment practice will be examined and issues to be faced in implementation will be discussed.

The context for these issues is familiar to many within higher education, and that is the relationship between learning at university and the skills demands of employers. As Barnett (1994) has suggested: 'Employers want expertise, but the expertise they are crucially after is not the expertise of propositional knowledge. It is forms of knowing how'. Thus the traditional concept of a qualified graduate has not been accepted by those increasingly dominant players in the educational debate, the employers and that argument can be traced back to the early 1980s and before, with this quote from the Confederation of British Industry being illustrative:

> The complaint which we frequently hear from companies is that ... the people who come out of technological disciplines are all too often less lively as people than those who have done less relevant subjects like the arts.... We would like to see somewhat greater weight attached to the personal qualities, such as motivation, ability for original thought and ability to get at and solve problems. This does not necessarily come out in some applicants (Confederation of British Industry Select Committee on Education, Science and the Arts, 1980).

We have seen over two decades of education and training initiatives in schools, further education colleges and universities that have sought to establish a stronger link between the formal sector and the workplace (see Dockrell et al, 1997). The Employment Department (ED) within the UK took the lead in the early 1990s by encouraging a greater responsiveness to the needs of individuals and the labour market by funding a number of initiatives in higher education (ED, 1993). This included the Enterprise in Higher Education Initiative whereby over 80 projects in higher education institutions were supported, each aiming to provide as Hale (1994) states '"models of good practice" for the introduction of activities that enabled students to enhance their transferable skills[2] as a preparation for their life long learning in the world of work'. As higher education has expanded and diversified through the 1990s into the new century there have been continuing attempts to

---

[2] Educationalists make the distinction between specific disciplinary skills, often referred to as vocational skills, and general skills, frequently referred to as transferable skills. Traditionally, however, the emphasis has been on the acquisition of discipline knowledge; the development of the other intellectual and social abilities is seen as 'a bonus within the curriculum' (Hare and Powrie, 1992).

increase its accountability to society (eg Scott, 1995; Barnett, 1994; Kogan and Hanney, 2000; Brennan and Shah, 2000) and a greater attention has been afforded to the employability of graduates (eg HEFCE, 2001), with a more public concern with the nature of 'graduateness' (HEQC, 1995).

In tandem in the UK we have seen the emergence of 'learning outcomes'[3]. Within an influential paper by Otter (1992) the perceived deficiency of the undergraduate curriculum is captured in the statement that 'a degree is currently described in terms of the process - three years full-time study - rather than achievement, and, that as a result, notions of quality are based on the processes rather than the outcome'. In HE perhaps the most important policy development deriving from the debates about quality has been the publication of subject benchmark statements by the Quality Assurance Agency for Higher Education (QAA, 2000). Benchmark statements represent explicit formulations of the academic community's tacit knowledge about what is learned in higher education. They define learning outcomes intended to reflect the 'academic' and to guide students as well as wider stakeholders (Jary, 2002).

Thus in order to assess and to accredit learning, including that achieved in a multiple range of environments, the re-configuration of programmes to include explicit learning outcomes has become common place. As is evident in Biggs' model, such a re-configuration is a pre-requisite of his alignment model. That being said, many would suggest that the advantages of alignment have a downside. As with the related notion of 'competence', (see Ashworth and Saxton, 1990) namely 'the ability to perform activities within a given occupation or function to the standards expected in employment' and itself an outcome, it has been argued that learning outcomes ignore process and are mechanist and reductionist. Furthermore, systems in which outcomes are rigorously defined can become a straightjacket for assessors in certain circumstances, and in many situations students demonstrate unintended learning outcomes that are as legitimate as those envisaged in our planning. If teachers and learners are influenced by context in learning outcomes and processes then as Boud (1999) suggests 'context impacts on…learning outcomes…in what teachers and learners accept as legitimate goals and what outcomes are valued over others'. Any learning outcomes model might reasonably be expected to accommodate such contingencies.

Sue Drew, in the second of the presentations, focused on the particular strategies at an institutional level that Sheffield Hallam University had developed in the area of alignment over the last 10 years. In her own words she described her presentation as follows:

> These consist of a number of complementary strategies to encourage and support the alignment in courses of the aims, learning outcomes, assessment tasks and assessment criteria, and learning and teaching methods. The Academic Board has required the use of learning outcomes in course validation documents since 1994-95 and the inclusion of pass descriptors (assessment criteria) since 2001. Course planning and validation templates have been amended and refined over the years on an iterative basis, to encourage course teams to develop

---

[3] For concrete examples of Scottish curricula constructed around a learning outcomes mode, the following may be instructive: the *Professional Competence Course* of the Law Society of Scotland (www.lawscot.org.uk/pdfs/PCCguidelinesfor applicants_March2002.pdf); the Scottish Deans' Medical Curriculum Group (SDMCG) work on *Learning Outcomes for the Medical Undergraduate in Scotland: A foundation for competent and reflective practitioners* (www.scottishdoctor.org).

courses where there is alignment. Policies, frameworks and exemplars have been developed, informed by practice. There is a validation process involving peers, which reviews course planning against University requirements, provides feedback to planning teams, and encourages dialogue and debate.

The University's Learning and Teaching Institute (LTI) works with University schools and departments (particularly the Registry) in developing policy, strategy and documentation. It also supports course teams in their course planning. The work of the presenter, an LTI member, has been informed by her research into student perceptions of their learning outcomes and what helps their achievement.

Her presentation will briefly outline the findings from her research into student perceptions relating to their learning outcomes, will indicate the University's main current requirements in course planning, will describe the aspects of the University's infrastructure which support the course planning process, and will go on to explore the main issues in supporting staff in their development of 'joined up' courses.

In the final presentation, Iddo Oberski presented a case study of a particular professional training programme offered at the University of Stirling in which alignment issues take the form of the ways in which different forms and traditions of assessment practice can be accommodated in one programme. He summarised his presentation as follows:

The Teaching Qualification (Further Education) (TQFE) programme at the University of Stirling is now in its fifth year of delivery and highly rated by its students and by the FE sector in Scotland. The core programme was initially designed for part-time students employed in the sector as lecturers, but has more recently also been developed as a route into FE lecturing, similar to initial teacher education programmes for schools, with student placements in FE colleges. To accomplish the TQFE, students must successfully complete six core units, each of which addresses a key area of FE practice, such as 'The FE context', 'Curriculum Design', 'Assessment' and so on. The programme is offered at undergraduate and postgraduate levels, in full or part-time mode and in face-to-face and distant delivery mode. Once completed, students may enrol on a Research & Enquiry unit and, depending on their prior entry qualifications, this unit allows them to graduate with a degree or a postgraduate diploma with the TQFE.

The programme is modular so that each unit is assessed at the end of a semester. The assessment of each unit consists of two main parts, a university-based assessment and a college-based assessment. The latter is competence based and requires the student to meet the occupational standards published by the Scottish Executive in 1997 (currently under review). Teacher Fellows, who are themselves working in the FE sector and who are employed by the University assess the students' performance in their place of work or at their placement. The university-based assessment is not competence based. Instead it focuses on theoretical perspectives, covering the areas specified as 'underpinning knowledge' in the occupational standards as a minimum, but in more depth and breadth, through academic essay assignments. So assessment on the TQFE seems basically divided in line with the traditional theory-practice theories of professional knowledge and it could be argued that this reflects a model of professionalism aligned to the novice-expert model developed by the Dreyfus brothers (Eraut, 1994).

Formal and informal evaluations with students and staff on the programme have regularly involved discussion around this strong division between the two types of assessment. So, for example, students have expressed wonder at having to write essays, because they may not see how the skill of essay writing is relevant to their role as FE lecturer in their particular subject. Students have also expressed the view that much of the theory coming from educational research seems far removed from the realities of practice. Of course these arguments and views are not new, with similar ideas having emerged from the areas of initial teacher education and nursing education for example. Although it is quite possible to put up a strong argument in defence of using essays, staff on the programme have recently experimented with the nature of the university-based assessment in an attempt to overcome this perceived gap between theory and practice. In this short presentation I will attempt an initial comparison between student work on the university-based assessment for the unit on 'Professional Development'. I will explain how and why the assignment was changed and illustrate the old and new assignments with some quotes from student's work. I will conclude by speculating to what extent constructive alignment has been achieved.

In subsequent papers, further details are provided by each of the three speakers. There follows a summary of the main issues that arose in subsequent workshop discussions.

References

Ashworth P and Saxton J (1990) On competence, *Journal of Further and Higher Education*, 14, 2, 3-25

Barnett R (1994) *The Limits of Competence: knowledge, higher education and society*, Society for Research into Higher Education and Open University Press, Buckingham

Biggs J B (1995) Assessing for learning: Some dimensions underlying new approaches to educational assessment, *Alberta Journal of Educational Research*, 41, 1-8

Biggs J B (1996) Enhancing teaching through constructive alignment, *Higher Education*, 32, 3, 347-364

Biggs J B (2002) Aligning the curriculum to promote good learning, Constructive *Alignment in Action: An Imaginative Curriculum Symposium*, 4 November 2002, London www.heacademy.ac.uk/resources.asp?process=full_record&section=generic&id=167

Biggs J B (2003) *Teaching for quality learning at university*, Open University Press, Buckingham

Bloom B S (1965) *Taxonomy of Educational Objectives*, Longman, London

Boud D (1999) Avoiding the traps: seeking good practice in the use of self-assessment and reflection in professional courses, *Social Work Education*, 18, 2, 121-132

Brennan J L and Shah T (2000) *Managing Quality in Higher education. An international perspective on institutional assessment and change*, OECD, Society for Research into Higher Education/Open University Press, Buckingham

Confederation of British Industry (CBI) (1980) *Towards a Skills Revolution* - a youth charter, Interim report of the vocational task force, CBI, London

Curzon L B (1985) *Teaching in Further Education*, third edition, Cassell, London

Dart B and Boulton-Lewis G (1998) *Teaching and Learning in Higher Education*, ACER, Melbourne

Dockrell R, Hettich P and Osborne M (1997) Transferable Skills - some UK and US models, *Continuing Higher Education Review*, 61, 12-27

Employment Department (1993) *Prosperity through Skills: The National Development Agenda*, Employment Department, Sheffield

Eraut M (1994) *Developing Professional Knowledge and Competence*, pp 123-127, The Falmer Press, London,

Hale R (1994) Foreword to Proceedings of EHE Directors' Conference, in R Hale and N Pope (eds) *EHE - A Vision for Higher Education*, University of Edinburgh, Edinburgh

Hare C and Powrie P (1992) Developing Transferable Personal Skills in a Business Studies Option of a French Undergraduate Degree, in J A Coleman and G Parker (eds), *French and the Enterprise Path: developing transferable and professional skills*, CILT, London

HEFCE (2001) *Indicators of Employment*, Report 21, Higher Education Funding Council for England, Bristol

HEQC (1995) *What Are Graduates? Clarifying the Attributes of 'Graduateness'*, The Higher Education Quality Council: Quality Enhancement Group
www.city.londonmet.ac.uk/deliberations/graduates/starter.html

Jary D (2002) *Subject Benchmarking in the 'Audit Society' - towards an evaluation of subject benchmarking*, Working Paper for the First C-SAP Conference on Learning and Teaching, Birmingham, Jan 2002

Kogan M and Hannay S (2000) *Reforming Higher Education*, Jessica Kingsley, London

Otter S (1992) *Learning Outcomes in Higher Education: Executive summary*, UDACE, Leicester

QAA (2000) *Subject benchmark statements,* Quality Assurance Agency for Higher Education
www.qaa.ac.uk/academicinfrastructure/benchmark/

Scott P (1995) *The Meanings of Mass Higher Education*, Society for Research into Higher Education/Open University Press, Buckingham

SOEID (1997) *National Guidelines on Provision Leading to the TQ(FE) and Related Professional Development*, HMSO, Edinburgh

References specific to David Boud's work

Boud D (1990) Assessment and the promotion of academic values, *Studies in Higher Education*, 15, 1, 101-111

Boud D (1995) Assessment and learning: contradictory or complementary? in Knight P (ed) *Assessment for Learning in Higher Education*, pp35-48, Kogan Page, London

Boud D (2000) Sustainable assessment: rethinking assessment for the learning society, *Studies in Continuing Educatio*n, 22, 2, 151-167

Boud D and Falchikov N (1989) Quantitative studies of student self-assessment in higher education: a critical analysis of findings, *Higher Education*, 18, 5, 529-549

Boud D, Cohen R and Sampson J (1999) Peer learning and assessment, *Assessment and Evaluation in Higher Education*, 24, 4, 413-426

Falchikov N and Boud D (1989) Student self-assessment in higher education: a meta-analysis, *Review of Educational Research*, 59, 4, 395-430

McDonald B and Boud D (2003) The effects of self-assessment training on performance in external examinations, *Assessment in Education*, 10, 2, 210-220

References specific to Sue Drew's work

Drew S (1998) Students' perceptions of their learning outcomes, *Teaching in Higher Education*, 3, 2

Drew S (2001) Student perceptions of what helps them learn and develop in higher education, *Teaching in Higher Education, 6, 3*, 309-331

Other useful references for thinking about assessment

Biggs J (2003) *Teaching for Quality Learning at University*, second edition, Society for Research into Higher Education and Open University Press, Buckingham

Black P and Wiliam D (1998) Assessment and classroom learning, *Assessment in Education*, 5, 1, 7-74

Falchikov N (forthcoming) *Improving Assessment through Student Involvement*, RoutledgeFalmer, London

Knight P T and Yorke M (2003) *Assessment, Learning and Employability*, Society for Research into Higher Education and Open University Press, Maidenhead

Other resources on assessment and related matters

www.davidboud.com
Publications of David Boud with some available for downloading

http://ahe.cqu.edu.au/
An Australian site about student assessment in higher education

www.ied.edu.hk/loap/index.html
A Hong Kong site about student assessment in higher education

www.keele.ac.uk/depts/aa/landt/links/assessment.htm
A useful site about assessment at the University of Keele

www.shu.ac.uk/services/lti/
Learning and Teaching Institute, Sheffield Hallam University

www.heacademy.ac.uk
Higher Education Academy

Ensuring alignment in the curriculum: aligning policies, processes and practice

Sue Drew, Learning and Teaching Institute, Sheffield Hallam University

Introduction

This paper is based on current practice at Sheffield Hallam University, a post-1992 university with approximately 28,000 students. It focuses on one process that aims to encourage constructive alignment in the curriculum, that of course planning. There are other supportive processes, for example, those relating to quality review and to the ongoing iterative development of practice, but the focus here is on course planning.

What is constructive alignment and why does it matter?

According to Biggs (2003), constructivism has as its focus learners constructing their own knowledge, rather than being passive recipients of the knowledge created by others. For Biggs, '...education is about conceptual change, not just the acquisition of information' (Biggs, 2003).

Biggs considers that: 'They (different views of constructivism) have in common the idea that what the learner has to do to create knowledge is the important thing' (Biggs, 2003).

It follows, therefore, that what students are asked to **do** within the curriculum must align with what those designing the curriculum intend them to learn. This may sound obvious, but it might be seen as in opposition to more traditional approaches, where what students are **told** aligns with what they are intended to learn. It also requires some hard thinking. Biggs sees the curriculum as a system composed of elements which need to be in balance, if the intended outcomes are to be achieved.

My own research supports this view (Drew 1998, 2001). I carried out structured group sessions with students from a wide range of courses (14 course groups with a total of 263 students) and at different levels in Sheffield Hallam University. The structured group session format was originally developed for the UDACE Project on learning outcomes and was described as a 'robust tool' (Otter, 1992). The session is a very different evaluative tool from traditional questionnaires about course provision, which assume the aspects of delivery which are important to students and then ask for ratings of those aspects (eg teacher performance, resources etc). In the structured group session, students were asked to discuss in sub groups three questions.

- What are the most important learning outcomes from your course?
- What helped you in achieving them?
- What hindered you in achieving them?

The sessions generated a considerable amount of data (flip charts from sub groups, individual summaries). An analysis suggested that, a mix of factors within the students

interacted with a mix of contextual factors, and that these factors did, indeed, need to be in balance. The main factors were as follows.

Student factors	Contextual factors
Self management	Course organisation/resources/facilities
Motivational needs	Assessment
Understanding	Learning activities and teaching
Support needs	

For example, students' abilities to manage their work depended on course organisation/resources/facilities (the timetable; access to resources; clarity and stability in assessment schedules), and on assessment (the assessment load and its distribution). Whether or not students felt supported related again to organisation/resources/facilities (clear course aims; the course being well organised), assessment (useful and timely feedback on assessed work) and on learning activities and teaching (peer support being seen as vital, so that activities encouraging peer interaction were important).

How can alignment be encouraged?

University level processes

The above suggests that these are complex interactions that need careful consideration and planning. Sheffield Hallam University has, over a number of years, developed approaches and structures which, in turn, have allowed it to develop strategies to support coherent planning. The Learning and Teaching Institute (LTI) has been a key player here.

The LTI is an educational development unit that is part of the Learning Centre and incorporates educational developers, researchers, staff supporting the implementation of the University's virtual learning environment and courseware developers. The LTI works closely with Registry to ensure that course planning and validation processes comply with Quality Assurance Agency for Higher Education requirements, promote good practice in learning, teaching and assessment (LTA) and are helpful to staff. The LTI has also led on the development of the University's LTA strategy, developed in conjunction with University schools, and is responsible for supporting its implementation. The LTI has encouraged the establishment of LTA coordinators in the University schools. These coordinators lead school LTA committees, support the implementation of the LTA strategy and meet together in an LTA coordinator forum facilitated by the LTI. From September 2004, the University's 10 schools will be reformed into four faculties and the LTA structures will be more formalised, with each faculty having a head of LTA. Each faculty will also have a head of quality whose role includes responsibility for course planning and validation processes.
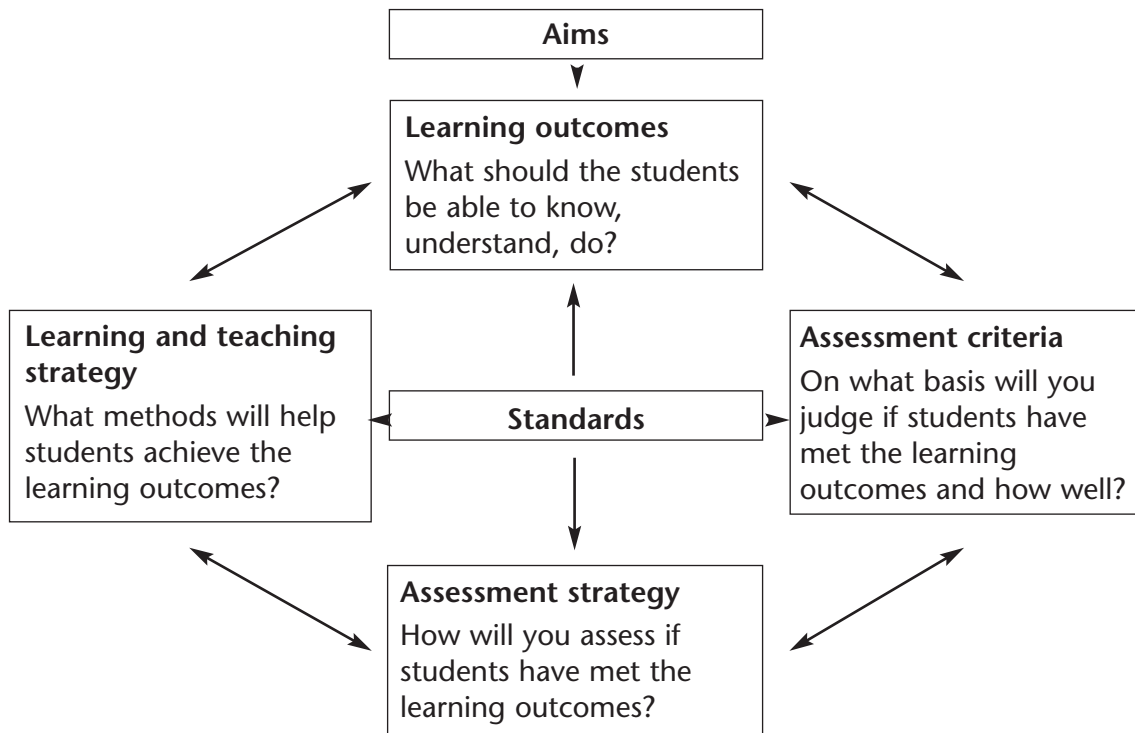
The University, therefore, has LTA working relationships and structures which allow for the development and implementation of good LTA practice.

A clear vision

For the infrastructure described above to lead to the development of coherent, aligned courses it needs to be underpinned by shared visions, aims and views of good practice.

One aspect of the shared views of good practice is the following model of course planning (Bingham, 2001), which is clearly articulated at the start of the University's module description template.

```
                        ┌─────────────────┐
                        │      Aims       │
                        └─────────────────┘
                                 │
                                 ▼
                    ┌─────────────────────────┐
                    │   Learning outcomes     │
                    │ What should the students│
                    │ be able to know,        │
                    │ understand, do?         │
                    └─────────────────────────┘
```

Learning and teaching strategy		Assessment criteria
What methods will help students achieve the learning outcomes?	Standards	On what basis will you judge if students have met the learning outcomes and how well?

```
                    ┌─────────────────────────┐
                    │   Assessment strategy   │
                    │ How will you assess if  │
                    │ students have met the   │
                    │ learning outcomes?      │
                    └─────────────────────────┘
```

The above is a basis for developing aligned courses. Learning outcomes should reflect the aims (of a course or a module), the assessment methods should assess those learning outcomes with assessment criteria linked to the outcomes, and learning and teaching methods should support the development of the learning outcomes and the achievement of the assessment tasks. It is by specifying course and module requirements in this way that standards can be articulated.

The model provides those involved in planning (the course team; the LTI; validation panels) with questions to ask or points to look out for. For example, in reviewing a module description you might ask: why are the learning outcomes all knowledge based if the aim is to develop professional skills; if there is a learning outcome about reflection, why is there no support for developing reflective abilities in the learning and teaching methods and why is it not assessed?

Enabling strategies

A number of strategies have been developed (and are still developing) to encourage use of this model.

Registry, supported by the LTI, provide templates, examples and guidance for course planning teams. These templates are designed to encourage aligned course design. They reflect the above model and draw on the aims of the LTA strategy and on University policies and frameworks, which in turn allow for external requirements, for example the *Code of practice for the assurance of academic quality and standards in higher education* published by the Quality Assurance Agency for Higher Education

(QAA, 2004). The policies include, for example, one on integrating key skills through the curriculum and policies on assessment and on learning from work that articulate good practice.

Course teams have access to LTI support for their course planning, for example, LTI staff join course planning teams or run workshops for them on specific topics (such as the development of assessment criteria) and the LTI also produces guidance booklets.

Validation panels of peers review course teams' plans, again in relation to University policies and requirements. The Registry and LTI support validation panels with the aim of encouraging consistent practice across the University. Quality processes review practice against the documentation approved at validation.

It sounds like a very joined up, cut and dried approach. The reality is a bit more 'messy'. The LTI does not reach all courses and validation panels sometimes approve documentation that does not reflect policies or fully follow through on all the elements of the model outlined above. The approach may not always lead to perfect results, but it does enable the achievement of good and often very good results.

Shifts in practice

Learning outcomes

Our initial challenge was the shift towards defining courses in terms of learning outcomes rather than content. Using learning outcomes enables course planners to be more explicit about desirable processes, skills, attitudes and values, and facilitates the development of aligned courses. If a course is described in terms of content (eg 'behaviourist theories') it is more difficult to identify how far learning, teaching and assessment methods are aligned than if it is described in terms of learning outcomes, for example, 'students will describe the principles of behaviourism'. Framing the learning desired in this way makes it more obvious if the learning and teaching and assessment methods are appropriate (will the learning and teaching methods enable students to understand the principles and will the assessment task identify how well they can describe those principles?).

At Sheffield Hallam University we have faced a number of challenges: encouraging staff to see courses in terms of learning outcomes; writing the learning outcomes in an appropriate way for the level; and writing learning outcomes which are assessable. These challenges have largely been met. The critical aspect of a learning outcomes is the verb used, since it shows what the students should do and to what level. Biggs (2003) sees verbs as critical in writing learning outcomes. Will the student describe, explain, explore, critique or synthesise? Some verbs present difficulties because they indicate neither what the student needs to do nor to what level (eg 'know', 'understand', 'be aware of').

Support for staff in writing learning outcomes has been in the form of an LTI booklet (a much appreciated aspect is a list of verbs!), workshops with course teams and feedback on drafts from the LTI, quality chairs and validation panels. The University has been requiring courses to be described in terms of learning outcomes for over 10 years.

A major supportive strategy was the development of a set of generic learning outcomes by level for the University. Staff can use these as a starting point, to help them see how they might frame their learning outcomes. Course teams adapt them for their context. They were developed in 2001 when the new academic framework for the University was approved and were based on *The framework for higher education qualifications in England, Wales and Northern Ireland* (QAA, 2004). The generic learning outcomes include knowledge and understanding, intellectual skills, subject skills, professional skills and key skills and relate to the sections in the University's programme specification template. An example of the learning outcome which refers to 'argument' is given below. The University levels are: level 3, access; level 4, undergraduate first level; level 5, undergraduate, second level; level 6, undergraduate, third level; level 7, master's; level 8, doctorate.

Level 3	Level 4	Level 5	Level 6	Level 7	Level 8
Present aspects of the subject in an order which enables understanding, using given procedures/ formats.	Sort and order aspects of the subject into a logical line of argument.	Produce a line of argument supported by relevant evidence.	Devise and sustain an argument, supported by valid/ significant evidence.	Devise and sustain an argument, supported by valid/ significant, evaluated evidence. Include some elements which are new/original/ unusual and may offer new insights or hypotheses.	Meet the standards set by peers in the discipline, in terms of devising and sustaining a new or, original argument.

(Sheffield Hallam University, 2001)

Assessment criteria

Since 2000, there has been an increasing focus in the University on assessment criteria as being crucial in clarifying standards. Module descriptions must now give pass descriptors linked to learning outcomes.

Staff find writing these descriptors difficult and time consuming, and the University is possibly at a similar stage to the one it was at a few years ago with the use of learning outcomes. Work on assessment criteria does, however, create another huge shift. The very difficulty of writing them has meant that course teams have worked together to support each other, and this leads to valuable conversations about what is really required; what students really need to do to pass; how important is x as opposed to y; and must all criteria be met and should they be weighted. It forces a rethinking of what assessment is about and what it is for. It has proved important to write learning outcomes and assessment criteria together. Often, when staff write criteria they realise that the learning outcome is imprecise or difficult to assess and must be reworded.

Support for staff in this area has included, again, an LTI booklet on writing criteria, workshops and feedback to staff from the LTI and staff involved in 'quality'. Although staff are required to give only pass descriptors on the module description template, they are encouraged to provide grids of descriptors against each outcome in grading bands. This both helps clarify to students what is needed and enables the giving of feedback. We have developed exemplars of pass descriptors against the generic learning outcomes. Again, these are intended to be a starting point for staff, to be contextualised by them. The intention is not for them to use them as they stand.

The following are the exemplars for the level 4 and 5 learning outcomes which refer to argument.

Learning outcome	Exemplar pass descriptor
Level 4 Sort and order aspects of the subject into a logical line of argument.	Elements with some commonality are grouped together and are placed in an order which makes sense for the topic and which leads to and provides some reasons for a claim.
Level 5 Produce a line of argument supported by relevant evidence.	Claims made are applicable to the topic. Main reasons and information to support the claims are given, in an order which has some logic. Some relevant, current and valid evidence is given for key reasons and claims.

(Sheffield Hallam University, 2003)

Focussing on students

A third major shift is the result of a change in focus. In 2002 the template for module descriptions was amended to be aimed at students. Wording the headings or 'lead-in' statements so that what follows is student orientated has been a simple technique which has led to a great improvement in the way the module descriptions are written. Guidelines to help in their completion are within the template and are simply deleted when the document is produced.

Previous template headings	Current template headings
Rationale Aims Anticipated learning outcomes Indicative content Teaching and learning strategy and methods Assessment strategy Assessment criteria Indicative reading	These are the aims of this module... The reason for having this module and for having it at this level or point in the course is... By the end of the module you will be able to... These are the main ways of learning and teaching which will help you to achieve the learning outcomes... This is how the learning outcomes will be assessed... This is how and at point(s) you will be given feedback on your performance... To achieve a pass... These are examples of the content of the module... These are examples of the main learning resources you will use...

(Sheffield Hallam University, 2004)

We are currently in the process of redesigning the template for the programme specification so that it will also be directed at students and so that it will encourage good practice in course design.

Changes in practice

Section 4 above indicates three major vehicles which, with support, have encouraged shifts in staff practice towards courses being more aligned. There seem to be some key, complementary elements:

- a shared understanding that the elements of courses should be aligned (ie learning outcomes related to aims, assessment criteria related to the outcomes, outcomes supported by learning and teaching methods and appropriate resources and validly assessed)

- central requirements (course planners must adhere to these if their courses are to be validated)

- ongoing liaison between Registry staff, who articulate the requirements, and the LTI, which supports good practice in LTA

- strong faculty LTA and quality structures to support and encourage staff in good practice

- central support for course planners (not only from the LTI but from the Learning Centre, Students Services and our computing infrastructure).

Have there actually been changes in practice - or do we just have a set of very well-crafted course planning and validation documents?

Where course teams engage actively with the process encouraged by the templates, it clearly leads to changes in practice. Merely being part of a discussion on writing assessment criteria shifts perceptions and challenges assumptions. Where teams regard the process as a hurdle to be overcome, a 'cutting and pasting' exercise, then impact is less. Course planning is a pressurised time and it does not always allow staff to have opportunities to really engage with changes to practice. We (the LTI) are looking at other points where we might engage with staff to encourage aligned, balanced curricula, for example quality reviews.

Balancing elements of the system which is the curriculum is complex. New perceptions, realisations, or developments (such as e-learning) shift the balance, so that there is a constant need to review practice and the most effective ways of encouraging and enabling staff to do so.

References

Biggs J (2003) *Teaching for Quality Learning at University*, second edition, Society for Research into Higher Education, London

Bingham R (2001) *Course Planning Model*, Sheffield Hallam University, Sheffield

Drew S (1998) Students' perceptions of their learning outcomes, *Teaching in Higher Education*, Vol 3, No 2

Drew S (2001) Student perceptions of what helps them learn and develop in higher education, *Teaching in Higher Education,* Vol 6, No 3, pp309-331

Otter S (1992) *Learning Outcomes in Higher Education,* UDACE, Leicester

Sheffield Hallam University (2001) *Generic Learning Outcomes*, Sheffield Hallam University, Sheffield

Sheffield Hallam University (2003) *Exemplar Assessment Criteria*, Sheffield Hallam University, Sheffield

Sheffield Hallam University (2004) *Module Description Template*, Sheffield Hallam University, Sheffield

Websites
Quality Assurance Agency for Higher Education
www.qaa.ac.uk

Constructive alignment in the Teaching Qualification (Further Education) at the University of Stirling

Dr Iddo Oberski and Dr Kathy Nicoll, Institute of Education, University of Stirling

Introduction

The Teaching Qualification (Further Education) (TQFE) programme at the University of Stirling is now in its fifth year of delivery and highly rated by its students and by the further education (FE) sector in Scotland. The core programme was initially designed for students already employed in the sector as lecturers, but has more recently also been developed as a pre-service route into lecturing (ie an initial teacher education programme) with student practice placement in FE colleges. To accomplish the TQFE, students must successfully complete six core units, each of which addresses a key area of practice, such as 'The FE context', 'Curriculum Design', 'Assessment', and so on. The programme is offered at undergraduate and postgraduate levels, in full or part-time mode and in face-to-face and distant delivery mode. Once completed, students may enrol on a Research & Enquiry unit, and depending on their prior entry qualifications, this unit allows them to graduate with a degree or a postgraduate diploma with the TQFE.

The programme is modular, and learning for each unit is assessed summatively, at the end of a semester. The assessment consists of two parts: a college and a university-based assessment. The former requires the student to meet the occupational standards, published by the Scottish Executive (SOEID, 1997) (currently under review), for the unit. Teacher Fellows working in the sector are employed by the University to assess the students' performance in their place of work, or at their placement. The latter, the university-based assessment, focuses on theoretical perspectives and contemporary debate, covering the areas specified as 'underpinning knowledge' in the occupational standards, but in depth and breadth, and requiring critical engagement over the various perspectives and rationales that are available. This component is assessed through the writing of an academic essay. So assessment on the TQFE could be argued to reflect a theory-practice binary between professional knowledge and practice, and a model of professionalism aligned to the novice-expert model developed by the Dreyfus brothers (in Eraut, 1994).

Formal and informal evaluations with students and staff on the programme have regularly involved discussion around this strong division between the two types of assessment. For example, students have expressed wonder at having to write essays, because the skills of essay writing do not appear immediately relevant to their role as FE lecturer. Students have also expressed the view that much of the theory coming from educational research seems far removed from the realities of practice. Of course, these arguments and views are not new, with similar ideas having emerged, for example, from the areas of initial teacher education and nursing education. While we concur that a theory-practice perspective is possible, and it might appear that alignment of the essay component with practice is not strong, we would like to explore this view critically by looking at the work that is done through it. Staff on the programme have recently experimented with the nature of the university-based assessment by replacing some essays with variant forms, and it is therefore important to look in some detail at what these alternatives make possible. In this short paper we explore examples of student work on the university-based essay assessment for the

unit 'Professional Development'. We explain how and why the assignment was changed, and illustrate the old and new assignments with some extracts from two students' work. We will present an initial rhetorical analysis of these pieces of work, and conclude by speculating to what extent constructive alignment may, or may not, be indicated by them. Of course, a detailed and systematic piece of research would be necessary to come to any conclusions. What is presented here is preparatory to a more formal 'case study' exploration, it can be indicative only of practices promoted through particular assessment formulations and questions of alignment that might arise from them.

Constructive alignment: rationale for changing the assessment description

Constructive alignment is about more than validity, it is about ensuring that provisions engage practitioners, or practitioners to be, in the activities that they require for professional practice. It is about supporting practitioner/students in active engagement (Biggs, 2002). 'Constructive', then, does not mean that the alignment is simply beneficial. Rather, that learners construct 'functioning knowledge' (Biggs, 2002) - in this context meanings of practice - through their assignment activity, instead of allowing the potential that they might learn knowledge which remains unusable. 'Alignment' ensures that programme outcomes are clearly specified, curriculum content is clearly matched with intended outcomes, learning is actively oriented, and assessments are valid in the sense that they assess specifically whether or not the student has achieved the outcomes.

The 'Professional Development' early version of the unit aimed to enable students to meet the mandatory Units 5.1, 6.1, and 7.1 & 7.2 in the Occupational Standards (SOEID, 1997). These prescribe that the lecturer must be able to 'Evaluate the Teaching and Learning Experience', 'Support Continuous Quality', 'Develop Effective Working Relationships' and 'Manage Self' respectively, with the latter two units under the heading of 'Continue to Develop Professionally'. In further development of the unit we decided to align the content and assessment directly with a reflective practice model of professionalism and with the specific practices of reflection denoted by this model. Practices of reflection are synonymous with those of construction within constructive alignment, in that they require a process of the construction of meaning within practice. Thus, the college-based assessment, requiring a reflective account (written and oral) of work towards these outcomes, remained essentially unchanged. However, the university-based assessment was rewritten to require students to do reflective practice, and demonstrate this through their written work. While the Occupational Standards under the heading of 'Teaching and learning approaches' state that the TQFE programmes should promote 'the concept of the reflective practitioner' (SOEID, 1997), none of the actual performance criteria listed in Annex B explicitly require reflection as practice. Only section 7 on Professional Development states that 'The units in this section constitute the first steps towards addressing the areas of personal and career-long development which the lecturer requires as a reflective professional' (ibid, p63).

There are two alignment issues here. First, reflection can be viewed as a process of thought rather than written word. To require a student to demonstrate success in reflection through a written text may, through one interpretation, be a barrier to those who do not write as well as they reflect. Thus we introduced the option of an oral account supporting a written one within our college-based assessments. However, if the intention is that the competent practitioner must also be literate in reflective writing as well as in reflective practice, then to require a written assignment is still an alignment.

In the early version of the assessment for the unit, there was a range of essay topics available to the student. Each of these required critical engagement with issues of practice. The pertinent aspects of the 'old' and 'new' assignment description are given in Figure 1.

OLD

Essay topic

a   Critically discuss the models of professionalism and professional development that dominate within the FE context with reference to alternatives.

b   Critically evaluate a 'case study' course or programme evaluation.

c   Identify and critically discuss contributions that a lecturer can make to support continuous quality improvement within the FE context.

d   Negotiate a title with your tutor.

(Student Handbook, Unit: Professional Development and Evaluation, Spring 2003:4)

NEW

Assignment - critical incident analysis

If you have not yet had experience working in a lecturing role in an FE college, you are asked to:

1   interview an FE lecturer to identify a critical incident that you can explore and gain their analysis of the incident

2   analyse the incident on your own using the guidelines offered by Tripp (1993) to help you

3   inquire into the incident by finding literature that will help inform your analysis

4   work with University based colleagues to analyse the incident further and with reference to the literature

5   develop practical knowledge regarding the critical incident

6   write a description of the incident using the above headings as a guide.

For this assignment, you will need to explore what others might have already written about the type of critical incident identified.

(Student Handbook, Unit: Professional Development, Autumn 2003:4)

Figure 1 Old and new assignment descriptions on the unit Professional Development

The essay topic 'c' within the old description relates to a critical engagement with the 'additional knowledge' component, as indicated in Figure 2. The performance aspects of it are, as previously outlined, assessed through the demonstration of competent performance within the college environment.

Outcome 6.1.4 Contribute to effective quality assurance and quality improvement

Performance criteria

a   The outline of quality assurance policies and initiatives is accurate.

b   The explanation of the role of the lecturer in quality assurance is correct.

c   The action taken to implement quality assurance procedures in own work area is in accordance with college procedures and practices.

Range statement critical terms

QA Initiatives: SQMS; IIP; ISO90000; Chartermark; SCOTVEC quality criteria; HMI quality elements

Evidence requirements

Practical evidence that the candidate has implemented quality assurance procedures and has effected improvement in his or her area of work as specified in performance criteria c.

Written or oral evidence of the identification of quality assurance policies and the explanation of the role of the lecturer within that as specified in performance criteria a and b.

Additional knowledge evidence of: Concept of continuous quality improvement; quality assurance and initiatives; College quality procedures.

Figure 2 Occupational standard from the National Guidelines (SOEID 1997: 57)

Our old assessment (Figure 1, Essay topic a) required a critical analysis of the concept of the reflective practitioner. It is for closer alignment with reflective practice itself, rather than with the concept of it, that we devised our new assignment.

Evidence of alignment? Initial rhetorical analysis of student work

For the purpose of this paper, we will present an initial rhetorical analysis (Nicoll and Harrison, 2003) of the work that the students produced in response to these very different assignment descriptions. A more detailed analysis will be published elsewhere. We extracted one piece of student work from the archives for each of the old and new assignment. (The piece of work submitted for the new assignment happened to be by a pre-service student and thus the pre-service assignment description is given in the Figure 1 above. The in-service assignment was similar, except that students were asked to draw more on their own experience of teaching in FE). We then analysed these pieces of writing in relation to the work required by the assignment topics. Figure 3 illustrates two extracts from a student's work in response to the old assignment (Essay topic, c). Figure 4 illustrates two extracts from a student's work in response to the new assignment.

'I agree with Salis (1996) 'argues (sic) that quality is a dynamic idea and that too much definitions may kill it' I also agree with the multi-layered approach suggested and this is where we as a college fall down. The simplest tasks are not achieved and frustration sets in, for example we ask for something to be typed and it comes back with spelling and grammar mistakes, this shows that it has not been spell checked or proof read. The lecturer then adopts the attitude that we would be quicker to do this task ourselves which leads to an increased work load which means less time for other functions which leads on to stress and eventually time is take of (sic) work. In reality we should sent the document back with feedback about the document, however, it is college policy that the word processor type what they are given and they do not need to proof read or spell check the document. This would suggest that policy and procedures have to be adapted which allows personnel to reduce the amount of mistakes that are being made.'

'As a lecturer I look at quality after the event, this is to (sic) late, and West-Burnham (1994, P167) argues that "Evaluation is a classic exemplification of a reactive culture at the micro level however significant it may be at the macro level" I have to agree with this statement. As a lecturer I have to become proactive and I believe TQM would allow me to do this. I could do this if I was allowed to carry out management by process. It is achieved by realising that results come from process. In other words rather than focussing on the Performance Indicator Data we focus on the process, which has caused the result…Using the model of TQM and by managing by process as a lecturer I would still be working towards quality improvement through self-evaluation but I would be being proactive rather than reactive.'

Figure 3 Two extracts from one student's work on the old assignment

'Interpretation of the incident within the immediate context…

In examining the incident in the immediate context the student teacher felt that the incident had indeed been a typical event in that she was not overly surprised that the incident had occurred. In many ways she expected some students to be disruptive in class and did not feel that this was particularly unusual behaviour. By examining the incident in the immediate context two main initial interpretations were drawn:

1   The lesson was boring and did not meet the learning needs of the group
2   The two students were trouble makers, de-motivated and did not want to learn.'

Figure 4 Extract from one student's work on the new assignment

Rhetorical analysis can reveal detail in the way in which particular kinds of descriptions of the world are required and built up within and through assessments. Assignment descriptions and their associated criteria for assessment require the learner to construct knowledge in ways that are persuasive in terms of these, and analysis may show the means whereby this is done. Assessments are thus mechanisms whereby power is exercised in the production of knowledge - of the objects to be known, the subject who knows, and specific modalities of knowledge (Foucault, 1972): 'The learner is "trapped", and cannot escape without learning what is intended' (Biggs, 2002), and this is both productive and constraining.

We can thus draw on some resources from the work of Michel Foucault (1972) and rhetorical analysis (Solomon et al, 2004) for an analysis, to examine in some small detail the work of the two assessments; to see what can be said of their rhetorical accomplishment. As descriptions, they stand in differing relationships with the government frameworks of occupational standards, and academic knowledge, as already indicated. Through both assessments, specific objects of knowledge, forms of knowledge, and a knowing subject, are constituted and reinforced.

In the old assessment description (Figures 1 and 3), knowledge is framed as knowledge of 'the contributions…to support continuous quality improvement'. This is positioned through our assessment criteria, as that which can be identified and critiqued 'with reference to appropriate literature'. The objects constituted, then, are those of 'contributions', 'continuous quality improvement' 'the lecturer's role in support' and 'appropriate literature'. The subject is thus one who knows the contributions to support continuous quality improvement, and can refer to the appropriate literature to identify and critically discuss this knowledge.

It can be seen that this essay topic constructs demonstration of the achievement in terms of the occupational standard and in a specific way. 'Constructive alignment' here takes on a new meaning. Meaning cannot be constructed freely by learner activity, but is constrained by the assignment description. By requiring **identification** of knowledge of individual contribution to continuous quality improvement within the essay, the object of knowledge that is constituted within the occupational standard is reinforced. Within the range statement, quality assurance and improvement is that which is already determined through 'college procedures and practices', and wider policies and initiatives, as identified within the 'Range statement critical terms'. Demonstration of achievement of the standard is to be made by effecting improvement to practices, and by providing written or oral evidence of knowledge of the role of the lecturer in these terms. The subject is one who knows their predetermined role and activities within a whole network of such roles and activities. This position and knowledge is made possible by the knowledge objects 'continuous quality improvement', 'quality assurance and initiatives', and 'College quality procedures' which are made secure through the range statement critical terms.

It is by requiring **critical discussion** of the object of the occupational standard that the essay topic does quite different rhetorical work in the constitution of knowledge and subject positioning. This is made possible by through a further object, in what might be identified as a 'tactic' or 'manoeuvre' (Foucault 1979) in the exercise of power, which is the 'appropriate literature'. This intersection and relationship, both allows the constitution of a **critique** of the predetermined roles and activities of the lecturer, and binds him or her in a focus on the further education context.

The student (Figure 3) demonstrated that they have interrogated knowledge derived from the literature, and has drawn on it to build up a description of potential institutional and personal development. This is a specific rhetorical style and process of dynamic interrogation. The student positions themself in agreement with object representations and knowledge put forward by the authors Salis and West-Burnham. These are represented within the student text, as modified and alternative to those implicit within their own context. To warrant the narrative and acceptance of these

representations and objects, the student presents an argument of their capacity to offer solutions to problems of institutional practice. The notions of 'quality is a dynamic idea', 'the multi-layered approach' and of evaluation as 'a reactive approach at the micro-level', are these solutions. They emerge through a comparison of the objects of quality that are implicit with a specific institutional context with those described within 'appropriate literature'. Coherence of argument, and a problem-solving future-oriented correspondence of knowledge with reality, is the rhetorical devices drawn upon here.

Explicit within this description is also a logos of agreement/disagreement - 'I agree with Salis…I also agree with…I have to agree with this statement'. The authors of the literature are positioned as those producing the knowledge. But at the same time the knowledge is represented as that which is to be accepted or rejected, agreed or disagreed with, in terms of its capacity to offer solutions to practical problems. The self-identity constructed through this description is one valorising a critical cognitive process whereby knowledge produced elsewhere is accepted where it affords such potential solutions.

In the new description (Figure 1), students were assessed through an assignment that attempted to reverse an emphasis on theory within a theory/practice binary, through a focus on practice. Here, pedagogical discourses of reflective practice point to the appropriateness of cognitive processes which identify practical knowledge, and work to make this more explicit and to develop it critically by drawing on the knowledge of others, including theoretical knowledge. Here the knowledge produced is emphasised as practical.

Implicit within this assessment description are the same two forms of knowledge that were found within the essay topic - practical and theoretical. Although theory is implicit, in that 'the literature' is to be drawn upon to help critically analyse the incident, this latter is oriented, through Tripp's (1993) process, towards the 'root cause' of the incident. The student is required in this to identify the literature that is appropriate to the incident, rather than it being identified as literature related to the unit material. Pedagogy is thus not positioned as the transmission of knowledge, at least directly through its selection by the academic. Rather, the student must actively search out knowledge, and construct their own meaning with reference to this and their own prior assumptions. The cognitive capacity is the ability to 'develop practical knowledge' through a process that requires that the student come to their own decisions about the persuasiveness of knowledge from differing sources, albeit through a predetermined process of inquiry and analysis that reifies certain sources and forms of knowledge.

The assignment was not aligned with any description of occupational standard in this case. The incidents that emerge to be analysed do not therefore arise as 'objects', from those already determined through the assignment or occupational description. Rather, they emerge from the descriptions that the individual can 'imagine' through those discourses that they are already subject to, within the further education or wider context.

This student (Figure 4) drew on literature that explicated the processes of incident analysis to inform her assignment work. She wrote about this: '… through the analysis process, the interpretations of the incident began to change as the analysis progressed through the various stages.' By recording a self-evaluation of her lesson and examining this document, she came to the conclusion that '… she had incorporated a variety of learning and teaching strategies to make the lesson as interesting as possible to meet a diverse range of learning needs'. 'Vicarious analysis', based on the experience of colleagues, led the student to consider various possibilities of seating arrangement and group work as a way of countering what colleagues considered as the bad behaviour of known 'troublemakers'. The student then explored alternative interpretations '… based on a more general meaning…This usually involves adopting a wider perspective, perhaps social view of the situation (Tripp, 1993)'. Various interpretations arose through this process, leading the student to consider that she may have '… labelled and stereotyped the two students as trouble makers due to her own personal values and previous experiences as a teacher/learner' or that '… the students had no autonomy over their learning and were therefore de-motivated. This was contextualised by the classroom institution and in turn contextualised by the college curriculum which was contextualised by the Scottish Qualifications Authority unit descriptor'. The student went on to consider each of three interpretations in detail by exploring literature that she took as relevant. She concluded by identifying 'The practical knowledge that has emerged from the analysis process and its implications on future practice'. This knowledge was informed by the literature upon which she had drawn. A further extract from this student's work is offered in Figure 5.

'Despite a diversity of teaching approaches the student teacher has come to realise that it is not necessary a failing of her own teaching when all of the learner needs are not met all of the time. In this respect the practical knowledge that has emerged is that she has learned not to be too overly critical of her own abilities, realise her own limitations and accept that she cannot take on the responsibility for every learner's motivational levels…This was felt to be a very important implication for future practice because as a pre service student, it is often easier to apportion blame on to one's inadequacies and inexperience as a teacher as opposed to other complex motivational factors.

From reviewing current literature and obtaining the vicarious experiences and practical solutions from fellow colleagues, it was felt that the seating arrangements within the classroom are extremely important in the success/failure of lessons. Another important theory to practice link which has emerges is that whilst it is commendable and learner centred to encourage students to be independent in their choice of seating arrangements and groups, it is nevertheless sometimes necessary as a teacher to control or direct these arrangements to accommodate the needs of the lesson and the learners. In this respect, the student teacher is hoping to take more control of the physical environment in which she teachers by arranging the classroom to encourage group work and participation among (sic) all members of the group…

Perhaps the most important aspect of practical knowledge which has emerged in respect of the student teacher's personal experience is that in her future practice she will try to engage in critical reflection with other peers, who are perhaps not of a traditional FE background. It was felt that this was an important aspect of future

> practice as without wishing to appear overly critical of colleagues, sometimes their own stereotyping is often exacerbated by cynical views and de-motivated attitudes towards their institution/profession… [this]… can and does significantly effect (sic) how we label our students and for this reason she would like to involve other members such as fellow pre-service student (sic) within her future evaluations. It is hoped that this will help to provide a wider perspective on future issues of practice.'

Figure 5 Further extract of student work in response to the new assessment description

Figure 5 suggests that a quite different form of knowledge, identity and performance is made possible through an assignment description requiring reflective practice. The student identifies her practical knowledge, and embarks on a process for the critical scrutiny and development of this. The warranting devices drawn upon are those of a situated consensus and corroboration, first in terms of the descriptive interpretations of experienced colleagues. This constructs the potential for a community of practitioners with shared norms, values, and knowledge. Second, however, by testing knowledge further, and this time in relation to codified knowledge, the knowledge of this community and prior personal knowledge is found wanting, the student clearly begins to define professional values and norms around specific **processes** of critical interrogation.

The authors of the literature and, potentially, experienced professionals are positioned as experts within this description. The participant is positioned as a novice who will achieve expertise through a continuation of such processes of reflective practice as a defining engagement within practice. The binary between theory and practice is upheld, with the differing roles that are allocated through this. Theoretical knowledge is again valorised. However, warranting of the descriptions of practical knowledge is achieved through its concurrence with relevant theoretical knowledge, and in the testing of peer knowledge against individual practical knowledge and that found within the literature. This is a quite distinctly different style and process of dynamic interrogation from the previous one. The object domain, and narrative of the description is not confined through a relationship with a competence statement. What is to count as an 'incident' was not prescribed in the same way as in the previous assignment, which had required the narrative to address 'continuous quality improvement', and where the alignment of this object was pre-determined.

Conclusion: constructive alignment?

From the previous analysis it might be thought that alignment between the assessment and the Occupational Standards has improved (has been constructive) with the redesign of the assessment description in the Professional Development unit on our TQFE programme. However, it must be remembered that our analysis draws on the work of only one student in each case and it is likely that a greater range of rhetoric strategies would be evident in if we had been able to examine a larger sample of students' work on the different assignments. Furthermore, seeing an improvement in alignment assumes that the kind of reflective practice in the second piece of work accords to a higher degree than the first with the requirements for reflective practice as specified in the Occupational Standards. However, it should be

clear from our analysis that the very notion of reflective practice itself, as actually exhibited through student work, requires much deliberation. In making the shift from notions of the competent and technical expert to the reflective practitioner as a framing for assessment descriptions, certain potentials of professional identity and comportment are reinforced and others elided. In this, a realist and very specific view of the world and practice is reinforced. In other words, notions of competence, expertise and reflection are rhetorical achievements, which have significant rhetorical effects within and through pedagogy, and elsewhere, which are important to explore. Pedagogical descriptions can thus be explored for their work in rhetorically building up the most 'appropriate' or 'effective' work and comportment of the teacher, and in undermining alternative discourses. It implies - as does expertise and competence - an empiricist and realist view of the world. While persuasive and in tune with many common sense views of the world, this is a rhetorical achievement that fails to reflexively articulate the rhetoric of its own invention.

Finally, a similar conclusion can be reached about the notion of constructive alignment, which is itself a rhetorical device belying an empiricist and technical-rational perspective on pedagogical practices. In conjunction with learning outcomes, performance criteria and range statements, it oversimplifies the complexity of learning situations by offering cause and effect solutions that do no longer require professional autonomy. Perhaps what is needed is a more detailed breakdown of the process of reflective practice into performance criteria, in order to further improve constructive alignment!

References

Biggs J (2002) Aligning the curriculum to promote good learning, *Constructive Alignment in Action: An Imaginative Curriculum Symposium*, 4 November 2002, London www.heacademy.ac.uk/resrouces.asp?process=full_record&section=generic&id=167 [Accessed June 2004]

Eraut M (1994) *Developing Professional Knowledge and Competence*, pp123-127, The Falmer Press, London

Foucault M (1979) *Discipline and Punish: the Birth of the Prison*, Peregrine, London

Nicoll K and Harrison R (2003) Constructing the good teacher in higher education: the discursive work of standards, *Studies in Continuing Education*, 25, 1, 24-35

Scottish Office Education and Industry Department (SOEID) (1997) *National Guidelines on Provision Learning to the Teaching Qualification (Further Education) and Related Professional Development, Annex A&B*, SOEID, Edinburgh

Solomon N, Nicoll K and Edwards R (2004) *Rhetoric and Educational Discourse: Persuasive Texts*, Routledge, London

Tripp D (1993) Interpretation: Creating critical incidents, in *Critical Incidents in Teaching: Developing Professional Judgement*, Routledge, London

Constructive alignment of learning outcomes to assessment methods - Post-workshop report

Professor Mike Osborne, Institute of Education, University of Stirling and Workshop Director

In the first breakout session, four parallel groups were asked to explore issues pertinent to constructive alignment of learning outcomes to assessment methods in general terms, based upon their individual and collective expertise and experience, using questions 1 and 2 below as a starting point.

1   What are learning outcomes?

2   What currently happens in your institution to support and encourage the alignment in courses of the aims, learning outcomes, assessment tasks and assessment criteria, and learning and teaching methods?

In the second session, groups were asked to identify one major problem/issue pertinent to the workshop theme then go on to develop ways in which this may be addressed and overcome by developing new ideas and learning practices that can be recommended for adoption by institutions. Questions 3 and 4 below were the stimuli for these discussions.

3   What else/more could happen in your institution to modify our current assessment practices to equip students to be more effective learners?

4   What institutional barriers need to be removed to ensure that assessment contributes actively to learning rather than merely certifying performance?

Each group was asked to write, on a flipchart (provided), the problem(s) they have identified and a summary of ways in which this may be resolved with a focus on enhancing future learning and teaching standards and the student learning experience.

The flipcharts will all be posted up in the final plenary session where a rapporteur (also the note-taker in the session) from each group summarised the group's findings very briefly. There followed general discussion from the floor.

The three speakers acted as 'floating ears' throughout both workshop sessions and moved between the breakout groups listening to the discussions and providing some expert input, as required.

Based upon the group's discussions the facilitator and note-taker produced a series of bullet points that set out their identified problem(s) and the proposed model(s) of tackling these. These notes are incorporated into the account below.

What are learning outcomes?

The summaries of discussions produced by the workshop rapporteurs raised a number of issues pertaining to the applicability and suitability of learning outcomes (LOs).

At a fundamental level questions were raised as to whether LOs were a new concept and whether they were they different from learning objectives. This, in turn, in one group, led to some discussion about differences between aims and objectives. In the same group, the question of whether LOs for a programme and course were different was also discussed.

The following were common issues raised by more than one group.

Learning outcomes tend to be prescriptive and focus on the particular. How could practitioners value and measure unintended LOs, including outcomes that students had identified, but which were not recognised and not assessed? Similarly how can that learning that is unassessed (eg tutorials) be linked to LOs?

Are LOs helpful for practitioners? There was some agreement that at a micro-level LOs could be helpful. For example, one participant commented that for laboratory work such micro-level LOs had focused students in each practical. However, others comment that a set of LOs don't necessarily aggregate into the graduate 'product' at end of a programme of study. Related matters included the implication that LOs could be used to assess low-level outcomes and routine factual information, but could not capture the more sophisticated learning processes and outputs. Here differences in disciplines become important. For example, another participant commented that art students are often encouraged to surprise their instructors and explore new areas; therefore specific LOs stated at the beginning of the course are sometimes difficult to formulate. Furthermore, there was some feeling that LOs stressed a surface-level of learning and that depth of knowledge would be compromised. By contrast LOs' ability to capture generic outcomes of higher education accorded for some with the modern mission of a mass system.

Are outcomes useful to students? The disaggregation of learning into multiple LOs, and the implication of certain models that all LOs should be demonstrated, was felt to be unhelpful and problematic for students. However, the explicitness of LOs did have possible advantages. For example, it was suggested that it be beneficial to map the whole curriculum in LO terms to show explicit relationships between sub-components and how these develop at a programme level. The idea of the list of graduate skills which was mentioned by Boud in the first presentation found support from some students and it was commented these were outcomes that could certainly be useful for fourth year students thinking of applying for jobs and even first year students in terms of what to expect and look out for throughout their time at university. The sheer explicitness of LOs was also perceived by some as a means of democratising higher education. They allowed the production of a 'shared space' where students understand what and how they are to be taught and see an alignment between that and the learning that they are expected to achieve and what they will be assessed upon.

Are LOs useful for employers? In contrast to the aforementioned comments the utility of LOs to employers was questioned. For example, an individual from an engineering background pointed out that in his field outcomes were generic across the range of engineering subjects that meant that they had become almost meaningless. As a result difficulties were experienced not only by academics attempting to construct programmes based on guidelines informed by LOs, but also by employers in their interpretation of the value of qualifications derived from such generic outcomes.

Can the attainment of all LOs be meaningfully graded on a scale? There was discussion about the implications of LOs in relation to grading students since there is potential a mismatch between the principles underpinning achievement of LOs and

the level of that achievement. For instance, as one group reported what would a mark of 40 per cent signify in terms of LOs achieved? Furthermore, how might we distinguish between **levels** of LOs?

What currently happens in your institution to support and encouragement the alignment in courses of the aims, learning outcomes, assessment tasks and assessment criteria and learning and teaching methods?

Questions raised in a number of workshops related to the issue of **modularisation**.

Key questions posed included the following.

- Is the rigidity of a modular structure a hindrance? It may be too big an administrative hurdle to change assessment part way through if things are not going as planned.

- Within a modular structure, there may be different rules of engagement between modules and coordinators. How do we ensure consistency and alignment in the greater sense?

- How can progression be measured in a modular system and are we losing a sense of progression by atomising learning?

This later question also related to the difficulties of incorporating **formative assessment** within a modular structure. Possible ways suggested of approaching such difficulties included introducing mid-way assessments and breaking down assessments into specific tasks that can be built on. One group discussed the need for students to take responsibility for their own learning and the benefits of peer assessment. While the difficulties of peer assessment were also acknowledged, there was general agreement that it was a useful formative tool. Another possibility discussed was the delivery of assessments during contact time, ie presentations that had the same criteria as assignments. Furthermore, interrogating students about criteria and performance was seen as a way to enable students and staff to have confidence in the students' judgement.

Some workshop participants suggested that the **structure of modules** themselves was an issue. For example, long thin modules over the whole year rather than short fat modules over one semester could aid alignment between teaching, learning and assessment.

Other comments related to the **nature of assessment.** For example, while in certain course outcomes are written in terms of the demonstration of specific skills, the necessary adaptation of assessment, namely the use of process-based criteria, rarely occurs.

On a positive note examples were cited of **problem-based learning** within medicine scheme in which students were directed towards metacognitive approaches (learning how to learning) in tandem with developing an appropriate knowledge base. The success of the scheme however depended upon introducing the approach early enough in the degree course in order for the students to find their feet and get the right approach.

Alignment seemed quite clearly depends on the type of programme. In one group it was suggested that alignment is best achieved in those courses if driven by a

professional body/vocational courses, especially where courses are more prescriptive, the same staff teach across the course and where staff are tuned in to accreditation needs.

A number of groups made remarks about **student attitudes and responsibilities** and issues pertaining to the **increasing diversity** of the student cohort. Some comments expressed the view that certain students are more conservative than staff, and took a mechanist and strategic view of study. They want to be told what to do and what they need to know. Others expressed these concerns in questions such as:

- how do you counteract students' apathy and resistance to changes in assessment approaches, which make new and unfamiliar demands on them?
- how do you use constructive alignment to promote deep learning?

It was suggested that in those courses where reflective activities of students mapped against content are seen as key (eg physiotherapy) students are able to actively critique assessment criteria, although they may need training. Personal development plans could also be used to encourage students to question assessment rationale and alignment.

Others commented that it was difficult to engage students in the learning process given the diverse needs and experience of students and constraints on resources. Nonetheless, there was a common feeling that although students find it difficult to pinpoint their achievements, strategies that overtly aim to align learning outcomes and assessment might help.

A number of suggestions were made in relation to institutional change. Overall the sheer choice of possible routes, namely the flexibility of many higher education institutions, sat uncomfortable with ensuring consisting across institutions in relation to constructive alignment. There was a general call for explicitness in processes to demonstrate to students, what, how and when something happens. In this context it was pointed out in one group that it wasn't simply a question as to whether students know, but also as to whether staff knew as well.

What else/more could happen in your institution to modify our current assessment practices to equip students to be more effective learners?

In relation to this question, a number of issues arose and a brief summary of points follows.

- Action at an institutional level needs to address conflicting value systems/agendas.
- Learning and teaching should be a priority, but this requires time, resources and support services eg professional help in educational development.
- The language of assessment needs to change, for example, by making expectations transparent.

- Sharing ideas of what is expected at each level is a necessity. In some institutions this is clearly delineated eg by matching the grade achievable in a particular unit and level against criteria. There are advantages (eg transparency) but also disadvantages (limits to what can be assessed).

- There are of course differences in learning, teaching and assessment across disciplines, but we can still have good institutional policy which is then reflected in each subject area.

- There might be a quality enhancement day at department, schools or faculty level each year so that quality is not just seen as a paper trail: communication within the institution is crucial.

- Ways should be created so that educational development units can support busy academics who do not think they need/want support.

- Rewarding academics for excellence in learning and teaching would help. Most get promoted for research, not teaching. Could we encourage staff to research in the pedagogy of their subject? View scholarship as not just subject scholarship.

What institutional barriers need to be removed to ensure that assessment contributes actively to learning rather than merely certifying performance?

One particular group commented extensively on this question and their points are summarised here.

- One of the ironies of alignment is that assessment often stops learning that the student might be motivated to do; assessment does not often measure development of an individual.

- It was felt that in relation to Drew's presentation on level descriptors there might be disciplinary differences in the verbs used: perhaps there was also more emphasis than necessary on the 'pass descriptors'.

- One suggestion was that perhaps we shouldn't tell students what the criteria were as figuring them out was part of the development of understanding of what a degree is - LOs may promote too mechanistic a response to the subject.

- In the interests of effectively assessing the 'processes' as well as the knowledge we should move away from written exams and move towards oral examinations as you can mark understanding more readily. It was noted however that at least one Scottish higher education institution had recently completely abandoned the practice of oral examinations even for borderline candidates.

- It was perhaps easier to assess students 30 years ago when there were fewer of them because of this perhaps they are more dependent upon what they are told, there was a feeling among some academic that students were less likely to bring something unexpected to their contact hours.

- There was a feeling that the only way to get students to undertake assessment that would have been formative in previous years was to make it summative. It was noted, however, that all forms of assessment, even summative elements, were formative. There was general agreement that there should be feedback on summative examinations although it was recognised that it would be problematic there was no reason for not doing it. It was pointed out that

students could benefit from formative assessment with a link to summative assessment by giving them feedback to an assessment, for example, an essay and then allowing them to re-submit it amended or not as a summative assessment.

- It was possible to give feedback to the entire course rather than individuals, or even get the previous year group to give advice.

- It was suggested that feedback should be returned to students in a typed format, as nearly all assessment was expected to be submitted in typed format.

- It was argued that self-assessment promoted a dialogue with the tutor as students could identify their strengths and weaknesses and then the tutors could comment upon students' perceptions of their work.

- Was it dismissive of students to say they are only interested in the mark? Some courses had only pass/fail marking with feedback but no mark.

- One participant pointed to the example of Reed College in the United States where students were told none of the grades on any piece of work unless they were in imminent danger of failing.

- It was commented that at one institution in the United States, the criteria given at the beginning of course and then each semester and at the end of the semester/year, students 'bid' with tutors to demonstrate with any relevant evidence that they had acquired the outcomes. The aim was to promote a reflective learner (which is what several courses are aiming to do, ie teacher training). It was commented that at Maastricht the medical examination is based upon the same each year through using a huge exam bank and then taking a sample of questions from the bank.

- Staff time was extremely imbalanced between assessment and teaching - sometimes getting up to a 50:50 ratio. It was commented that often institutions do not give its staff enough time to make proper assessment in formative manner in the time allowed.

- Certainty and competency of marking are major problems; criteria might go along way to change that and create a fairer and more consistent assessment of students' work.

- Issues were emerging regarding e-learning and online assessment: it should be noted that these approaches require much up-front time and effort.

Assessment workshop series - No 4

Developing a variety of assessment methods, including self and peer-assessment

Contents	Page

Developing a variety of assessment methods, including self and peer assessment - An overview

Professor David Lines, Centre for Enhancement of Learning and Teaching, The Robert Gordon University and Workshop Director

Introduction

The material that follows is designed to support participants at the workshop entitled 'Developing a variety of assessment methods' to be held at The Robert Gordon University on 19 March 2004. The first part consists of a paper, written by David Lines with a substantial input from Win Hornby, both from The Robert Gordon University, and the second is further reading supplied by Win Hornby and Chris Rust of Oxford Brookes University, in support of their talks at the workshop.

For some readers much of what follows will be very familiar and rather elementary, especially in the following seven sections. These are designed more as an aide memoir to the fundamental definitions used in assessment and can be skipped if they are not needed.

More important in the context of the workshop are the later sections. These examine alternative assessment methods that can and have been used both in higher education institutions and in the workplace. A necessarily subjective opinion, though one based on the literature, is offered on each method's advantages and disadvantages. It is to be hoped that the workshop will provide more first-hand evidence to support (or not) these opinions.

In researching this paper, it was fascinating to go back to the work of Dai Hounsell and the Assessment Strategies in Scottish Higher Education (ASSHE) project team, who reported in 1996. The research found a wide variety of assessment techniques in use at that time, but in a somewhat haphazard way, with developments left largely in the hands of innovative and creative 'pioneers'. One is left speculating that a similar survey undertaken today might find much the same. That said, the document is well worth reading and provides a wealth of interesting approaches to assessment.

The paper that follows starts with definitions of fundamental terms used in assessment, but looks in more depth at the tension that exists between validity and reliability, and in particular the trade-off that often has to take place between them. An attempt is then made to 'map' the assessment methods described later in the paper against validity and reliability. Finally, alternative methods are described and evaluated.

Formative, summative and 'high stakes' assessment

Formative assessment is ongoing. It provides both education providers and students with information about progress in order to support future learning. **Summative** assessment provides information about the level of a student's performance at certain points in the learning process, usually at the end of a course of study. **High stakes** is when the result of a summative assessment has the potential to alter the course of the candidate's life in some way: the greater its impact, the higher the stakes. Final degree examinations are therefore good examples of high stakes summative assessments.

The purposes of assessment

Though different writers use slightly different terminology, they generally agree that assessment has three purposes (inter alia, Brown, Bull and Pendlebury, 1997; Yorke, 1998; Black, 1998). Firstly, assessment is designed to support and thus enhance learning. Secondly, it provides certification for progress or transfer, and thirdly it is a form of accountability (quality assurance) to stakeholders.

Validity

An assessment task is said to be valid when it tests what it sets out to test, though this is neither as simple as it may sound, nor does it tell the whole story.

The best-known types of validity are predictive, concurrent, construct and content (for further references see Peter Knight's *Briefing on Key Concepts* in the Learning, Teaching and Support Network's Assessment Series No 7 - now part of the Higher Education Academy) with a fifth, unitary validity, added more recently.

The notion of unitary validity rests on what inferences and actions are drawn from tests. So, no matter how well constructed a test, no matter how *valid* it is in terms of the criteria mentioned above, if the results are not used appropriately it will not be valid. According to this idea, it is the **use** of test results that determines validity.

Reliability

Probably for the simple reason that it is easier to do, reliability has been researched more thoroughly than validity (Wood, 1991; Black, 1998). Like validity, reliability consists of a number of sub-sets with complicated-sounding names, but all of them rest beneath the fairly straightforward notion that if we test someone then the test will be reliable if the result is exactly the same across all occasions, tasks, observations and settings.

Reliability versus validity

The extent to which validity **depends** on reliability will alter according to circumstances. In the case of high stakes, externally set and marked examinations, reliability is an essential prior condition to validity because the decision cannot be changed very quickly. For example, if university entrance depends on particular grades or scores on a test, it is vital that the marking and grade awarding processes are reliable in a technical sense (or at least **believed** to be so), because if they are not the entire system will collapse, as both candidates and users sense that on a different occasion or with a different marker the result would be substantially different.

However, ensuring a high level of reliability is very expensive because, at just a basic level, scripts have to be marked, remarked, checked against other markers and so on, and even then in, say, an essay paper, different examiners could (legitimately) argue for two different marks for the same work. This complexity explains the pressure for multiple-choice tests (and the main reason for their popularity in the United States until quite recently). Because of their high reliability at relatively low cost, multiple-choice tests have obvious attractions. But they are often not **valid** - they cannot test certain domains and are not necessarily good predictors. Therefore, given limited resources, an externally set and marked examination will often contain different types of papers, as a compromise or trade-off between reliability and validity.

There is yet another consideration and that centres on the cost of the people doing the marking. Jay Parks (2000) examined the relationship between examination reliability and the cost of performance. He argued that the notion of increasing reliability through more training and the provision of prompts for markers (ie mark schemes and hints on what to look for in answers) must also be limited by considerations of cost. Interestingly, he suggested that there need be no automatic assumption that one test equals one marker equals one testing occasion. It might be possible, for example, to increase the number of markers rather than the number of tests and thereby increase reliability at lower cost, while at the same time reduce the stress on candidates. Intuitively, this would seem to be the case, but an unreliable test remains unreliable no matter how many times it is marked. For instance, Newton (1996) investigated the consistency with which marks were awarded in mathematics and English examinations and found reliability to be consistently higher in the former. This is as one might expect, since tests in mathematics require fewer judgements on the part of markers. While increased training and guidance might close the gap, the intrinsic nature of the style of mathematics and English examinations means it can never be closed completely, and this is a price that has to be paid if validity is paramount.

Measuring validity and reliability in such a way that a meaningful figure can be obtained is impossible in practice because of the vast number of variables involved. The same test taken by the same candidate will yield a different result because it will be sat at a different time. Apart from the obvious impact of practice, the candidate will have changed - they may be more or less tired, more or less interested in the subject, more or less stressed by outside factors and so on. The same applies to markers, of course; only in machine-marked tests can this change factor be eliminated.

Because it is impossible to ascribe numerical values to validity and reliability in any meaningful way, Table 1 has been constructed to give readers an idea of **relative** validity and reliability for the different types of testing types that we will be examining in later sections.

Manageability

In simple terms, the notion of manageability means that the assessment task must not take excessive time or be overly complicated so that the costs involved, in the widest sense, do not outweigh the benefits.

Assessment conservatism

Despite the large number of alternative methods of assessment available, in practice few are used. Brown and Glasner (1999), for example, found that 90 per cent of a typical UK degree depends upon unseen time constrained written examinations and tutor marked essays and reports. It may be, however, that this balance is changing. Hornby and Laing (forthcoming) report that at The Robert Gordon University, based on a sample of 168 modules at both undergraduate and postgraduate levels, 50 per cent of assessments are now conducted by way of coursework with only three per cent entirely by unseen examination. That said, unseen examinations still play a part in the remaining 47 per cent of assessment, so if there is a movement towards change, it is slow.

	Low	Medium	High
High	● Portfolios ● Essays	● Critical incident accounts ● Direct observation ● Orals and presentations ● Learning contracts/logs and diaries	● Self and peer assessment
Medium	● Cases Open problems In-tray exercises ● Annotated bibliographies Articles and book reviews	● Extended comp exercises ● Short answer tests	● Computer aided assessment
Low			
	Low	**Medium**	**High**

(vertical axis label: Increasing validity — horizontal axis label: Increasing reliability →)

Table 1 The relative validity and reliability for the different types of testing.

Notes

1 Examiner training can increase both validity and reliability. Of the two, reliability is probably easier to improve by means of marker training/better mark schemes/hints at likely answers etc. Validity can also be improved, however, through paper setting training - involving the identification of the domains that are to be examined and appropriate questions that will test those domains.

2 Improved technology is increasingly enabling examiners to set questions that are valid and, because they are machine-marked, highly reliable. In the future, therefore, CAA is likely to move to the upper right quadrant. Machine-marked essays are currently subject to a great deal of research and, though some disagree, their proponents predict an imminent breakthrough.

3 There is an apparent trade-off between validity and reliability (Newton, 1996). However, low reliability is deemed 'acceptable' in many systems because validity is seen as more important, especially in a formative/training context. Such a position is controversial, and might be subject to a lengthy debate.

A survey of assessment methods

Sources

The section that follows examines a range of alternative assessment methods. Each method is explained and the main advantages and disadvantages are outlined.

The methods are divided into three. The first fall into what are here called 'conventional' or non work-based, the second, primarily work-based and the third, those that can be used both inside and outside the workplace.

This section draws on work by Rowntree (1987); Wood (1991); Gibbs et al (1993); Brown et al (1994); Cotton (1995); The ASSHE Inventory (1996); Wakefield (1999); and the UK's Learning and Teaching Support Network (LTSN) Generic Centre (2001). Where appropriate, reference is made to empirical and theoretical work as it relates to each assessment method.

'Conventional' or traditionally non work-based assessment methods

Annotated bibliographies, articles and book reviews

In order to encourage students to read more widely, they can be set the task of preparing annotated bibliographies and/or book reviews. The principal advantage of this method of assessment is that it will reward students who undertake wide reading. In the case of a book review, it also encourages them to critically evaluate the work. Students might also be asked to prepare an article for a journal or periodical which has to be accessible to a general reader by a deadline and within a word count. Such an activity encourages students to manage their time and to produce work which avoids jargon. It also rewards the ability to condense complex ideas in a simple and effective way. At its best it can be an effective method of getting students to undertake wide reading for which they receive an immediate and direct benefit. If the required learning outcomes in a course stress that students should display evidence of wide reading and research, this method is ideal and is 'constructively aligned' (Biggs, 1996).

The danger of this method is that at its worst it may encourage students simply to paraphrase the articles or books and not apply their knowledge appropriately or evaluate the material. It also suffers from the fact that it requires considerable staff time to assess, particularly if the bibliographies are substantial. Staff may also need to undertake the reading themselves to be able to judge whether the work has been correctly represented in the bibliography.

Cases and open problems

Case studies are a popular method of both teaching and assessing, particularly in a business and management setting (Shapiro, 1994; Boehrer, 1995). Case study assessment has three main components: the case material, the students' preparation of that material and an examination based on the case. The questions may be seen prior to the examination, but more commonly they are not. The case study may be given out prior to the examination or may be issued on the day. If it is issued in advance then there is clearly an expectation that the students will have undertaken some preparation. Unlike problem-based assessments (see below), the case traditionally does not set out what the problem is. Nor does it lead to a single correct solution. A classic case study examination attempts to assess the student's ability to recognise the nature of the 'problem' first, before suggesting alternative solutions. To

that extent, at its best, it has the potential to test a wider range of higher cognitive skills in a more 'realistic' setting than more conventional methods.

Cases may also be used to assess group work. In these circumstances, other transferable skills can be assessed such as those of team working, presentation, research and time-management. Group work may be assessed by peers (see below) or conventionally, by a tutor.

Velenchik (1995) argues that the case method of teaching and assessing can provide a powerful incentive for students to learn theoretical concepts. It encourages the application of theory to practice and the use of empirical evidence and data to support recommendations and conclusions. Crucially, it offers students the opportunity to explore theoretical limitations.

The case method turns the traditional sequence of theory followed by practical examples on its head, instead moving from practice to theory. By so doing it tests the ability to sift and search through the candidate's theoretical 'tool box', to select the appropriate models to understand the problem and find solutions. It can therefore be used to move students up the Bloom's cognitive skills ladder from knowledge and comprehension, to analysis, synthesis and evaluation (Bloom, 1956).

The use of the case study method as a vehicle for assessment is described by Carlson and Schodt (1995) who report assessing their students through a combination of examination (70 per cent), coursework and case study work (30 per cent). They point out that there are costs, both in educational and resource terms to assess through cases. Firstly, it takes much more time (and sometimes, money) to prepare an appropriate case and to obtain the necessary copyright permissions. If the case is to be issued before the examination there are issues surrounding the distribution of the material and the administrative and delivery costs involved. Generally speaking, case studies can focus on fewer areas of the syllabus than other methods of assessment such as multiple-choice tests or computer assisted assessment (CAA) methods (see below), so arguably case study assessment sacrifices breadth for depth.

From a student's perspective there can be greater risks with a case study as a form of assessment. It places less of a premium on knowledge and understanding, and instead emphasises higher cognitive skills. In addition, if students make an incorrect judgement as to the nature of the problem presented, the risks of failure may be greater than with other forms of assessment that are less 'all or nothing'.

In-tray exercises represent a form of case study. Here the candidate is faced with a number of memos (or emails) that have arrived at once. No specific instructions are given, since the exercise is a business simulation designed to test the ability to synthesise information, evaluate it and then make decisions. Generally, the memos have no apparent connection, but the careful reader will spot, for example, that certain names keep appearing, or that some of the data is less 'credible' than others. Not only do in-tray exercises test the same high level skills as case studies, they probably simulate real-world environments better, simply because they are played out in the first person - the decision-maker is the candidate.

Because there is no single 'right answer', it may be that in-tray exercises are better used formatively than summatively, since reliability is an issue. However, the learning

experience can be profound, an experience that can be heightened still further with careful feedback.

There is another danger, and one common to all forms of examination where marking is undertaken on the basis of percentages or other numeric forms and that is the ease with which high marks (and low ones) can be justified where clear 'right' and 'wrong' answers can be identified, but which are not given in more discursive papers. If such a situation is not handled with care, in a multi-component examination, consisting of papers involving calculations and others requiring extended writing or essays, the numerical papers may distort the overall mark. Take, for example, the following figures from a pre-degree course for entry into a British university, sat by students in the Far East in 2003:

Paper title	Mean	Standard deviation	Minimum mark	Maximum mark
IT for Business (N)	80.35	13.72	41.00	94.00
Principles of Accounting (N)	77.12	12.74	36.00	92.00
Mathematical Techniques (N)	77.56	21.96	7.80	100.00
Principles of Business (D)	57.65	9.64	41.00	71.00
The Local Economy (D)	53.65	9.51	34.00	69.00

N=Numerical; D= Discursive

Table 2 The impact of numerical and discursive questions on an examination

In each of the numerical papers the mean mark is considerably higher, as is the standard deviation. What this means is that markers are confident about awarding high and low marks (note in one paper 7.8 per cent and 100 per cent were given to two candidates) whereas in discursive papers, not only is the mean mark lower, there is considerably less discrimination (more 'bunching' around the mean). While it could be argued that the range of marks awarded on the numerical questions are more valid, as it differentiates between the strong and weak student, if all the results are averaged without any form of weighting, they benefit those candidates who are numerically gifted at the expense of others who are not.

Computer assisted assessment

CAA refers to the use of computers to assess students progress. The format of CAA can vary considerably, from paper and pencil tests which are processed automatically using an optical mark reader, to students inputting their responses directly into a computer terminal. The nature of the assessments can also vary. The conventional view that CAA consists entirely of four or five multiple-choice options is not borne out by the range and variety of questions that can be set. Reason-assertion, data response, mix and match, click and drop menus with options, as well as short answer questions are now commonplace in most CAA packages. They may be used diagnostically before a student embarks upon a course, formatively to assess what progress a student is making, or summatively to give a final grade or mark. Tests can be supervised or unsupervised depending on the circumstances. If used either formatively or for diagnostic purposes, then students need not be supervised. Used summatively, supervision does become an issue, however.

There are several powerful advantages both educationally and in resource efficiency terms in using CAA. Educationally, if used diagnostically or as formative assessment, it can give students important feedback about what learning they have achieved. Not only do they obtain marks and grades quickly, they can also receive help on how to improve performance. In addition, staff can also get important information about how well the course is being taught. Content analysis of responses to questions can reveal which areas are causing students the most difficulty and appropriate modifications can be made to the teaching. At its best, CAA can be a powerful tool for enhancing the learning experience.

From a resource perspective, CAA can be very attractive. It saves considerable time in supervision, invigilation and marking. It also improves marking reliability. When dealing with large groups it makes it easier to give quick and effective feedback.

However, there are some disadvantages of CAA. Although it is **possible** to test higher order skills with CAA, it is more difficult. For instance, it is harder for students to demonstrate their ability to communicate or their propensity for original thinking.

It has been claimed that there may be some cultural or gender bias in using CAA. In particular, it is claimed that female students do less well in CAA although McKendree (2002) suggests that current evidence indicates only a weak link.

CAA also relies on a certain level of information technology skills and competence and it is easy to overestimate students' familiarity with systems, especially under examination conditions. Plagiarism can be a serious issue as students can potentially view the work of other students more easily than in conventional examinations. To some extent this can be overcome by randomising the questions, thereby eliminating the possibility of machines in close proximity having identical screens. However, unless blocking mechanisms are put in place, students may still have access to information 'outside' the test situation via the internet. Enhanced computer security or encrypted password access needs also to be in place to ensure that students cannot hack into the system and access the test database.

There are also technical issues associated with computerised examinations; since there is always the danger of 'personation', in other words that the candidate is not the individual striking the keyboard. There are ways around this, such as taking digital photographs of candidates or taking their thumbprints, but all of these devices add complexity and cost to the system. Furthermore, although this aspect has been little researched, there is always the possibility of what Huff and Sireci (2002) call 'construct-irrelevance variance' in computer-based tests. What they mean by this is the computer offers the examiner the ability to construct tests that could only be done in a computer-enabled environment. For example, it would be possible to introduce video clips and interactive assessments that would be impossible in a 'normal' examination. There is always the danger then that the test becomes invalid as it tests computer skills, or familiarity with software, rather than the skills and knowledge of the subject itself.

There is also the important question of the reliability of the technology or the 'platform' that is used to support CAA. This can be especially critical if CAA is used for

large numbers of students and arrangements need to be in place to have spare terminals and machines available in the case of breakdown or to allow tests to be undertaken by more conventional 'paper and pencil' means.

There are also staff development costs. People need to be trained in the design of suitable questions, on the technology for loading them onto the system and for releasing the tests. Development costs can be significant, although these can be reduced by creating and sharing banks of questions across a number of institutions. Set up costs can be significant as CAA will involve cooperation across several different groups of professionals (academics, educationalists, examination or awarding bodies and information technologists). There are also issues of space and access to terminals, especially if there are large groups of students being tested at the same time. CAA may also conflict with other policies on assessment, for example those on anonymous marking.

Multiple-choice, objective tests

One of the most widespread forms of CAA is multiple choice, or objective tests. Although multiple choice tests need not be administered or marked by computer, they offer virtually 100 per cent reliability if they are, and this is their great strength. In addition, they can pose a large number of questions in a relatively short time, thus enabling an extensive coverage of the course content. But at the same time, should the examiner wish to focus on specific knowledge and skills that can also be done.

Multiple-choice tests do not offer choices to candidates. That means that the test is identical for all. In addition, because items are pre-tested, their difficulties will be known in advance. The test can therefore be modified to suit the required assessment need.

It is often the case, (though not always) that questions can be banked and re-used many times. So, multiple-choice questions score highly in terms of efficiency - they are easy to administer and mark, and candidates can sit them at different times and different locations.

Multiple choice, objective tests do have drawbacks, however. They are difficult and expensive to construct initially, and if they are 'banked' they must be withheld from candidates, so transparency falls. Pre-testing is also required, which is again resource intensive. From the candidates' point of view, it is not always possible to see why a question was wrong, although there are differences between using multiple choice in formative as opposed to summative settings, as was emphasised under CAA. Furthermore, no credit is given for workings that display knowledge of the principles: answers are either right or wrong.

Among candidates, guessing a multiple-choice answer is common practice and a good tactic so long as wrong answers are ignored. Allen, Swain, Sudweeks and Schaalje (1998) compared multiple-choice with 'free response' questions given to 187 introductory accounting students. The students performed better on the multiple-choice tests, which indicated that guessing took place and 'enhanced' the test score. Furthermore, the more intelligent the student, the lower the impact of guessing (because there were fewer items that the candidate was unsure about). These two factors together result in 'bunching', with 'bottom end' students doing better than they would on a free response test.

Overall, multiple-choice tests offer considerable advantages in a formative setting, especially if computer assisted. That they have a summative role is less certain, however.

Short-answer tests

Short-answer tests are composed of items that are similar to objective items, in that a clearly defined answer is required, but they differ from them in that the answer has to be supplied by the person being tested rather than simply chosen from a number of options provided. Such items are much easier to write than objective items and are also much more versatile, since they can take a wide range of forms. They range from items that have the same almost perfect reliability as true objective items, because the answer provided will unambiguously be either right or wrong, to items where the reliability is somewhat lower because there is some degree of subjectivity in deciding whether the answer is correct or not.

Like multiple-choice tests, short answer tests can offer extensive syllabus coverage, they can focus on specific knowledge and skills, questions can be 'banked' and they offer students no choice.

They are also relatively quick and easy to mark, though they cannot be machine marked in the same way as multiple-choice tests can be. Certainly, they are easier to write then multiple-choice tests and they are more versatile.

On the other hand, short answer questions are less reliable than multiple-choice tests, though they are better than, say, essays in this department. They generally score quite well on validity, except where 'short answers' simplify what are in reality quite complicated ideas.

Short-answer tests thus have most of the advantages of objective tests without suffering from the same weaknesses (difficulty of writing items, need for items to be tested before use etc). Indeed, virtually anything that can be assessed by means of an objective item can equally well be tested by making use of a corresponding short-answer item.

Extended-answer tests

Until comparatively recently, these were by far the most commonly-used assessment method in many institutions, with the 'answer any five out of eight questions in three hours' examination paper being typical. Such extended-answer papers were and remain popular with setters, since questions are relatively easy to construct (if not so easy to mark), and they were, and perhaps still are, also fairly popular with students - mainly because they afford considerable scope for question spotting and ignoring sections of the syllabus that are found to be difficult.

The reliability of extended answer questions, on the other hand, is highly questionable, especially if only one marker is involved. Increasing use of 'blind, double-marking' has addressed some of the concerns expressed over this issue, but has added cost and complexity.

Work-based assessment methods

Critical incident accounts

Critical incident accounts are used to assess the lessons that can be learned from a key incident that has occurred in the work place. Such incidents often relate to a

problem, breakdown or other crisis. They may deal with life threatening situations or with more mundane everyday issues which can nevertheless affect the organisation or have an impact on the learner. They can often be used effectively in the workplace to encourage reflection on ways a situation might have been avoided or how the reaction to a crisis might have been improved. They require learners to display a whole range of self-assessment and self-evaluation skills that are simply not possible to assess in any other way. They have considerable advantage in that such incidents will not be artificial but instead will be based on real examples. Used effectively, not only does the trainee learn from the experience of compiling these accounts, but if used in conjunction with others, different perspectives on the same incident can be compared and contrasted. The organisation may also benefit from taking time to reflect on the lessons learned. Critical incident accounts can form part of a portfolio of evidence obtained from the workplace, which can then be used as evidence of the competencies sought by employers (see below). Therefore there are considerable educational advantages to using this method of assessment.

There are disadvantages, however. It is difficult to programme 'critical incidents' if they are to appear to be 'naturalistic'. Before learners embark on such assessments they must have an opportunity to develop their self-evaluation and self-assessment skills. In some cultures and in some organisations it may be difficult to obtain frank admissions of limitations or to be self-critical.

Nevertheless, the administrative costs of setting up such assessments are not as high as, for example, CAA (see above). If one imagines a number of trainee accountants in the same working environment, a given incident can give rise to multiple critical accounts, providing a cost-effective means of assessment, quite apart from the educational benefits.

Direct observation

In some professions, such as teaching and medicine, and in some vocational situations, direct observation of a candidate is a central element within the overall assessment regime (Wolf, 1995; Leask, 1999). Such observations may be formal, in the sense that they contribute to the overall score of the candidate, with observations made at an agreed, pre-determined time and place, or informal, meaning that the assessor will watch the student over a period of time and make a judgement based on a summation of what has been seen.

Direct observation has considerable benefits in a formative setting because the context will be shared by both the assessor and the candidate, giving the opportunity for a constructive reflection on what was learned, an examination of alternative strategies that could have been applied, any theoretical concepts illustrated by the situation and so on.

When used summatively, the pressure on the candidate increases considerably, especially if the number of observations is limited. It might be possible, for instance, for a single observation to encounter an entirely atypical situation resulting in a distorted view of competence. The candidate might not be on their best form and so on (driving tests and job interviews fall into this category). In addition, in both formative and summative settings there is a danger of the so-called 'Hawthorne effect', that is to say the observations are affected by the very act of observation.

The fact that observations are by definition, unique, makes reliability a major problem. Two observers of an identical situation may, for quite legitimate reasons, have completely different views on what happened, how the situation could or should have been dealt with, the lessons learnt and so on. Extensive training, including teamed observation, is required in order to counter this problem. It also means that any attempt to grade such observations, other than on a pass/fail basis, should be treated with the utmost caution. Nevertheless, they can have an important role, either as a formal or informal element within a work-based assessment regime.

Learning contracts

Learning contracts can be used both in an educational environment and in the work place. There are a significant number of studies which report their use principally as part of the assessment of the 'sandwich' element of degree programmes (Dorsman, 1985; Chatterton et al, 1988; Ashworth and Saxton, 1989; Davies, 1990; Stephenson and Laycock, 1993; Hornby and Gammie, 1994).

Learning contracts are used to construct an individual's programme of learning based upon an assessment of the learners' current competencies, compared to the level they wish to achieve.

There are significant advantages in using learning contracts. For example, Goodwin and Forsyth (2000) argue that the contracts allow learners to negotiate non-standard programmes reflecting both their own professional needs as well as those of the employer. In addition, the contract helps to spell out the respective roles of the learner, the employers and the educational provider.

The major disadvantage of learning contracts is that they require the learner to have the necessary skills to identify where they are placed in terms of their own skills and competencies. It also requires students to be able to carry out a needs analysis. This may require a good deal of staff input at the outset to achieve the required level of competence, even before the learner can identify competence gaps.

There is not usually a problem with verifying that the contract is the learner's own work, since the process of drawing it up is necessarily a cooperative one, usually with the employer and the educational provider involved. Learning logs and diaries (see below) are often used in this context. However, although on the face of it much of the curriculum planning and execution passes to the learner, the assessment load on tutors is shifted from being a provider of knowledge into that of supporter and facilitator of student-directed learning.

Learning logs and diaries

Learning logs and diaries have a similar potential to critical incident accounts insofar as they encourage reflection on what has been learned. Self-reflection and self-evaluation are again required to make this method of assessment effective. With this method students keep a note over an agreed period of what goals or objectives have been set and how they have been achieved.

The key educational advantage of logs and diaries is that they provide a tangible record of what learning has taken place during a work placement. They are therefore constructively aligned with the learning outcomes associated with work-based

learning. This makes them a valid method of assessing work-based learning and as a result they have long been recommended as an important tool in this area (Nixon 1990; CNAA, 1991) especially if they are used as reflective journals in which students view events through a self-assessment prism (University of Huddersfield, 1992).

The major drawback with the use of such methods is that they can simply become a descriptive diary of events. They may become repetitious and show no evidence of learning by doing. If students have not been trained nor had any experience of self-assessment then they may simply lack the necessary skills to undertake the task. Alternatively, achievement logs set out prescribed competencies which require to be achieved, as evidenced by a tick in the requisite box. While this alleviates repetition, arguably it is weak on evidence and hence the potential for lack of reflection.

Assessment of journals and logs can become problematic. Grading using this assessment method is difficult, especially if it is used summatively. Based on an informal review of those courses in accounting, business and management in higher education in Scotland that employ such methods and from a review of the literature (Smith and Wilson, 1992), it appears that they are most commonly used in sandwich programmes that have a placement element as part of the assessment of work-based learning. They are more commonly used as supporting evidence of completion of the placement where their fidelity and veracity are often confirmed by an in-company mentor who is required to 'sign them off'. They are also commonly assessed on a completion/non-completion basis and no grade is assigned to them. One of the issues however, is the potential variability in mentor evaluation of competence which can call into question the reliability of this method of assessment from one work experience to another.

To some extent the evaluation of such logs and diaries depends on how they are implemented. If they are seen as part of a suite of assessment instruments for assessing work-based learning that go to make up a student's portfolio (see below) then they can be highly effective. For example, students might be asked to produce reflective practice assignments (see below), which require them to produce a report on what has been learned after a period of practical experience. The diaries or logs therefore can provide the source documentation for such reports and can be referenced. Critical incidents accounts (see above) that are required as part of the assessment of experiential learning can be dated and referenced and recalled some time after the event using the learning log/diary. In addition, if they are verified by in-company mentors and form part of an appraisal system, the incentive to cheat is reduced and indeed becomes self-defeating for the learners. If they are assessed on a completion/non completion basis then time and effort in grading is significantly reduced. Used as formative assessment, learning logs/diaries can be extremely valuable for students. To use Brown's terminology, assessment by the use of learning logs/diaries will tend to focus on the acquisition of skills as opposed to knowledge, on the 'process' of learning as opposed to 'product' and to focus on 'competence' as opposed to 'grading'. In these circumstances learning logs and diaries can be both effective and efficient, especially in providing evidence to support the acquisition of the necessary skills and competencies gained in the workplace.

Assessment methods that can be used both inside and outside the workplace

Self and peer-assessment

Self and peer-assessment by students has long been considered an important method of assessment. For well over 20 years the literature on assessment has made reference to the advantages of students assessing themselves (see separate references).

In a number of the papers in the research literature, the emphasis has been less on how to do self and peer-assessment and more on empirically testing its reliability and validity. For example, Mabe and West (1983) undertook a comprehensive review of 55 studies in which the self evaluation of ability was compared with other tutor assessed measures of performance. They found a very close correlation between the two assessment measures. More recently, in one of the most comprehensive research reviews on self-assessment, Boud and Falchikov (1995) examined a total of 68 studies undertaken between 1932 and 1994. They looked at how frequently students over rate themselves (17 studies) as opposed to under rate (11 studies); at whether good students rate themselves more accurately than poorer students (11 studies in total, of which nearly all studies showed that better students were in fact more accurate or rated themselves below tutors grades than poorer students). They also reviewed seven studies on whether more advanced students rated themselves more accurately than students at an introductory level (undergraduates and 'freshmen' overestimated their abilities more than postgraduates and senior years). In addition, they reviewed seven studies which looked at whether students got better with practice (they did not). Finally, they reviewed six studies on gender differences and they discovered only three that found women more accurate self assessors than men, whilst more recent work by Stephani (1994) found no gender differences. Other studies looked at the effect of self-assessment on learning. For example, Fitzgerald et al (1997) undertook an analysis of medical students' self-assessment of knowledge and found significant improvements. Finally, in one of the most comprehensive and rigorous empirical studies to date, Rust et al (2003) report on the findings of a two year research project developing students' understanding of assessment criteria. By a planned intervention allowing students to 'engage' with the criteria on which they are assessed, Rust reports that students learning can be improved and that this improvement can not only be significant but it can also last over time (see also Hornby and Laing, forthcoming).

There are considerable advantages in using self and peer-assessment methods. According to Brown and Knight (1994), the development of self-assessment skills are important in themselves in developing learners' own abilities. Indeed, it can be argued that becoming self reflective is the basis for becoming a life long learner. If successfully achieved, it is a skill that can underpin many valuable assessment methods, especially those in the workplace (see below).

Critics argue that self and peer-assessment are unreliable, but it should be noted that the benefit of self-assessment comes not so much in getting the grade or mark right (important though this might be), but in the process of reflection on how the work or performance in the task could have been improved. Peers, it is reported by Gray (2001) are more likely to over-grade a learner's work than under-grade it and the danger of collusion is always likely to be present. Weaker students can 'hide' in more proactive and stronger groups and so benefit from being a 'free-rider'. While

there are a number of strategies for countering such effects, they tend to add to the costs of implementation.

While recognising that if implemented properly, peer assessment offers valuable learning opportunities, it has a problem with basic credibility across many stakeholder communities, especially perhaps, outside of education.

Orals and presentations

Communication skills are highly valued with an increasing emphasis being placed on this area by professions throughout the world. However, while oral examinations are a good way of assessing communication skills, they are also a good way of examining students' problem solving abilities. Orals allow for follow-up questions and probing in a way denied to other forms of assessment. They also have the advantage of immediacy and the added advantage that it is difficult to cheat. However, orals are very time-consuming to stage as they are normally done on a one-to-one basis and they are therefore unsuitable where large numbers of students are concerned.

Portfolios

Assessment by portfolio has a long tradition, stemming from its use in the visual arts. However, there have been a number of factors in recent years that have prompted a growing interest in portfolio assessment (whether it is paper-based, multimedia or online) in other areas. There is increasing pressure from employers who want to see what applicants can do as well as what they know. As more courses become concerned with testing abilities and competencies that are relevant for the workplace, portfolio assessment is seen as an attractive option. There is also an increasing realisation among employers and professional bodies that conventional forms of assessment often test only a narrow range of knowledge and abilities.

A portfolio is essentially a collection of items, rather than a single piece of work. It therefore differs significantly from other forms of assessment insofar as it attempts to produce multiple sources of evidence to verify claims for achievement of learning outcomes. Typically, portfolios will be used to demonstrate the achievement of several learning outcomes over a period of time.

Portfolios offer several important educational advantages. Firstly, they are clearly a valid form of assessment especially if students are engaged and involved in their development alongside their tutor/mentors (Palomba and Banta, 1999). Secondly, they can allow the collection of evidence of sustained pieces of academic, professional and personal work, thereby demonstrating attainment of the course's specified learning outcomes, which is fundamental to any educational course (Baume, 2000).

Provided that there are not too many separate and prescribed learning outcomes that have to be attained, and that assessors are briefed and trained in portfolio assessment techniques, then portfolios will be reliable (inter alia Herman et al, 1992; Palomba and Banta, 1999; Palomba, 2002). On the other hand, if there are multiple outcomes of increasing detail with many hurdles that learners have to clear, if the criteria for assessing are unclear and if the assessors are not well briefed, then portfolio assessment will be unreliable.

As with all assessment methods openness is important. It becomes particularly critical with this form of assessment. For example, should students display just their best work in the portfolio, allowing them to make the final selection? Or should the portfolio consist of all the evidence, the good and not so good? The answer to these questions crucially depends on the perspectives of the various stakeholders in the assessment process and their relevant interest and influence. For example, if an institution wants to allow students to show the very best of what they are capable, then it will allow the selection of the best work to be presented. If, on the other hand, there is a concern for professional recognition then it may be necessary to have either the full range of evidence, or more reasonably, a representative sample of the range of the work presented to ensure that performance has never dropped below some minimum acceptable standard. Whatever is decided should be made explicit at the outset.

There are, however, significant management/resource issues to consider. The assessment process involved with portfolios moves out of the classroom or the examination hall and moves instead into the workplace. With that move, some of the costs of assessment are likewise transferred from one set of stakeholders to another. In addition, there are issues surrounding the authenticity of the evidence presented. This is something that applies to any form of assessment which is conducted in the students own time, but it has particular force in portfolio assessment where the assembling of evidence is entirely at the discretion of the student. Fortunately, there are ways of addressing such issues. Evidence can be authenticated by trusted third parties, such as recognised mentors or accredited in-company supervisors. Assessment criteria can be weighted in such a way as to reward referencing and especially critical review. Plagiarism becomes more difficult as the task is more geared to an individual learner, to their interests and experiences, which is exactly what portfolio assessment is designed to achieve (Baume, 2001). However, there does remain the problem of variability or indeed lack of work experience.

Not withstanding this and to summarise, the portfolio has considerable benefits, but as with most successful methods of assessment, such value-added is not costless. It is more complex to manage than some others because it is so individualised, which, somewhat ironically, is also its strength as far as the learner is concerned.

References

Allen R D, Swain M R, Sudweeks R R and Schaalje G B (1998) The impact of recognition opportunities and examination format on students' examination performance in introductory accounting, *Advances in Accounting Education,* 1, (Jan) 189-214

Ashworth P and Saxton J (1989) Experiential Learning during Sandwich Degree Placements and Questions of Assessment, *Sheffield Papers in Education Management*, Sheffield City Polytechnic, Sheffield

Baume D (2000) Dialogues: Assessment, *Educational Development,* 1(3)

Black P (1998) *Testing: Friend or Foe? Theory and Practice of Assessment and Testing*, Falmer Press, London

Bloom B S (ed) (1956) *Taxonomy of Educational Objectives Handbook 1: Cognitive Domain*, Longman, Green and Co, New York

Boehrer J (1995) How to teach a case, *Kennedy School of Government Case Programme*, Case No C18-95-1285.0
www.ksgcase.harvard.edu

Boud D and Falchikov N (1995) What does Research Tell us about Self Assessment? in D Boud (ed) *Enhancing Learning through Self Assessment*, Kogan Page, London

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

Brown S and Glasner A (eds) (1999) *Assessment Matters in Higher Education: Choosing and Using Diverse Approaches*, Open University, Buckingham

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

Chatterton D, Robert C and Huston F (1988) *Learning from Experience: assessing Supervised work experience*, Ealing College of Higher Education/CNAA, London

Council for National Academic Awards (1991) *Development of Placement and Assessment Procedures in Supervised Work Experience*, CNAA, London

Davies L (1990) *Experience-based Learning within the Curriculum*, ASET/CNAA, London

Dorsman M (1985) *An Investigation into the Measurement of Student Performance During Supervised Work Experience in Undergraduate Sandwich Courses*, MPhil thesis, Manchester Polytechnic, Manchester

Fitzgerald T J, Gruppen L D, White C and Davis W K (1997) Medical Students' Self Assessment Abilities: Accuracy and Calibration, *Proceedings of the Meeting of the American Educational Research Association*, Chicago

Gibbs G, Habeshaw S and Habeshaw T (1993) *53 interesting ways to assess your students*, Technical and Educational Services, Bristol

Goldstein H (1996) The statistical analysis of institution-based data, in H Goldstein and T Lewis (eds) *Assessment Problems: Developments and Statistical Issues*, Wiley, Chichester and New York

Goodwin M and Forsyth H A (2000) A Development of Professional Studies by Negotiated Work Based Learning, in *The Impact of Work Based Learning: Proceedings of a Conference, Cambridge 14/15 December 2000,* Work Based Learning Network of the Universities Association for Continuing Education, Cambridge

Gray D (2001) A Briefing on Work-based Learning, *Assessment Series no 11*, LTSN Generic Centre, London
www.heacademy.ac.uk/resources.asp?process=full_record&section=generic&id=11

Herman J L, Aschbacher P R and Winters L (1992) *A Practical Guide to Alternative Assessment, Association* for Supervision and Curriculum Development, Alexandria

Hornby W and Gammie E (1994) Learning contracts and sandwich education: the accreditation of work-based learning, *Capability,* 1(2) 46-58

Hornby W and Laing D (forthcoming) Efficiency and Effectiveness in Assessment: an empirical analysis of university practice (forthcoming at www.ssrn.com/)

Huff K L and Sireci S G (2001) Validity Issues in Computer-based Testing, *Educational Measurement: Issues and Practice,* 20 (3) 16-25

Learning and Teaching Support Network (2001) *Assessment Series*, LTSN, York

Leask M (1999) Improving your teaching: an introduction to action research and reflective practice, in S Capel, M Lesk and T Turner *Learning to Teach in the Secondary School,* Routledge, London

Mabe P A III and West S G (1983) Validity of Self Evaluation of Ability: A Review and Meta-Analysis, *Journal of Applied Psychology*, 67 (3) 280-296

McKendree J (2002) *Negative marking and gender bias,* CAAC http://caacentre.lboro.ac.uk/resources/faqs/negative-marketing.shtml

Newton P E (1996) The Reliability of Marking of General Certificate of Secondary Education Scripts: Mathematics and English, *British Educational Journal,* Vol 22, No 4, 405-20

Nixon N (1990) *Assessment issues in relation to experience-based learning within courses*, Council for National Academic Awards, London

Palomba C A (2002) Scholarly Assessment of Student Learning in the Major and General Education in T W Banta (ed) *Building a Scholarship of Assessment,* Jossey-Bass, San Francisco

Palomba C A and Banta T W (1999) *Assessment Essentials,* Jossey-Bass, San Francisco

Parkes J (2000) The Relationship between the Reliability and Cost of Performance Assessments, *Educational Policy Analysis Archives,* 8 (16)

Rowntree D (1987) *Assessing Students, How Shall We Know Them?* Kogan Page, London

Rust C, Price M and O'Donovan B (2003) Improving Students' Learning by Developing their Understanding of Assessment Criteria and Processes, *Assessment & Evaluation in Higher Education,* 28 (2) 147-164

Shapiro B P (1984) Hints for case teaching, *Harvard Business School,* 9-585-012

Shermis M D and Daniels K E (2002) Web Applications in Assessment, in T W Banta (ed) Jossey-Bass, San Francisco

Smith D and Wilson B (1992) The development and assessment of personal transferable skills during work-based placements, *Assessment and Evaluation in Higher Education,* 17 (3) Autumn, 195-208

Stephani L A J (1994) Self, peer and group assessment procedures, in I Sneddon and J Kremer (eds) *An Enterprising Curriculum: Teaching Innovations in Higher Education*, Kogan Page, London

Stephenson J and Laycock M (1993) *Using Learning Contracts in Higher Education*, Kogan Page, London

University of Huddersfield (1992) *A self-assessment model for the integration of work-based learning with academic assessment,* University of Huddersfield, Huddersfield

Velenchik A D (1995) The case method as a strategy for teaching policy analysis to undergraduates, *Journal of Economics Education,* 26 (1) 29-38

Wolf A (1995) *Competence-base Assessment*, Open University Press, Buckingham

Wood R (1991) *Assessment and Testing: A Survey of Research*, Cambridge University Press, Cambridge

Yorke M (1998) Assessing capability, in J Stephenson and M Yorke (eds) *Capability and quality in higher education*, pp174-191, Kogan Page, London

References for Chris Rust's presentation

Angelo T and Cross K P (1993) *Classroom Assessment Techniques*, Jossey-Bass, San Francisco
Features 50 techniques, presenting step-by-step procedures for adapting and administering them, and practical advice on how to analyse the data, pros, cons, caveats and other useful information.

Atkins M J et al (1993) *Assessment Issues in Higher Education*, Department of Employment
An excellent and comprehensive summary of the key issues.

Banta T W et al (1996) *Assessment in Practice,* Jossey-Bass, San Francisco
An American compendium of detailed examples of assessment practices and strategies from a wide range of disciplines and institutions.

Brown S et al (1994) *Strategies for Diversifying Assessment*, OCSD
A practical workbook to help you consider alternative strategies for assessment, whether they would be appropriate in your context and why you should consider them.

Brown S et al (1996) *500 Tips on Assessment*, Kogan Page, London
Exactly what the title says, this book can provide both useful checklists and a starting point for ideas you might want to follow-up and explore in more detail elsewhere.

Crooks T (1988) *Assessing Student Performance,* Higher Education Research and Development Society of Australasia (HERDSA Green Guide)
This guide to assessment gives brief and practical information on how to organise assessment across a range of learning outcomes.

Gibbs G (1992) *Assessing More Students: Book 4 of the Teaching More Students project,* PCFC - OCSD
Again, this book was originally designed to support a workshop, but it is a useful free-standing publication. It includes a review of students' and lecturers' experiences of assessment and current assessment methods together with a section on assessment strategies. Graham Gibbs edited the five books for this series; the design team for this one was Alan Jenkins of Oxford Centre for Staff and Learning Development, Oxford Brookes University and Gina Wisker of Anglia Polytechnic University.

Gibbs G, Habeshaw S, Habeshaw T (1993) *53* interesting ways to assess your students, third edition, *The Times Educational Supplement*

Hounsell D et al (1996) *The ASSHE Inventory*
An inventory of assessment strategies being used in Scottish higher education presented as case study examples.

LTSN Generic Centre Assessment Series, various authors, 2001
12 volumes covering a wide range of assessment topics, available at
www.heacademy.ac.uk/resources.asp?id=1&section=generic&process=filter_fields&type=all

Miller C M L and Parlett M (1974) *Up to the Mark: A study of the examination game,* Society for Research into Higher Education, London
A research study of students' strategies in coping with assessment. The authors identify three types of students - the cue seekers, the cue conscious and the cue deaf - and reveal some disturbing conclusions.

Nightingale P et al (1996) *Assessing Learning in Universities,* University of New South Wales Press
A wealth of detailed case studies from Australia organised under sections depending on the nature of what it is you are trying to assess.

Rowntree D (1987) *Assessing Students: How Shall We Know Them?* Harper & Row Ltd
This contains an analysis of the rationales of four different assessment methods and their effects on learning.

References for Win Hornby's Case Study

Bloom B S et al (1956) *Taxonomy of Educational Objectives I: The Cognitive Domain,* Longmans, London

Boud D (1986) *Implementing Student Self Assessment,* HERDSA Green Studies, Vol 5, University of New South Wales

Boud D and Falchikov N (1995) What Does Research Tell us about Self Assessment? in Boud D (ed) *Enhancing Learning through Self Assessment,* Kogan Page, London

Boud D and Lubin J (1983) *Self Assessment in Professional Education*, University of New South Wales

Bridges P et al (1999) Discipline-related Marking Behaviour Using Percentages: A Potential Cause of Inequity in *Assessment, Assessment and Evaluation in Higher Education,* 24 (3) 285-300

Brown S and Glasner A (eds) (1999) *Assessment Matters in Higher Education - Choosing and Using Diverse Approaches,* Open University Press, Buckingham

Fitzgerald T J, Gruppen L D, White C and Davis W K (1997) Medical Students' Self Assessment Abilities: Accuracy and Calibration, *Proceedings of the Meeting of the American Educational Research Association*, Chicago

Hornby W (2003) Assessing Using Grade Related Criteria: a Single Currency for Universities? *Assessment and Evaluation in Higher Education,* Vol 28, No 4, 435-454, 20

Larres P M, Ballantine J A and Whittington M (2003) Evaluating the validity of self-assessment: measuring computer literacy among entry-level undergraduates within accounting degree programmes at two UK universities, *Accounting Education*, 12 (2) 97-112

Mabe P A III and West S G (1983) Validity of Self Evaluation of Ability: A Review and Meta-Analysis, *Journal of Applied Psychology,* 67 (3) 280-296

Rowntree D (1989) *Assessing Students: How Shall We Know Them*? Kogan Page, London

Rust C, Price M and O'Donovan B (2003) Improving Students' Learning by Developing their Understanding of Assessment Criteria and Processes, *Assessment & Evaluation in Higher Education,* Vol 28, No 2, 147-164

Salmon G (2002) e-tivities: The Key to Active Online Learning, Kogan Page, London

Stephani L A J (1994) Self, peer and group assessment procedures, in Sneddon I and Kremer J (eds) *An Enterprising Curriculum: Teaching Innovations in Higher Education*, Kogan Page, London

Yorke M, Bridges P and Woolf H (2000) Mark distributions and marking practices in UK higher education: some challenging issues, *Active Learning in Higher Education,* Vol 1 (1) 7-27

Developing assessment methods at classroom, unit and university-wide levels

Professor Trudy W Banta, Indiana University-Purdue University Indianapolis

Definitions of outcomes assessment

In the early 1980s, the term assessment, or more accurately, outcomes assessment, was adopted in the United States to refer to information obtained from students, graduates and other stakeholders that may be used to improve academic programs and student services within universities. In many other countries, this process is called evaluation, or programme evaluation. The term assessment is preferred in the United States to distinguish the process designed to improve programmes and services from evaluation, a process designed to gauge the achievements of academic staff for purposes of awarding promotions, tenure and merit pay.

This author views outcomes assessment as a prudent step in a process that begins with planning what we wish to do. Plans are implemented and simultaneously appropriate data can be collected for use in assessing progress. If assessment findings are used to improve our processes, our plans may be adjusted and the cycle of planning, implementing, assessing, and improving begins anew. Assessment in this context may be defined as a process of providing credible evidence of the outcomes of higher education that is undertaken for the purpose of improving programs and services within an institution. A second, simpler definition focuses squarely on the paramount college outcome, student learning. Former vice president of the American Association for Higher Education, Theodore Marchese, calls assessment 'a rich conversation about student learning informed by data' (personal communication, 7 January 2004). This definition may provide the best context for the study of assessment currently underway in Scotland.

Assessment of individuals and groups

When academic staff hear the term assessment, they think most often in terms of assessing individual student development. They assess basic skills such as the ability to write, communicate orally, or use mathematics, for the purpose of advising students about appropriate placement in courses. They review student performance in their classes or modules using assignments, papers and projects. And as students complete some programmes, they are given comprehensive written and/or oral exams that test what they have learned throughout their years of study. Important outcomes of assessing individual student development include the following:

1   faculty can assign marks or grades to students

2   students learn about their own strengths and weaknesses so that they can correct them and improve their future performance, and

3   students acquire skills in self-assessment that they can use throughout their lives. Assessment of individual student development is a critically important component of the higher education experience.

For purposes of conducting outcomes assessment, we need a second look at aggregated student work in a class or module, in sections of the same class and even

across classes in a curriculum. Looking at student work collectively, we can tell where learning is satisfactory and where gaps in learning exist. We may also obtain some clues about which approaches to instruction produce the most learning for which students. These group assessment activities consist of classroom assignments, tests and projects - all the same sorts of measures that are used to assess individual student development. But with group assessment we can add a variety of other measures, such as questionnaires for students, graduates and employers. Interviews and focus groups yield helpful data. We can look at programme completion data to see how many students complete our courses and curricula and how long it takes them. We can look at the placement of students in further education or careers. By tracking our graduates, we can see how successful they are in postgraduate programmes or on the job and if they have received awards or recognition for their performance. Finally, we can use the results of group, or outcomes, assessment to improve our programmes and to demonstrate accountability to external stakeholders.

To summarise, assessment of individual student development can assist students in mastering content as well as in learning to assess their own strengths. Group, or outcomes, assessment can help faculty improve instruction and enable institutions to demonstrate their accountability.

Good assessment, or evaluation as many call it, embodies the same principles as does good research. In both we pose an important question, determine an appropriate approach to answering the question, collect data, analyse the findings and issue a report. Assessment goes a step farther in that the findings are utilised to improve instruction in individual classrooms as well as entire academic programmes and university-wide services.

Preparing academic staff to conduct assessment

Since most academic staff are not trained as teachers, faculty development is an important prerequisite for conducting good assessment. Faculty development can help instructors:

- write clear objectives for student learning in modules and curricula
- individualise instruction using a variety of methods and materials, and
- develop assessment tools that test higher order intellectual skills.

In determining appropriate approaches to assessment, it is very helpful to write goals and objectives for student learning using action verbs. For instance, if we want students to improve their writing skills, an appropriate assessment of their progress would be a written assignment. If we want them to develop skills in locating reliable information, we could give them a project incorporating the use of such skills in order to assess their internet search and analysis strategies.

Bloom's *Taxonomy of Educational Objectives* (Bloom, 1956) consists of six increasingly complex categories that describe what Bloom has called the cognitive domain. These extend from knowledge and comprehension at the lowest level of complexity through application, analysis, synthesis and evaluation. Action verbs may be associated with each of these levels of the domain. For instance, if we develop an

objective for students using a verb such as identify, define or describe, this learning objective is at the knowledge level. If we ask them to demonstrate, compute or solve, students will be performing at the application level. If we expect them to criticise, compare or conclude, the students will be developing skills at the evaluation level. In faculty development, discussing the use of verbs from the various levels of Bloom's taxonomy can be a helpful step in developing the ability to assess learning outcomes.

The use of action verbs in learning objectives may be illustrated more specifically as follows: If we ask a student in an English course to demonstrate how language influences intellectual and emotional responses, we are testing the student's application skills. Synthesis skills would be illustrated in the following objective: Synthesise diverse issues and responses raised in collaborative discussions of texts. Learning outcomes in science might include the following: Define and explain basic principles, concepts, and theories of science (knowledge level); solve theoretical and experimental problems in science (application level); and evaluate scientific arguments at a level encountered by informed citizens (evaluation level).

A matrix can be useful in a number of ways in promoting conceptual thinking about assessment. A matrix format with six columns that has been used successfully at many colleges and universities in the United States is one that has as a heading for the first column of 'What general outcome are you seeking (eg critical thinking)?' The second column is headed 'How would you know it (the outcome) if you saw it - that is, what would the student know or be able to do?' The third column heading is 'How will you help students learn the concept, in class or out of class?' And the fourth heading is 'How could you measure each of the desired behaviours listed in column 2?' The fifth column heading reads 'What are the assessment findings?' And the sixth asks 'What improvements are or might be based on assessment findings?' Completing such a matrix can enable faculty to explain to students and other stakeholders:

1   specific learning outcomes of a module or a course of study
2   collective student outcomes, and
3   actions undertaken to improve student learning based on assessment findings.

Classroom, unit and university-wide levels of assessment

Outcomes assessment occurs at a number of levels. It begins with the individual student in a classroom. Aggregating the work of all students in a classroom will provide information to inform classroom assessment. Aggregating student work across various classes or modules can provide assessment (evaluation) of the impact on learning of an entire course of study. Looking at student products across the disciplines in a college provides assessment at that level. Assessment findings from various academic units within a university can provide a measure of institutional effectiveness that can be used to demonstrate accountability at the state, regional or national level.

A distinction must be drawn between direct and indirect measures of student learning. Direct measures are those assignments, exams, projects and papers that enable us to see what students actually know and can do. Indirect measures include questionnaires, interviews, and focus groups that enable us to assess the process of learning or other aspects of the student experience. Direct measures of learning are critical if we are to

assess acquisition of knowledge and skills. But no test score will tell us why certain components of students' knowledge are strong or weak. Thus indirect measures are needed to help us understand why weaknesses are occurring and what might be done to address them. Good assessment includes both direct and indirect measures.

Citing some examples of assessment at various levels may add clarity to this concept. Fast feedback, or classroom assessment, can be used at the individual classroom level. Students are asked during the last five minutes of a classroom session to state the most important thing they learned in the class that day and to tell the instructor what is still unclear. Then they may be asked about the helpfulness of the advance reading assignments for the day's work. Finally, they may be asked for suggestions for improving the class and/or the assignments. In an illustration from the Graduate School of Business at the University of Chicago, students responded to the last question in that sequence by suggesting the following improvements:

● install a portable microphone

● increase the type size on transparencies

● leave lights on when using a projector

● don't cover the assigned reading in great detail but instead

● provide more examples from actual practice in class lectures and discussion (Bateman and Roberts, 1993).

We can adapt the typical course evaluation to include questions about the student experience. Are students encountering in the course principles of good practice in undergraduate education (Chickering and Gamson, 1987)? We might ask, for instance, if in a given module or in an entire curriculum:

● learners held high expectations for one another

● learners interacted frequently with academic staff in and outside class

● learners participated in learning teams

● learners respected diverse talents and ways of learning (Cournoyer, 2001).

Primary trait scoring

Primary trait scoring is an assessment method that can be used in both direct and indirect measures, and at all levels (Walvoord and Anderson, 1998). Instructors identify the traits or attributes that are necessary for success in an assignment, then compose a scale or rubric that gives clear definition to each point, and finally evaluate student work according to the rubric. For example, a project that involves developing and presenting a research paper encompasses at least the following primary traits:

● an appropriately narrow topic or purpose

● a bibliography

● an outline

● a first draft

● a final draft, and

● an oral defence.

For each of the traits of this assignment we might develop a three-point rubric, defining each point carefully and explicitly. The bibliography, for instance, might be assessed as follows.

3   (Outstanding) References current, appropriately cited, representative and relevant.

2   (Acceptable) References mostly current, few citation errors, coverage adequate, mostly relevant.

1   (Unacceptable) No references or containing many errors in citation format, inadequate coverage, or irrelevant.

If one creates a matrix containing the primary traits of an assignment as row titles and the levels of each rubric as column headings, such a matrix can serve three purposes. First, it can be shared with students prior to an assignment so that they will understand the criteria being used to judge their work. Second, it can be completed for each student on the basis of the work submitted and thus provide detailed feedback when returned to the student. Third, if the instructor places a check mark in the appropriate box of the matrix for every mark assigned in evaluating the work of all students, the matrix can indicate to the instructor where there are weaknesses in student learning and suggest what changes may need to be made to enable every student to reach the desired learning outcomes.

Another matrix might list principal outcomes as row titles and courses in a curriculum as column headings. Placing check marks in the matrix to demonstrate which outcomes each course addresses will help students understand where they will learn specified knowledge and will assist instructors in spotting gaps in the curriculum.

Primary trait scoring can be used in virtually any field. For instance, at Ball State University in Indiana, sophomore competence in mathematics was tested. Students were asked to turn in their supporting work in connection with their item responses on a math test. Then instructors used a four-point scale to score responses in terms of conceptual understanding, consistent notation, logical formulation and completeness of the solution (Emert and Parish, 1996).

At North Dakota State University, staff in sociology and anthropology developed scenarios appropriate to the discipline, then asked graduating students to respond to the scenarios in groups (Murphy and Gerst, 1997). A staff facilitator asked questions related to outcomes staff had identified in three areas - concepts, theory and methods. Then two staff observing the group work used a 0-3 scale to rate each student on each question. Looking at aggregate scores across all student groups enabled staff working together to ascertain strengths and weaknesses of their curriculum.

Group interaction also can be assessed using primary traits and scoring rubrics. Faculty at the Purdue University College of Pharmacy in Indiana developed a five-point scale ranging from 5 = consistently excellent to 1 = inconsistent and/or inappropriate to judge the performance of students working in groups (Chalmers and Mason, 1994). The characteristics faculty were observing included the following:

- listened to others

- actively contributed to discussion

- challenged others effectively

- was willing to alter own opinion
- effectively explained concepts/insights
- summarised proposed solutions.

Involving stakeholders in assessment

In a comprehensive assessment programme, it is important to involve as many stakeholders as possible. Such groups include students, staff, student affairs professionals, administrators, graduates and employers.

An example of involving students may be drawn from the experience of the Department of Psychology at Montevallo University (Judith Rogers, personal communication, 4 May 1994). Students were asked to form an advisory council and to provide continuous assessment of the psychology major. Students made a number of important recommendations for improving their programme. They asked that a statement of expected ethical behaviors for students be drawn up and volunteered to do this themselves. They suggested that a second research course with a laboratory be added to the curriculum because they felt they needed more research experience. They asked for more comparative psychology; additional terminals for the statistics laboratory; and more opportunities in all their courses to write, make oral presentations and conduct research projects.

At Harvard University, Richard Light (1992) has involved students in teams to evaluate both courses and programmes. In addition to the good information about the strengths and weaknesses of courses and curricula that the students provide to academic staff, the students experience learning benefits as they engage in the assessment process. As they interact with their peers, they strengthen their communication skills and develop competence in working in a team. Their self-confidence increases and they improve their understanding of others' perspectives. Finally, student involvement in learning increases.

At Indiana University-Purdue University Indianapolis, staff across the institution have agreed on student outcomes in six areas, including:

1   core communication and quantitative skills

2   critical thinking

3   integration and application of knowledge

4   intellectual depth, breadth, and adaptiveness

5   understanding society and culture

6   values and ethics.

Efforts are underway to develop an electronic portfolio that will give students the responsibility for demonstrating their skills in connection with each of the six principles of undergraduate learning. The electronic format will give students the opportunity to use multiple media to demonstrate unique individual skills and achievements. Written reflections on the material they put into the portfolio will enable students to enhance their metacognitive skills.

Involving student services professionals with staff in the assessment of learning is illustrated in an experience at Virginia Commonwealth University (Fuhrmann, 1995). There first-year students in English composition wrote a theme each week for 15 weeks. The subject of each theme was the student's experience with, and opinions of, some aspect of campus life, such as the admissions process, the new-student orientation programme, or academic advising. Teams of staff and student affairs professionals read and evaluated the essays. Staff from the Department of English assessed the student work in terms of the quality of the writing, while student affairs professionals gained valuable information about students' perceptions of the quality of various campus programmes and services. As a result of this team approach to assessment, a number of changes have been made at Virginia Commonwealth University, including provision of a new central advising centre, a revised advising handbook and enhanced multicultural workshops.

Kretovics and McCambridge (1999) at Colorado State University provide an example of involving employers in assessment. They have used a combination of surveys and focus groups for employers of business graduates to identify skills, knowledge, and personality attributes that are sought by employers. The findings yielded by these assessment techniques have encouraged staff to make curriculum changes; motivated students to develop needed skills; and strengthened ties among staff, students and employers. Staff have added a credit to the business communications course that increases time for students to work in teams and make more oral presentations. In addition, business ethics and social responsibility are now discussed in introductory courses and a new Introduction to Business course emphasises career decision-making.

Using assessment findings

The time and energy devoted to outcomes assessment can only be justified if the findings are used to improve student learning and development. A few examples illustrate such uses. At Columbia College, academic staff in social work wanted students to apply critical thinking skills in their clinical practice settings (Baskind, Shank, and Ferraro, 2001). More specifically, staff specified that students would use statistical analysis to inform practice decisions. Laboratory exercises were used to assess the skill and staff set as a standard that 80 per cent of all students would earn a score of at least 3.0 on a four-point scale on their laboratory exercises. Columbia College staff found that only 64 per cent of the social work students scored 3.0 or better. By asking students what would help them most, staff learned that students needed more time in the laboratory - more time to practice their skills. In response, staff made the learning laboratory available to students five days a week.

At Southern Illinois University-Edwardsville, for more than a decade staff have required every student to undertake and pass a 'senior assignment' in their discipline (Eder, 2001). In business, the senior assignment is a case study analysis. In education, students must complete a professional portfolio. In psychology, students must develop a poster presentation based on a research project. In engineering, there is a senior design project and in nursing, students must develop a plan of care for a patient. At the end of each academic year, staff meet to go over the senior assignments for that year and draw conclusions about strengths and weaknesses in students' knowledge and skills. Over the years, each discipline has made changes

based on the evidence provided in the student work. In business, more case studies and research projects have been required at earlier levels of the curriculum. In education, more practice in classroom management has been offered to improve student learning. In psychology, the statistics requirement has been changed. In engineering, students have been given more opportunities to practice their writing and speaking skills throughout the curriculum. And in nursing, an expensive simulation laboratory with computerised patients has been purchased so that students may practice their clinical and diagnostic skills without harming a patient!

At Eastern New Mexico University, portfolio review and/or an audition is conducted for every fine arts student every semester by a panel that includes staff, students, community representatives and professional staff or staff outside fine arts (Alec Testa, personal communication, 27 April 2004). As a result of this intensive assessment process, changes have been made in instruction and programming and now it is possible to see some results of those changes. Staff have observed that students' creativity, conceptualisation and technique have improved.

Characteristics of effective assessment

Over the years since 1980, assessment practitioners have developed a number of characteristics that define effective assessment (Palomba and Banta, 1999). First, there should be an overall plan for assessment, suggesting where and when students will be evaluated, the evaluation instruments that will be used, and what will be done to respond to the data collected. As indicated earlier, all stakeholders in higher education should be drawn into the assessment process. For instance, students and employers, as well as staff, can help to set goals and objectives for courses and curricula, select or design assessment methods, and even collect some of the data. Assessment findings then should be reported to all stakeholders and their use encouraged. Finally, the assessment programme itself should be evaluated periodically, preferably through peer review.

Successful assessment initiatives are led by committed individuals and include collaboration between staff and student affairs leaders. Assessment depends upon effective teamwork in planning, implementation and utilisation of findings. Assessment thrives in a supportive campus climate where faculty and administrators demonstrate their concern for students and their commitment to continuous improvement. Once assessment results are reported, campus leaders must follow up periodically to make sure that the findings are being used to improve practice.

If the experience of the last quarter-century is any guide, the need for providing evidence of accountability in higher education will not diminish, but rather will increase, as it has virtually every year since 1980. As scholarship reveals how assessment can improve instruction and student learning, more staff will realise its benefits. Additional electronic assessment methods will be developed - already this is occurring at a rapid pace. More sharing of assessment methods will take place among groups of staff. And ultimately, assessment will enable staff to gain a deeper understanding of student learning and student learning will improve as assessment findings are used to effect continuous improvement of the environment for learning.

References

Baskind F R, Shank B W and Ferraro E K (2001) Accountability for Professional Practice: Assessment in Social Work Education, in Palomba C A and Banta T W (eds) A*ssessing Student Competence in Accredited Disciplines,* Stylus Publishing, Sterling, VA, 95-120

Bateman G R and Roberts H V (1993) *TQM for Professors and Students,* Chicago Graduate School of Business, University of Chicago, ERIC Document 384-319

Bloom B S (ed) (1956) *Taxonomy of Educational Objectives: The Classification of Educational Goals, Handbook I Cognitive Domain,* Longmans, Green, New York

Chalmers R and Mason H (1994) *Assessing Strategies for Enhancing Ability-Based Learning in Pharmacy Schools*, presented at the Assessment Conference in Indianapolis, 8 November 1994

Chickering A W and Gamson Z F (1987) Seven Principles for Good Practice in Undergraduate Education, *AAHE Bulletin,* 39 (7) 3-7

Cournoyer B R (2001) Assessment of Student Learning in Social Work Education: The Indiana Model, *Advances in Social Work,* 2 (2) 128-151

Eder D J (2001) Accredited Programs and Authentic Assessment, in Palomba C A and Banta T W (eds) *Assessing Student Competence in Accredited Disciplines,* Stylus Publishing, Sterling, VA 95-120

Emert J W and Parish C R (1996) Assessing Concept Attainment in Undergraduate Core Courses in Mathematics, in Banta T W, Lund J P, Black K E and Oblander F W *Assessment in Practice: Putting Principles to Work on College Campuses*, pp104-107, Jossey-Bass, San Francisco

Fuhrmann B (1995) The "Prompts" Project Prompts Academic and Student Affairs Collaboration, *Assessment Update*, 7 (6) 10

Indiana University-Purdue University Indianapolis (IUPUI) *IUPUI Principles of Undergraduate Learning*
www.iport.iupui.edu/teach/teach_pub.htm

Kretovics M A and McCambridge Jr (1999) A Seven-Step Approach to Developing an Outcomes Assessment Program, *Assessment Update,* 11 (2) 10-11

Light R J (1992) *Harvard Assessment Seminars: Explorations with Students and Faculty about Teaching, Learning, and Student Life: Second Report*, Harvard University, Graduate School of Education, Cambridge, MA

Murphy P D and Gerst J (1997) Beyond Grades and "Satisfactory" in Assessing Graduate Student Learning, *Assessment Update,* 9 (3) 12-13

Palomba C A and Banta T W (1999) *Assessment Essentials: Planning, Implementing and Improving Assessment in Higher Education,* Jossey-Bass, San Francisco

Walvoord B E and Anderson V J (1998) *Effective Grading: A Tool for Learning and Assessment*, Jossey-Bass, San Francisco

Developing a variety of assessment methods

Dr Chris Rust, Oxford Centre for Staff and Learning Development, Oxford Brookes University

Validity - an essential foundation

Before we even start to consider variety, and what that might mean or look like, there are certain basics that need to be got right first. And probably the most important and fundamental issue that needs to be got right is the question of validity. Are we assessing what we believe or claim to be assessing? And over the past decade there has been quite a lot of criticism that traditional assessment practices in higher education have not been valid. In the United States, Tom Angelo claimed in 1996 that:

> We continue to assess student learning - and to graduate and certify students much as we did in 1986, 1966, or 1946, without meaningful reference to what students should demonstrably know and be able to do (Angelo, 1996).

A little earlier, in the UK, Graham Gibbs argued that:

> Assessment systems dominate what students are oriented towards in their learning. Even when lecturers say that they want students to be creative and thoughtful, students often recognise that what is really necessary, or at least what is sufficient, is to memorise (Gibbs, 1992a).

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

In order to address this fundamental issue and ensure validity, we need to ensure that we have what John Biggs has termed a 'constructively aligned' teaching system (Biggs, 1999). The fundamental principle of constructive alignment is that a good teaching system aligns teaching method and assessment to the learning activities stated in the objectives so that all aspects of this system are in accord in supporting appropriate student learning.

In order to achieve constructive alignment, and hence valid assessment, Biggs proposes a three stage model of course design, namely establish:

1   What are 'desired' outcomes?

2   What teaching methods require students to behave in ways that are likely to achieve those outcomes?

3   What assessment tasks will tell us if the actual outcomes match those that are intended or desired?

So the fundamental message is that, whatever else we do regarding assessment, we must have clear explicit learning outcomes and whatever assessment method we choose must validly and transparently assess whether those outcomes have been met or not.

Why variety?

The title of this paper implicitly suggests that we should want to develop a variety of assessment methods and that this therefore is a good thing but why should this be? There are arguably a number of reasons.

Constructive alignment and validity

As has already been argued above, in order to achieve constructive alignment and to validly assess all, of the course outcomes we may need to increase the range of assessment methods used.

Traditional assessment only assesses a narrow range of skills

Linked to the argument about validity, it is arguable that traditional assessment only assesses a fairly narrow range of skills and with the current expectations to develop students as lifelong learners with a range of transferable as well as disciplinary skills it is increasingly unlikely that a narrow range of assessments can validly assess a wide range of outcomes.

Fairness

It can be argued that every form of assessment will place some students at a disadvantage to some extent. Exams may put students who suffer from exam nerves at a disadvantage, oral presentations may disadvantage the shy, projects may disadvantage those who are bad at time management, etc. A variety of assessment methods offers all students disadvantaged under one the opportunity to possibly excel in the others.

Logic of outcomes-based course design

The move from content-based to outcomes-based course design means that it is no longer sensible or logical (if it ever was) to see assessment in terms of measuring student achievement by grades and numbers (and this clearly links with the arguments for constructive alignment already discussed).

Interest and motivation (and flexible learning)

The links between interest and motivation and learning are well established in the research literature. As is the fact that choice and variety are two factors that can increase interest and motivation, and this is especially true when it comes to assessment. Choice may also assist in the provision of more flexible learning and modes of learning (part-time, distance, e-learning etc).

Be strategic - economical and efficient

With the increase in student numbers and decline in unit resource there has been an increasing imperative to be more economical and effective in everything we do, and assessment can be especially expensive in terms of staff time. But we also know that many students are 'assessment driven'.

> Assessment defines what students regard as important, how they spend their time and how they come to see themselves as students and then as graduates...If you want to change student learning then change the methods of assessment (Brown G et al, 1997).

> To use assessment strategically, to be effective and efficient and to positively influence student learning, may well require changes in assessment practice (see Gibbs, 1992b; Rust, 2001).

Be strategic - student learning and skills development (and retention)

With ever widening participation, support for student learning and skills development (and trying to improve retention) has become even more important and assessment clearly has a crucial role to play in this, in a number of ways.

1   Clarify expectations/keep the students busy

> Setting students assignments as soon as soon as they arrive at university could help cut dropout rates…integrate students into university life as quickly as possible. This involves making them aware of the quality and quantity of work expected from them…The freedom…is too much for some and they probably need more structure in the first year (Higher, 2003).

This assertion was based on conclusions from the Student Transition and Retention project which examined practice in five UK universities.

2   Pace student learning

Rather than having the only assessment at the end of a module, which encourages students to leave doing any real work until the last minute, which may have disastrous outcomes for weaker students, assessment that is staged through the module can help the students to pace their learning. Fears that this may increase the assessment workload for staff can be addressed through the strategic use of self and peer-assessment, and also computer-aided assessment, which once set up may require little or even no staff time (see below and Rust, 2001).

3   Skills development

The literature on skills development is clear that 'bolt-on' study skills courses are not very effective because students are not motivated and fail to see the relevance or to transfer what is learnt. Far more effective is to integrate the skills you want to develop in the curriculum, and especially in the various assessment tasks. So a sensible strategy is to audit the assessment tasks across the programme and to ensure that all the skills you want your students to develop are being highlighted and assessed somewhere. And if not, you may need to consider changing various tasks to make sure that they do.

4   Positively reinforce/allow for early failure

If students have low self-belief (and this may be more likely for widening participation students) it can adversely affect their achievement leading to drop-out (Mantz Yorke, keynote presentation at ILT symposium, 27 September 2001, based on study of six institutions). Such students believe failure is due to lack of intelligence. This leads to a kind of fatalistic 'learned helplessness' with the belief that there is nothing that they can do about it. 'There is no point in me trying because I am just not clever enough,'

so difficult tasks prompt the student to give up. Such students are also likely to be overly concerned with 'saving face', and rather than suffer repeated humiliation they will quickly drop-out. So what can be done about this? Well there are some possible assessment solutions, some of which have already been identified, above. The curriculum, and in particular the assessments, can be adjusted and designed to foster skills development, and sub-goals can be set for longer assignments. Consideration should also be given to allowing for 'slow learning' in the programme's assessment strategy. The University of Luton's recent decision to reduce the number of modules passed in the first year necessary to progress to year two is a courageous example of this. You might also consider making the first semester assessments (or even the first year's) primarily formative; for example, at Massachusets Institute of Technology in the US all first year assessments are simply pass/fail. There is interesting research data (Black and Wiliam, 1998) that where assessment does not include grades, only qualitative feedback subsequent student work improves significantly.

5   Timely and effective feedback

There is also considerable research evidence that potentially the most important part of the assessment process, with regard to supporting learning, is feedback. A recent analysis of the literature (Gibbs and Simpson, 2002) identified 11 conditions in which assessment supported learning, and seven of these concerned feedback. But there is also considerable evidence that all too often, in practice, feedback is not timely or effective (eg Hounsell, 1987; Lea and Street, 1998; Maclellan, 2001). In thinking about variety in assessment, important consideration should therefore be given to the process of feedback.

Diversifying assessment - including the students in the process

There is no shortage of literature on strategies for diversifying assessment (see Brown S et al, 1994; Rust, 2001), and these approaches can be grouped under at least five headings:

1   change the criteria

2   change the task

3   mechanise assessment

4   assess groups

5   involve the students.

If we simply consider the last of these, involving the students, there is a range of possibilities, from marking exercises to peer-marking which can both improve the work of the students and in some cases reduce staff workload at the same time.

Marking exercises

One example of these is an intervention centred around an assessment workshop, intended to increase the students' understanding of the assessment criteria and what was being expected of them, and thus improve the quality of the work they produced.

The exact detail of the process was as follows.

a   A week before the workshop all students on the module were provided with two sample assignments (one excellent piece of work and one a borderline pass) and

marksheets including assessment criteria and grade definitions (an assessment grid). Students were asked to individually complete a marksheet providing a grade, marks and rationale and feedback for each of the assignments before coming to the workshops.

b   Workshops (90 minutes long) were held for all students in groups of 40. These were held within the module's time frame but were in addition to the weekly lecture and seminar. The workshops were structured in the following way.

    i   Small group discussion of initial marking of sample work.

    ii   Feedback of small groups' agreed grades and rationale to plenary.

    iii   Tutor-led comparison of provided rationales with criteria.

    iv   Tutor explanation of each criterion.

    v   Small groups review their initial assessment and grade.

    vi   Final small group report to plenary of grade for each piece of work.

    vii   Tutor provides tutor-marked and annotated versions of samples, and discusses reasons for the assessment and mark given.

The small group discussions allowed the students to compare and justify their initial assessment of the work against that of others as well as allowing the declared grade to be the responsibility of the small group. However, the students were asked explicitly not to change their initial grading on their individual sheets.

c   Three weeks later, students submitted their coursework along with a completed self-assessment sheet.

The same feedback sheet was used for the sample assignments and for the self-assessment. It incorporated comments, an assessment grid, a grade and a mark.

The published results of this intervention (Rust et al, 2003), repeated and replicated over two years, are that the participants at the assessment workshop subsequently achieved statistically significant better results in their assessed coursework. While in comparison, the performance of the participants and non-participants on a module prior to the intervention showed no significant difference in the performance of the two groups. Furthermore, one year later, the participants of the first cohort were still achieving significantly better results than the non-participants. And since publication, these outcomes have been replicated again for a further year's cohort.

Self-assessment

Some tutors may be put off the idea of self-assessment thinking it means allowing students to give themselves marks. While it could involve marks it certainly need not do, and far and away the most important aspect of self-assessment is in helping the student to think about the quality of their own work. And this might be as simple as providing a short list of headings which the students is expected to respond to, and submit with the completed piece of work, for example:

- strengths of this piece of work

- weaknesses in this piece of work

- how this work could be improved

- the grade it deserves is…

- what I would like your comments on (Brown S et al, 1994).

If this simply makes the students read their work through again it may enable them to spot and correct errors and omissions, and it can also enable the tutor to give more focused feedback. But as with the marking exercise described above, there is evidence that training students to assess their own work results, over time, in them producing significantly better work (eg McDonald and Boud, 2003).

Peer-marking using model answers (Rust, 2001)

This may not be possible in all disciplines but if it is possible to have model answers for work that is set this strategy can have a number of varied benefits.

One example is an engineering course where students who did badly in the exam were known to be failing because they needed more practice in working through numerical problems. There was no way, however, that the staff could contemplate more marking and if it wasn't assessed how could they get the students to seriously undertake the practice problems?

The solution they adopted was to set the students problems on a regular basis to solve in their own time, and to allow 20 minutes at the start of certain lectures for these to be marked. This is done by rows of students swapping work; the lecturer then leads them through a model solution. Although 170+ students are involved they have become increasingly efficient at doing this and needed less time - only seven minutes on one occasion. It was made a course requirement that all students had to have attempted 50 problems over the term (in order to be eligible to sit the end of term exam) but marks did not count. It would therefore have been possible for students to only make a perfunctory attempt at each problem and get them all wrong and still be able to sit the exam. But this did not happen. The students did take the problems seriously, presumably because they did not know which of their classmates would be marking their work and they did not want to be 'shown up'. And the performance in the exam has improved staggeringly.

This is because not only are they getting increased practice in undertaking problems, and the associated feedback, but in addition they are benefiting from:

- seeing the preferred solution, with the weighting of an examination marking scheme explained

- seeing the variety of approaches taken by the peers, and

- having to judge the degree to which the work of their peers does or doesn't meet the requirements of the marking scheme.

Peer feedback (Rust, 2001)

Even if model answers are not possible, students can also greatly benefit from getting involved in reading and commenting on the work of others. Students should therefore be encouraged to get into the habit of getting informal feedback from each other, and if instituted formally this may be a very effective way of ensuring the

students get feedback which the tutor has not got time to give. It can also have very definite educational benefits.

An example of this is a course in geography which originally required the students to write two essays, one towards the beginning of the course and one near the end. The tutor became increasingly despondent that despite all the efforts spent marking and writing comments on the first essay invariably there was little improvement in the second and as student numbers were rapidly rising it was becoming increasingly difficult to find the time to maintain the quality of the feedback given.

Instead of two essays, the course now requires only one, which is written in two stages. In the first stage, the students write a first draft of their essay by a given date. In a seminar session, they are then paired up and each pair reads and gives detailed feedback on the other's piece of work. In the light of this feedback, the students then redraft their essays. When the essay is finally submitted it is accompanied by an account of how the feedback has been used, eg 'I've included more sources because the first draft was criticised for using only two. I've kept the introduction the same even though it was criticised as unclear because I don't agree' etc.

As a result of this change the tutor has halved his marking load. It is true that only one topic is now assessed rather than two as before but the tutor believes that educationally this is more than compensated for by the following benefits.

- It develops the students critical faculties.
- Significantly better work is produced.
- It is more like the 'real world' - good writing involves redrafting in the light of criticism.

So why don't we?

Nothing in this paper is particularly knew, but nevertheless it was felt to be a topic worth focussing on. Which begs the question, why have not more colleagues been already persuaded by the arguments above and implemented more of these changes? I would suggest the following list of possible answers:

- not convinced (personally)
- institution/colleagues/students not convinced
- tradition/inertia/no incentive to change
- time and/or other costs
- too difficult (need help/training)
- not allowed (regulations/procedures).

If we are serious about improving student learning, and are not to ignore what the evidence from the literature suggest about assessment practice, we surely should consider seriously what are the blocks to change and how are we to overcome them.

References

Angelo T (1996) Transforming assessment: high standards for higher learning, *AAHE Bulletin*, 3-4 April

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

Black P and Wiliam D (1998) Assessment and classroom learning, *Assessment in Education*, 5 (1) 7-74

Brown G, Bull J and Pendlebury M (1997) *Assessing student learning in higher education*, Routledge, London

Brown S, Rust C and Gibbs G (1994) *Strategies for diversifying assessment in higher education*, Oxford Centre for Staff Development, Oxford

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

Gibbs G (1992a) Improving the quality of student learning, *The Times Educational Supplement*

Gibbs G (1992b) *Teaching more students 4: Assessing more students,* Oxford Centre for Staff and Learning Development, Oxford

Gibbs G and Simpson C (2002) *Does your assessment support your students' learning?* www.brookes.ac.uk/services/ocsd/1_ocsld/lunchtime_gibbs.html

Hounsell D (1987) Essay writing and the quality of feedback, in Richardson J T E, Eysenck M W and Warren-Piper D (eds) *Student Learning: Research in Education and Cognitive Psychology*, SRHE/Open University, Milton Keynes

Lea M and Street B (1998) Student Writing in Higher Education: an academic literacies approach, *Studies in Higher Education,* 2 (2) 157-172

McDonald B and Boud D (2003) The impact of self-assessment on achievement: the effects of self-assessment training on performance in external examinations, *Assessment in Education: Principles, Policy and Practice,* 10 (2) 209-220

Maclellen E (2001) Assessment for learning: the different perceptions of tutors and students, *Assessment and Evaluation in Higher Education*, 26 (4) 307-31

Rust C (2001) *A Briefing on the Assessment of Large Groups*, LTSN Generic Centre Assessment Series, No 12, LTSN, York www.heacademy.ac.uk/resources.asp?process=full_record&section=generic&id=12

Rust C (2002) The impact of assessment on student learning, *Active learning in higher education*, 3 (2) 145-158

Rust C, Price M and O'Donovan B (2003) Improving students' learning by developing their understanding of assessment criteria and processes, *Assessment and Evaluation in Higher Education,* 28 (2) 147-164

Using assessment to improve learning: the BEAR assessment system

Professor Mark Wilson and Kathleen Scalise, Graduate School of Education, University of California, Berkeley

Abstract

This paper discusses how assessment practices in higher education can improve or hinder learning. Elements of effective learning environments that may better address underlying metacognitive issues are discussed. The principles of the Berkeley Evaluation & Assessment Research (BEAR) Assessment system are introduced, and their use to improve learning is described in the context of the University of California, Berkeley ChemQuery project.

Introduction

One of the surpassing mysteries in higher education is how, every semester and all around the country, substantial numbers of students come into class with all the right prerequisites and grades to prepare them to handle their new coursework, but in fact they seldom know what they are supposed to know. Why don't they know it? And furthermore, what does the instructor, especially in large lecture classes where the teaching load is already substantial, do about it?

In *From Naïve to Knowledgeable,* Joseph Hesse (1989), an instructor and conceptual change investigator in the sciences, said that a usual explanation is a 'pass-the-buck' interpretation - somehow the student just didn't study enough, didn't remember enough, wasn't interested enough. Hesse illustrated this by quoting a colleague 'who stated that, in his opinion, 90 per cent of student mistakes could be attributed to a lack of study on their part. Blame the student! Period! End of discussion'.

Assumptions of lack of studying or insufficient engagement with the material are common explanations of student underperformance. Also, views of fixed intelligence are common (Dweck and Leggett, 1988), in which instructors and even students themselves sometimes take the view that at some point they have 'topped out' in their ability to master the material. Throw in the complications of concept retention and knowledge transference, and it is perhaps too easy to justify the existence of under prepared students, and to support 'natural' filtering mechanisms that eliminate students through attrition or failing grades.

However, this premise fails to address a set of important issues that we will take on in this paper: Whether the students really did know the material they were responsible for in the first place; how we know they knew it; and whether sound metacognitive principles are in place for instructors and students to monitor and improve student learning processes, optimising their ability to construct, learn, retain and transfer knowledge.

In other words, is the problem really low ability or disengaged students, or are educational practices contributing to under preparation and underperformance?

Current state of formative assessment practices

To illustrate what some problematic practices might be, consider the role of formative assessment and feedback, as outlined by the recent National Research Council report, *Knowing What Students Know* (Pellegrino, Chudowsky and Glaser, 2001):

> [A] major law of skill acquisition involves *knowledge of results.* Individuals acquire a skill much more rapidly if they receive feedback about the correctness of what they have done. If incorrect, they need to know the nature of their mistake. It was demonstrated long ago that practice without feedback produces little learning (Thorndike, 1931). One of the persistent dilemmas in education is that students spend time practicing incorrect skills with little or no feedback. Furthermore, the feedback they receive is often neither timely nor informative. For the less capable student, unguided practice can be practice in doing tasks incorrectly.

The use of homework, laboratories, papers, quizzes and other activities by which students practice what they have learned is commonplace in education. Often some form of credit is given for the work and feedback is offered to students, sometimes in the form of a grade, at other times with more extensive critiquing. Especially in large lecture courses, more detailed feedback is often limited by resource constraints.

A major literature survey of over 250 sources on formative assessment (Black and Wiliam, 1998) found that effective assessment practices can play a powerful role in the learning experience, moving an average student, for instance, to the top third of the class, but only if certain conditions are satisfied. Student tasks needed to be aligned, or on target, with learning goals, and students need to receive meaningful and timely feedback on their performance, as well as targeted follow-up work. To regulate learning effectively, students need to understand three things:

a   the measures on which they will be judged

b   where they stand on these measures, and

c   how they can improve (Black and Wiliam, 1998).

Elements of effective learning environments

Clearly, a single summative score in the form of a grade can do little to inform mastery of complex material. Add to this the confounding effect of incorporating effort into this single grade, often on a basis that is not clearly defined for students, and one can see that the metacognitive 'signal' by which students 'tune' their performance has been weakened to the point of failure. The goal should be 'rigorous and wise diagnostic information' (Wolf, Bixby, Glenn and Gardner, 1991), but this is seldom made available.

Approaches exist that might facilitate metacognition in large lecture classes. An 'embedded assessment' system designed and used in assessment development at the University of California, Berkeley, called the BEAR assessment system (Wilson and Sloane, 2000) is described in the following section. It consists of easy-to-use tools for generating solid diagnostic information and feedback, perhaps especially useful in large class settings. The BEAR assessment system is a comprehensive, integrated system for assessing, interpreting, monitoring, and responding to student performance. It provides a set of tools for instructors and students to:

● reliably assess performance on central concepts and skills in curriculum

● set standards of performance

● validly track progress over the year on central concepts, and

● provide mechanisms for feedback and follow-up.

A note about embedded assessment

The term **embedded assessment** means just what it says: activities are 'embedded' or become part of, class learning activities. Instructors do embedded assessment all the time: a homework assignment, a laboratory procedure, a classroom discussion, an essay. Any of these and many more can be considered embedded assessment activities if a student produces something that can be rated, or observed and assessed

in some manner. The difference between these examples and what we discuss here as more formal embedded assessment is that the latter calls for attention to task design and formal 'calibration' of assessment tasks in relationship to a framework that describes the learning to take place.

The assessment triangle and the BEAR approach

Three broad elements on which every assessment should rest are described by the assessment triangle (Pellegrino et al, 2001), shown in Figure 1.

observations   **interpretation**
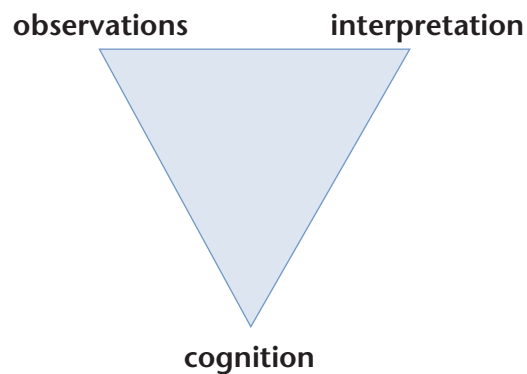
cognition

Figure 1 The *Knowing What Students Know* assessment triangle

According to the Committee Report, an effective assessment design requires:

- a model of student cognition and learning in the field of study

- well-designed and tested assessment questions and tasks, often called items

- and ways to make inferences about student competence for the particular context of use.

These elements are of course inextricably linked. Models of student learning should specify the most important aspects of student achievement to assess and they provide clues about the types of tasks that will elicit evidence and the types of inferences that can relate observations back to learning models and ideas of cognition. To serve as quality evidence, items themselves need to be systematically developed with both the learning model and subsequent inferences in mind, and they need to be tried out and the results of the trials systematically examined. Finally, the inferences provide the 'why' of it all - if we don't know what we want to do with the assessment information, then we can't figure out what the student model or the items should be. Of course, context determines many specifics of the assessment.

The BEAR assessment system is based on the idea that good assessment addresses these considerations through four principles:

1   developmental perspective

2   a match between instruction and assessment

3   the generating of quality evidence, and

4   management by instructors to allow appropriate feedback, feed forward and follow-up.

See Wilson (in press) for a detailed account of an instrument development process that works through these steps. Below we take up each of these issues in turn.

Principle 1: Developmental perspective

A developmental perspective of student learning means assessing the development of student understanding of particular concepts and skills over time, as opposed to, for instance, making a single measurement at some final or supposedly significant time point. One strategy to address the developmental perspective is specifying a set of 'progress variables' (Masters, Adams and Wilson, 1990; Wilson, 1990). These variables define the most important student growth goals of the curriculum and change from course to course as different areas of knowledge are the focus of interest and thus assessment.

With a progress variable approach, every instructional unit is seen as contributing to student progress on at least one of these variables and every assessment is closely aligned with one or more variables. This alignment allows the creation of a calibrated and meaningful scale to map the growth of students, so that instructors can track the progress of individual students and groups of students as they engage in learning.

Level of success	Big ideas	Descriptions of level	Item exemplars
10-12 Construction Why composition, structure, properties and amounts? (Using models)	The composition, structure and properties of matter are explained by varying strengths of interactions between particles (electrons, nuclei, atoms, ions, molecules) and by the motions of these particles.	Students are able to reason using normative models of chemistry, and use these models to explain and analyse the phase, composition and properties of matter. They are using accurate and appropriate chemistry models in their explanations, and understand the assumptions used to construct the models.	a Composition: How can we account for composition? b Structure: How can we account for 3D structure (eg crystal structure, formation of drops)? c Properties: How can we account for variations in the properties of matter (eg boiling point, viscosity, solubility, hardness, pH etc)? d Amount: What assumptions do we make when we measure the amount of matter (eg non-ideal gas law, average mass)?
7-9 Formulation How can we think about interactions between molecules? (Multirelational)	The composition, structure and properties of matter are related to how electrons are distributed among atoms.	Students are developing a more coherent understanding that matter is made of particles and the arrangements of these particles relate to the properties of matter. Their definitions are accurate, but understanding is not fully developed so that student reasoning is limited to causal instead of explanatory mechanisms. In their interpretations of new situations students may over-generalise as they try to relate multiple ideas and construct formulas.	a Composition: Why is the periodic table a roadmap for chemists? (Why is it a periodic table?) How can we think about the arrangements of electrons in atoms? (eg shells, orbitals) How do the numbers of valence electrons relate to composition (eg transfer or sharing)? b Structure: How do connections between atoms (bonds) and motions of atoms explain 3D structure (diamond rigid, water flows, air invisible)? c Properties: How can matter be classified according to bonds (ionic solids dissolve in water, covalent solids hard, matter phase)? d Amount: How can one quantity of matter be related to another (eg mass/mole/number, ideal gas law, Beer's law)?

Level of success	Big ideas	Descriptions of level	Item exemplars
4-6 Recognition How do chemists describe matter? (Unirelational)	Matter is categorised and described by various types of subatomic particles, atoms and molecules.	Students explore the language and symbols used by chemists to describe matter. They relate numbers of electrons, protons and neutrons to elements and mass, and the arrangements and motions of atoms to composition and phase. Ways of thinking about matter are limited to relating one idea to another at a simplistic level of understanding.	a Composition: How does the periodic table show trends? How are elements, compounds, and mixtures classified by letters and symbols? b Structure: How do the arrangements and motions of atoms differ in solids, liquids and gases? c Properties: How can the periodic table be used to predict properties? d Amount: How do chemists keep track of quantities of particles (eg number, mass, volume, pressure, mole)?
1-3 Notions What do you know about matter?	Matter has mass and takes up space. It can be classified according to how it occupies space.	Students articulate ideas about matter, and use experience, observation and logical reasoning to provide evidence. Focus is largely on macroscopic (not particulate).	a Composition: How is matter distinct from energy, thoughts, feelings? b Structure: How do solids, liquids and gases differ from one another? c Properties: How can you use properties to classify matter? d Amount: How can you measure the amount of matter?

Figure 2 Perspectives of chemists framework, matter variable

In this approach, the idea of a progress variable is focused on the concept of progression or growth. Learning is conceptualised not simply as a matter of acquiring quantitatively more knowledge and skills, but as progress toward higher levels of competence as new knowledge is linked to existing knowledge and as deeper understandings are developed from and take the place of earlier understandings. To use the BEAR assessment system in any given area it is assumed that learning can be described and mapped as progress in the direction of qualitatively richer knowledge, higher-order skills and deeper understandings.

Variables are derived in part from research into the underlying cognitive structure of the domain and in part from professional opinion about what constitutes higher and lower levels of performance or competence, but are also informed by empirical research into how students respond to instruction or perform in practice (Pellegrino et al, 2000). To more clearly understand what a progress variable is, let us consider an example. A university chemistry assessment project at University of California, Berkeley called *ChemQuery* recently developed a framework of progress variables called 'Perspectives of Chemists' that attempts to embody understanding of chemistry from a novice to expert level of sophistication. Its three variables describe chemistry views regarding three 'big ideas' in the discipline: matter, change, and stability. The **matter** strand is concerned with describing atomic and molecular views of matter. **Change** involves kinetic views of change and the conservation of matter during chemical change. **Stability** considers the network of relationships in conservation of energy. The Matter progress variable is shown in Figure 2. It describes how a student's view of matter progresses from a continuous, real-world view, to a particulate view and then builds in sophistication.

An example of a ChemQuery assessment prompt and actual student answers at Levels 1 and 2 is shown in Figure 3, along with interpretation. When students can begin to combine and relate patterns to account for, for instance, the contribution of valence electrons and molecular geometry to dissolving, they are considered to have moved to Level 3, Formulating, on the matter variable. Remaining levels of the framework represent further extensions and refinements on Level 3 and are not expected to be mastered at the introductory undergraduate levels, so are not addressed here.

This example shows how a progress variable can generate information on student mastery. Creating the developmental progress variables is not a trivial task. But having succeeded in adapting this approach to a given curriculum, the instructor will be well situated to address many of the issues raised in the first part of this paper. This approach assumes a match between instruction and assessment, which we address next.

Principle 2: Match between instruction and assessment

The match between the instruction and assessment in the BEAR assessment system is established and maintained through two major parts of the system: progress variables, described above, and assessment tasks or activities, described in this section. The main motivation for the progress variables so far developed is that they serve as a framework for the assessments and a method of making measurement possible. However, this second principle makes clear that the framework for the assessments and the framework for the curriculum and instruction must be one and the same. This is not to imply that the needs of assessment must drive the curriculum, nor that the curriculum description will entirely determine the assessment, but rather that the two,

assessment and instruction, must be in step - they must both be designed to accomplish the same thing, the aims of learning, whatever those aims are determined to be.

Using progress variables to structure both instruction and assessment is one way to make sure that the two are in alignment, at least at the planning level. In order to make this alignment concrete, however, the match must also exist at the level of classroom interaction and that is where the nature of the assessment tasks becomes crucial. Assessment tasks need to reflect the range and styles of the instructional practices in the curriculum. They must have a place in the 'rhythm' of the instruction, occurring at places where it makes instructional sense to include them, usually where instructors need to see how much progress their students have made on a specific topic. See Minstrell (1998) for an insightful account of such occasions.

Question:
You are given two liquids. One of the solutions is butyric acid with a molecular formula of C4H8O2. The other solution is ethyl acetate with the molecular formula C4H8O2. Both of the solutions have the same molecular formulas, but butyric acid smells bad and putrid while ethyl acetate smells good and sweet. Explain why you think these two solutions smell differently.

Student answers at Level 1 of visualising matter progress variable

Response: I think there could be a lot of different reasons as to why the two solutions smell differently. One could be that they're different ages and one has gone bad or is older which changed the smell. Another reason could be that one is cold and one is hot.

Response: Using chemistry theories, I don't have the faintest idea, but using common knowledge I will say that the producers of the ethyl products add smell to them so that you can tell them apart.

Response: Just because they have the same molecular formula doesn't mean they are the same substance. Like different races of people: black people, white people. Maybe made of the same stuff but look different.

Analysis: These students use ideas about phenomena they are familiar with from their experience combined with logic/comparative skills to generate a reasonable answer, but do not employ molecular chemistry concepts.

Student answers at Level 2 of visualising matter progress variable

Response: They smell differently because even though they have the same molecular formula, they have different structural formulas with different arrangements and patterns.

Response: "Butyric acid smells bad. It's an acid and even though they have the same molecular formula but they structure differently.

Analysis: Both responses appropriately cite the principle that molecules with the same formula can have different structures, or arrangements of atoms within the structure described by the formula. However, the first answer shows no attempt and the second answer shows an incomplete attempt to use such principles to describe the simple molecules given in the problem setup, which would have advanced response to the next level.

Figure 3 To match instruction and assessment, this LBC assessment question followed a laboratory project in which students explored chemicals that had different smells.

One good way to achieve this is to develop both the instructional materials and the assessment tasks at the same time - adapting good instructional sequences to produce assessable responses and developing assessments into full-blown instructional activities. Doing so brings the richness and vibrancy of curriculum development into assessment and also brings the discipline and hard-headedness of assessment data into the design of instruction.

By developing assessment tasks as part of curriculum materials, they can be made directly relevant to instruction. Assessment can become indistinguishable from other instructional activities, without precluding the generation of high-quality, comparative, and defensible assessment data on individual students and classes.

The variety of assessment tasks used by the BEAR assessment system can range widely, including individual and group challenges, data interpretation questions, and tasks involving student reading, laboratory or interactive exercises. In ChemQuery tasks, all assessment prompts are open-ended, requiring students to fully explain their responses. For the vast majority of assessment tasks, the student responses are in a written format.[1]

Whatever the form of instruction, if student work is generated or students can be observed at work and this work can be scored and matched to progress variables, then it is possible to consider use of an assessment system such as BEAR and to clearly match the assessments to instruction.

Principle 3: Management by instructors

For information from the assessment tasks and the BEAR analysis to be useful to instructors and students, it must be couched in terms that are directly related to the instructional goals behind the progress variables. Open-ended tasks, if used, must be quickly, readily and reliably scorable. Our response to these two issues are scoring guides (for instance, rubrics), scorable by people, such as students themselves, by readers, teaching assistants and instructors, or scorable by machine, say, using web-based interfaces with real-time delivery of instructional material and feedback, or more traditional machine-readable answer sheets.

Note that decisions on the structure of tasks and deployment of scoring and guides can be made course by course, but there should be a balance between the time constraints and needs of instructors for automatic machine scoring or reader scoring against the metacognitive needs of students to have instructors understand, engage and react to student levels of performance.

When scoring guides are used, instructors and students need concrete examples, which we call exemplars, of the rating of student work. Exemplars provide concrete examples of what an instructor might expect from students at varying levels of development along each variable. They are also a resource to understand the rationale of the scoring guides. Actual samples of student work, scored and moderated by those who pilot-tested the BEAR assessment system in ChemQuery, are available for each activity. These illustrate typical responses for each score level, as well as atypical responses that exercise the raters' skills.

---

[1] This is not a limitation of the BEAR system, but reflects the only practical way we then had available for instructors to attend to a full classroom of student work.

In addition to the scoring guides, the instructor needs a tool to indicate when assessments might take place, and what variables they pertain to. These are called Assessment Blueprints and are a valuable tool for keeping track of when to assess students. Assessment tasks are distributed throughout the course at opportune points for checking and monitoring student performance, as indicated in the Assessment Blueprints.

Principle 4: Evidence of quality

Technical issues of reliability and validity, fairness, consistency, and bias can quickly sink any attempt to measure along a progress variable as described above. To ensure comparability of results across time and context, procedures are needed to:

a   examine the coherence of information gathered using different formats

b   map student performances onto the progress variables

c   describe the structural elements of the accountability system - tasks and raters - in terms of the achievement variables, and

d   establish acceptable levels of system functioning, in terms of quality control indices such as reliability.

While this type of discussion can become very technical to consider, it is sufficient to keep in mind that the traditional elements of assessment standardisation, such as validity and reliability studies and bias and equity studies, must be carried out to satisfy quality control and ensure that evidence can be relied upon.

Our approach on this technical end of measurement is to use item response modelling, as described by Adams and Wilson (1992, 1996). These are examples of measurement models well developed enough for use in classroom-based assessment in a fairly routine and feasible way. The output from these models can be used as quality control information to address the concerns above, and to determine where individual students fall on a progress variable such as *ChemQuery's* matter variable, or any other progress variable that might be conceived and validated. Such output was used to validate and calibrate the matter progress variable, and to create the map of the progress variable in Figure 4. The left side of this map shows the measured distribution of students who responded to the matter items in 2001-02 trials, and the right side shows the calibrated difficulty of the tasks. Item response modelling can be used to locate a student or describe an entire class along a progress variable, as well as generate fit statistics and other indices for how well levels specified by the model fit classroom data. This becomes evidence for whether the measurement instruments are working well or not. For instance, in Figure 4, Task 6 at Level 2 appears lower than expected on the map. This means it was easier for students to successfully complete this task than for other similar tasks at this level, suggesting that the task may need to be examined to understand why it is measuring differently - are students readily able to guess on this item, for instance, or is there a clue to the answer in the question setup? Tables of reliability coefficients and standard errors are generated for further evidence, and inter-rater comparisons also can be made.

The formal nature of these models and their flexibility allows technical challenges inherent in the classroom assessment situation to be addressed, such as the maintenance of instructor rating consistency and the maintenance of a meaningful scale throughout the school year. This puts richer information into the hands of

instructors in the classroom. The central feature is the progress map, which provides a graph of the progress that students are making through the curriculum. Shown in Figure 5 are two types of progress map. The upper map shows the developmental progress of one student on a single variable over time. The first time point is a pretest score, followed by a first and second assessment of the same student further on in the curriculum and a final assessment at the end of instruction. The horizontal bands represent the score levels of the visualising matter progress variable, calibrated by item response modeling of data from classroom tests of the assessment system completed during curriculum development. The second map shows where the same student stands on several progress variables at a single timepoint, and offers brief advice for this student on how to improve in each area. Advice can be developed by teachers and many types of maps are available, derived from analysis of student data collected in coursework.[2]

Maps can be used to record and track student progress and to illustrate the skills a student has mastered and those that the student is working on. By placing students' performance on the continuum defined by the map, instructors can demonstrate students' progress with respect to the goals and expectations of the course. The maps, therefore, are one tool to provide feedback on how students as a whole are progressing in the course.
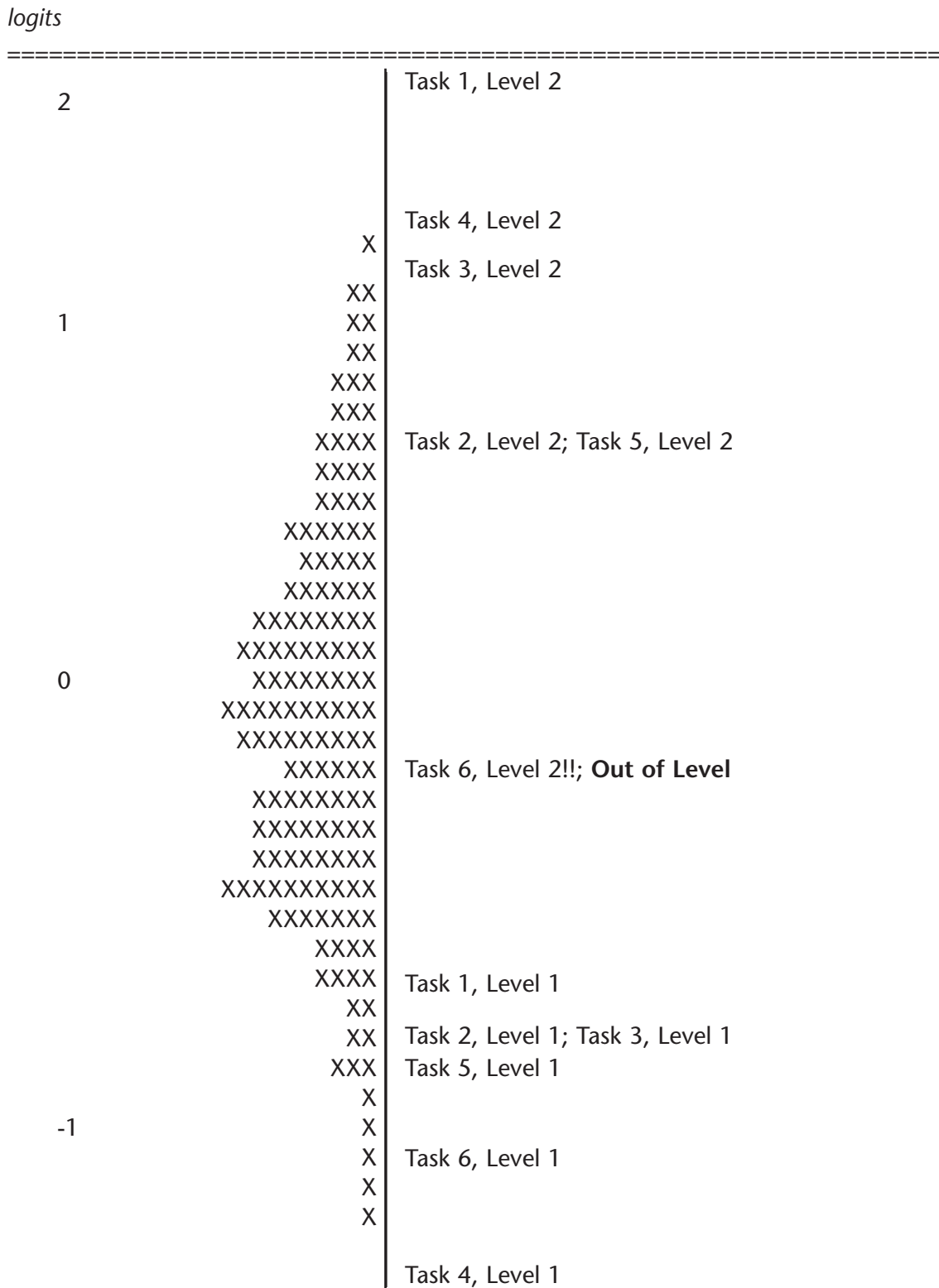
Maps, as graphical representations of student performance on assessment tasks, can be used to show how students are developing on progress variables throughout the course. This can then be used to inform instructional planning. For instance, if the class as a whole has not performed well on a variable following a series of assessments, then the instructor might feel the need to go back and re-address those concepts or issues reflected by the assessments. Additionally, during the development stage, unsatisfactory map results can indicate changes or additions to the curriculum.

Bringing it all together: assessment moderation

The four principles of the BEAR system are not designed to operate in isolation. Each of the principles provides a unifying 'thread' throughout the system, but their interrelationships also make the system more integrated. For example, the progress variables provide an initial unity to the curriculum materials, and define not only the content of student logits learning but also the paths over which student learning develops throughout the year. The implication is that each assessment, then, has a designated place in the instructional flow, reflecting the type of learning that students are expected to demonstrate at that point in time. Hence, scores assigned to student work can then be linked back to the developmental perspective and used both to diagnose an individual's progress with respect to a given variable and also to 'map' student learning over time.

---

[2] These maps were drawn in GradeMap, a software package developed at the University of California, Berkeley (Wilson, Kennedy and Draney, 2004). The analyses for these maps were performed using the ConQuest software (Wu, Adams and Wilson, 1998), which implements an EM algorithm for estimation of multidimensional Rasch-type models. For details on estimation and model-fitting, see Draney and Peres (1998).

logits

```
=====================================================================
                                 |  Task 1, Level 2
        2
                                 |
                                 |
                                 |  Task 4, Level 2
                            X    |
                                 |  Task 3, Level 2
                           XX    |
        1                  XX    |
                           XX    |
                          XXX    |
                          XXX    |
                         XXXX    |  Task 2, Level 2; Task 5, Level 2
                         XXXX    |
                         XXXX    |
                       XXXXXX    |
                        XXXXX    |
                       XXXXXX    |
                     XXXXXXXX    |
                    XXXXXXXXX    |
        0            XXXXXXXX    |
                   XXXXXXXXXX    |
                    XXXXXXXXX    |
                       XXXXXX    |  Task 6, Level 2!!; **Out of Level**
                     XXXXXXXX    |
                      XXXXXXX    |
                      XXXXXXX    |
                    XXXXXXXXX    |
                     XXXXXXX    |
                         XXXX    |
                         XXXX    |  Task 1, Level 1
                           XX    |
                           XX    |  Task 2, Level 1; Task 3, Level 1
                          XXX    |  Task 5, Level 1
                            X    |
       -1                   X    |
                            X    |  Task 6, Level 1
                            X    |
                            X    |
                                 |
                                 |  Task 4, Level 1
=====================================================================
```

*Numbers at left are in a unit called the logit, or log of the odds, with higher numbers indicating better student performance.

Figure 4 Calibration of LBC matter variable (partial credit, generalised-item thresholds)
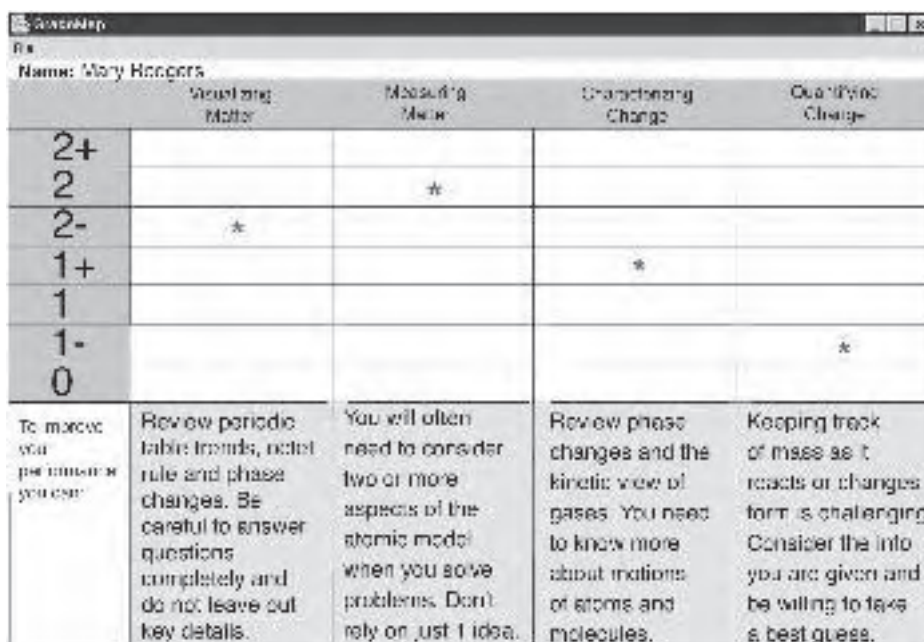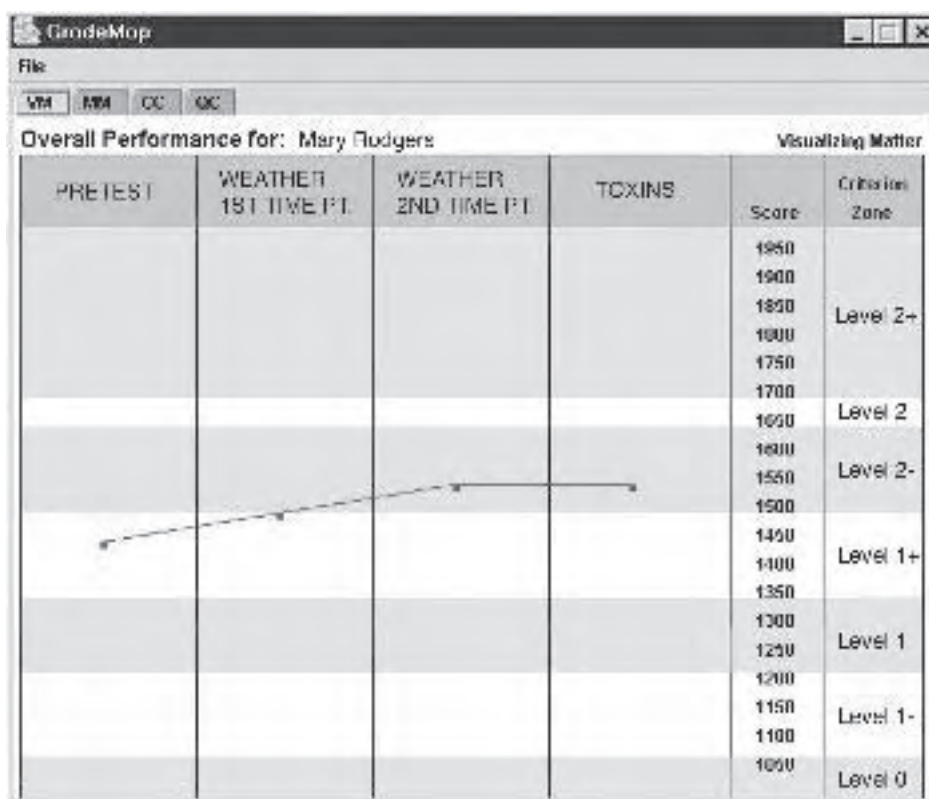
Figure 5 Progress maps for one student over time on a single variable (top) and same student at one time with score and advice for how to improve on several variables (bottom)

Adherence to each of the principles across each of the phases of the assessment process produces a coherence or 'internal consistency' to the system. Adherence to each of the principles within each phase of the assessment process produces a well-integrated system that addresses desired linkages among curriculum, instruction, and assessment.

Proper operation of the BEAR assessment system requires that instructors and students take control of essential parts of the assessment system, including the scoring process. We have devised the 'assessment moderation meeting' as part of our staff and student development strategy to accomplish these goals.

Moderation is the process by which instructors, teaching assistants, readers, students and others involved in a course discuss student work and the scores for work, ensuring that scores are interpreted in a similar way by all in the moderation group.

In instructor moderation sessions, instructors discuss the scoring, interpretation and use of student work, and make decisions regarding standards of performance and methods for reliably judging student work related to those standards. The moderation process gives instructors the responsibility of interpreting the scores given to students' work and allows them to set the standards for acceptable work. Instructors use moderation to adapt their judgments to local conditions. Upon reaching consensus on the interpretations of score levels, instructors can then adjust their individual scores to better reflect the instructor-adapted standards. The use of moderation allows instructors to make judgments about students' scores in a public way with respect to public standards and improves the fairness and consistency of the scores.

Moderation sessions also provide the opportunity for instructors to discuss implications of the assessment for their instruction, for example, by discussing ways to address common student mistakes or difficult concepts in instructional sequence. This last aspect of moderation is perhaps the strongest influence of moderation on instruction.

Moderation also can take place with students, so they better grasp what the instructor and course are valuing in terms of student learning. Students can score class work, if that is appropriate, or can score work provided as examples in the curriculum materials. They can map scores against progress variables and see more concretely the paths toward mastery of learning aims. (See video, *Moderation in All Things: A Class Act*, Berkeley Evaluation & Assessment Research Center.)

Conclusion

It seems clear that going beyond grades to map individual trajectories of learning is feasible, especially where computers and data collection devices are readily available in higher education and when instructional materials are restructured to use these tools to better accommodate improved formative assessment. In practice, the benefit to students is promising. Wilson and Sloane (2000) have documented evidence of the educationally important and statistically significant effects that use of the system can have on student performance. Furthermore, it is interesting to consider how theories of learning and theories of instruction may change as a result of better data and a clearer understanding of learning trajectories. We hope to see many embedded assessment efforts unfold in coming years, especially in large lecture classes, and invite those interested to, of course, consider using the BEAR assessment system.

But, moreover, it is important for educators to ponder the principles behind successful formative assessment. It is in satisfying these principles that the argument for formative assessment lies, and the metacognitive needs of students can be met.

'This kind of analysis gives me more than just a grade,' said one instructor using the BEAR system. 'I can diagnose a problem and move forward with a greater number of students. I can see the amount of time it takes for my students to learn, and find out how much of something they know, or how well they know it, not just whether they have a fact in their heads or they don't. It lets me value even wrong answers, because it shows me what in each answer I can value and support and work with. To me, it's a whole different way to truly value student thinking.'

Acknowledgements

References

Adams R J and Wilson M (1992) *A random coefficients multinomial logit: generalizing Rasch models*, paper presented at the annual meeting of the American Educational Research Association, San Francisco

Adams R J and Wilson M (1996) Formulating the Rasch model as a mixed coefficients multinomial logit, in G Engelhard and M Wilson (eds) *Objective measurement III: Theory into practice*, Ablex, Norwood

Black P and Wiliam D (1998) Inside the Black Box: Raising Standards through Classroom Assessment, *Phi Delta Kappan,* 80 (2) 139-148

Claesgens J K, Scalise K, Draney K, Wilson M and Stacy A (2002) *Perspectives of chemists: A framework to promote conceptual understanding of chemistry*, Paper presented at the annual meeting of the American Educational Research Association, New Orleans

Dweck C S and Leggett E L (1988) A social-cognitive approach to motivation and personality, *Psychological Review,* 95, 256-273

Hesse J (1989) From Naive to Knowledgeable, *The Science Teacher*, 55-58

Masters G N, Adams R A and Wilson M (1990) Charting student progress, in T Husen and T N Postlewaite (eds) *International encyclopedia of education: Research and studies, Supplementary volume 2*, pp628-34, Pergamon Press, Oxford

Minstrell J (1998) *Student thinking and related instruction: Creating a facet-based learning environment*, Paper presented at the meeting of the Committee on Foundations of Assessment, Woods Hole, MA

Pellegrino J, Chudowsky N, and Glaser R (2001) *Knowing What Students Know: The Science and Design of Educational Assessment*, NRC Center for Education, Washington DC, National Academy Press

Wilson M (1990) Measurement of developmental levels, in T Husen and T N Postlewaite (eds) *International encyclopedia of education: Research and studies, Supplementary volume*, Pergamon Press, Oxford

Wilson M (in press) *Constructing Measures: An Item Response Modeling Approach,* Mahwah, Erlbaum, New Jersey

Wilson M and Sloane K (2000) From principles to practice: An embedded assessment system, *Applied Measurement in Education*, 13 (2) 181-208

Wilson M C, Kennedy and Draney K (2004) GradeMap (Version 40) [computer program] University of California, Berkeley, BEAR Center

Wolf D, Bixby J, Glenn J III and H Gardner (1991) To use their minds well: Investigating new forms of student assessment, *Review of Research in Education,* 17, 31-74

Tutors, who needs them? Student self assessment in an accounting degree using university-wide common grade related criteria

Win Hornby and David Laing, Centre for Enhancement of Learning and Teaching, The Robert Gordon University

Abstract

Most previous research on self-assessment has been conducted in an assessment environment where percentage marks are used or where grades based on marks are generally employed. This paper considers whether the use of university-wide grade related criteria (GRC) in self-assessing performance with final year honours accounting students can improve the accuracy of self-assessment. Based on a sample of 48 students, a highly significant correlation was found between tutor and student grades ($r = 0.61$) with 53 per cent awarding themselves exactly the same grade as the tutor and a further 45 per cent at a one grade variance. Students significantly under assessed themselves on all of the criteria employed, with the exception of the evaluation dimension. No gender differences were found in the accuracy of the self assessment on overall grade, although female students under assessed their ability on the research dimension to a significantly greater extent than male students. The very best students were more able to self assess than poorer students. Qualitative evidence suggests that students now understand the basis on which they were assessed and learned more from the feedback they received. The use of self-assessment, using GRC has the potential to change the role of the tutor and to make assessment both more effective and more efficient.

Introduction

Most previous research on self-assessment has been conducted in an assessment environment where percentage marks are used or where grades based on marks are generally employed. This paper considers the use of university-wide grade related criteria in self-assessing performance on an economics of taxation module among final year honours accounting students. The paper examines the proposition that, given the alleged benefits of clear and transparent assessment criteria and student engagement with the criteria, there should be a significant improvement in the accuracy of self-assessment.

Self-assessment by students has long been thought to be an important method of assessment. Most of the previous research on students' self assessment has been conducted in an environment where students are marked using conventional percentage marks or grades which are derived from percentage marks. This paper looks at the use of self-assessment for undergraduate students using a university-wide GRC assessment scheme. This scheme, which was introduced across The Robert Gordon University in 1999, aims to provide greater transparency to all the stakeholders in the assessment process. One way to test whether such transparency exists, as far as the students are concerned, is to use it for the purposes of self-assessment. Having been introduced to the GRC and having had the basis on which they were to be assessed explained, would students be more able to assess themselves accurately? If GRC do provide greater accuracy in self-assessment how does this affect the tutor's role in the assessment process? Are tutors necessary? Would it, for example, mean that the tutor's role becomes less that of a conventional assessor and more that of a moderator/facilitator/trainer in the assessment process?

This paper is therefore concerned with examining the validity and reliability of self-assessment using a university-wide system of GRC. After reviewing recent research on self-assessment, the paper sets out the background to the introduction of the GRC system and gives the context within which the self-assessment was carried out. It then outlines the methodology and reports on the results of both the quantitative and qualitative data from a sample of 48 final year accounting and finance honours students. This paper is part of an ongoing longitudinal evaluation of the GRC system.

Review of the literature

For well over 20 years the literature on assessment has made reference to the advantages of students assessing themselves.[1] More recently, however, in a number of the papers in the research literature, the emphasis has been less on how to do it and more on empirically testing the reliability as well as the validity of student self assessment. For example, Mabe and West (1982) reviewed 55 studies from 1942 to 1977 involving a total of 267 correlations between students and tutor ratings in which self-evaluations of ability were compared with measures of performance. Their results showed that there was a poor relationship with self-evaluation of ability and more objective measures of performance with significant evidence of over-reporting of ability according to self-evaluations. Their review revealed a low mean validity coefficient (mean r =0.29) and a high degree of variability (SD = 0.25). According to their evidence, Mabe and West

---

[1] See for example Boud and Lubin (1983), Boud (1986), Rowntree (1989) and more recently Brown and Glasner

report that the key factors affecting the accuracy of self evaluations were:

a   the rater's expectation that the self evaluation would be compared with criterion measures

b   the rater's previous experience with self-evaluation

c   instructions guaranteeing anonymity of the self-evaluation and

d   self evaluation instructions which emphasised comparison with others.

Of the 22 studies which reported on over/under assessment, 15 reported evidence of over assessment by the self-evaluation, four produced mixed results and only three reported under assessment.

In one of the most comprehensive reviews of what research tells us about self-assessment, Boud and Falchikov (1995) reviewed a total of 68 studies from 1932 until 1994. They looked at how frequently students over rate themselves (17 studies) as opposed to under rate (11 studies); at whether good students rate themselves more accurately than poorer students (11 studies in total of which nearly all studies showed that better students were in fact more accurate or rated themselves below tutors grades than poorer students). They also reviewed seven studies on whether more advanced students rated themselves more accurately than students at an introductory level (undergraduates and 'freshmen' overestimated their abilities more than postgraduates and senior years). In addition, they reviewed seven studies which looked at whether students got better with practice (The evidence suggested that they did not). Finally, they reviewed six studies on gender differences and found only three studies where gender differences were important with women more accurate self assessors than men. More recent work for instance by Stephani (1994) found no gender differences.

More recently in the accounting field, in an evaluation of the validity of self-assessment, which measured computer literacy among entry level undergraduates in accounting degrees at two universities in the UK, Larres, Ballantine and Whittington (2003) reveal significant differences in the students perceived and actual computer literacy with the vast majority overestimating their computer knowledge, with the more able students being more accurate in their self-assessment. They conclude that self-assessment is not an appropriate means of determining computer literacy among entry level undergraduate accounting students when used in isolation. Used as an adjunct to a more robust measure of computer literacy they conclude that self-assessment provides a useful insight into students' attitudes to computing and stimulates reflection.

While research on the accuracy of self-assessment is clearly important, other studies have looked at the effect of self-assessment on enhancing learning. For example, Fitzgerald et al (1997) undertook an analysis of medical students self-assessment of knowledge and found significant improvements in learning. Finally, in one of the most comprehensive and rigorous empirical studies to date, Rust et al (2003) report on the findings of a two year study research project developing students' understanding of assessment criteria. Using a control and experimental design sample and a programme of planned intervention which allows students to 'engage' with the criteria on which they are assessed, Rust reports that students learning can be improved and that this improvement can be both significant and lasting over time.

Background

In 1999, The Robert Gordon University[2] introduced a university-wide system of GRC. The rationale for its introduction is outlined in Hornby (2003). It is well recognised in the research literature, that there is a considerable degree of heterogeneity in the way assessments are carried out, even within a given, fairly well defined and recognised cognate area such as business administration. Even in those universities which employed grading schemes, the underlying principles of such schemes are often still those of the percentage model. Based on the empirical evidence it is questionable that these scales are used consistently (Hornby, 2003). It is also debatable whether the percentage scale is in fact a 'common currency' at all. Percentage scales give a false impression of commonality and it can be argued that one way of overcoming this spurious commonality is to adopt a clearly defined criterion referenced grading system.[3] Based on evaluation data from students and survey data from staff, Hornby (2003) reports both students and staff viewed the introduction of the GRC favourably. Using a controlled sample over consecutive sessions, the research evidence shows that students consistently rated the clarity of the marking criteria employed by staff as better using the GRC system than the conventional percentage model. Staff also reported that the system of GRC was more reliable, valid and transparent. They also reported that it provided students with better feedback on performance.

However, the use of a university-wide GRC system in students' self-assessing is not well researched. Given that few universities have adopted institution-wide GRC systems, it was felt that the present study might reveal some interesting findings about reliability of self-assessment methods. Intuitively, it was felt that if the criteria were clearer then students should be more able to assess themselves. If it could be shown that the use of GRC systems significantly improved the reliability of self assessment and if students could be inducted successfully into this method of assessment, might it be possible to use students' self assessment as a method of streamlining the assessment process? It is to test these questions that the paper now turns.

Context

In session 2002-03 in the module on the Economics of Taxation and Corporate Taxes (BS 4214), 48 final year accounting students, as part of the assessment of the module, undertook a piece of independent research on a tax topic. In order to fulfil the coursework requirements, students were asked to submit a report of no more than 2,500 words on one of three topics on the economics of tax.

In previous sessions, tutors teaching this part of the course had reported that they had found difficulty in explaining the basis on which marks had been awarded. Accounting students, more used to computational type questions where they were involving staff from different departments, as was the case with module BS 4214, the problems of comparability were exacerbated. The introduction of the GRC scheme across the university thus provided an opportunity to use a common yardstick by thought to be

---

[2] The Robert Gordon University is located in Aberdeen, Scotland and was established in 1992. It consists of 10,000 full-time equivalent students and almost 600 academic staff. There are three faculties, of more or less equal size - the Faculty of Design and Technology, the Faculty of Health and Social Care and the Aberdeen Business School/Faculty of Management.

[3] For details of the scheme see www.rgu.ac.uk/academicaffairs/assessment/page.cfm?pge=2295

clearly defined 'right' answers, were often very critical and uncertain about the basis on which marks were awarded in this more discursive part of the module. In addition, it has been shown that different academic disciplines have different marking conventions (Bridges, 1999; Yorke, 2000; Hornby, 2003). Therefore while it was not uncommon for students to get percentage marks in a range 20-90 per cent in accounting subjects, in economics, marks ranges were generally speaking very much narrower. Where joint teaching and assessment on a module took place which students would be assessed.

Using the University's Assessment Template and the seven point scale (see Appendix 1) and adapting Bloom's taxonomy (Bloom, 1956), five criteria for assessing the students were identified and statements of expected student performance were drawn up and distributed (see Appendix 2). These were identified (with appropriate weighting) as:

- presentation 10 per cent
- research 10 per cent
- knowledge and understanding and application of key concepts 20 per cent
- analysis and synthesis 30 per cent
- critical evaluation 30 per cent.

Students were then briefed on the criteria in a one hour workshop. Each of the dimensions of the criteria was 'unpacked' and examples of what was implied by each of the criteria were explained. Although students were not given the opportunity to use the criteria themselves to assess examples of other students work[4], an online discussion forum was set up following the five step model proposed by Salmon (2002) to which all students subscribed and in which a number took part in discussions about the coursework and how it was to be assessed. This allowed all students, 'lurkers'[5] and participants alike, to share in the understanding of how they were to be assessed. Thus there was some significant engagement with the criteria by the students although not as intensive as that described by Rust (2003). There was, however, more engagement than would conventionally have been the case where normal custom and practice in previous years with this module was to issue the coursework brief with the criteria outlined with little or no further discussion taking place thereafter.

Students and tutor were then required independently to use the seven point scale to assess the work. A final overall grade was arrived at by looking at the profile of a student's performance over all five of the criteria using the algorithm that is used by the university to determine honours classifications. Thus if a student achieved 50 per cent of the criteria at a given grade (known as the provisional recommended grade or PRG) with no more than 25 per cent of the criteria two grades below this grade, the final grade they receive will be the PRG.[6]

They were also asked to identify what they considered to be the strengths and weaknesses of the work and to indicate what they might have done differently in order to improve the work. Thus it was possible to analyse the extent to which

---

[4] This was the first time this assignment had been set so no previous material existed for students to 'practice' assessing.

[5] This term is used to mean those students who observe the online discussion but do not participate.

[6] For details of the 'rules' governing profiling and examples of it working in practice see www.rgu.ac.uk/celt/quality/page.cfm?pge=6201 Strictly speaking it is not permissable to combine GRC since they are not marks. However, for practical reasons, it is often necessary to find a single grade for each module and for determining an honours classification. The rules quoted above apply across the university and thus give a consistent set of principles by which this done.

students ratings matched tutor ratings on each of the five dimensions and also on overall grade. The next section looks at the results.

The results

Quantitative analysis

In order to assess the accuracy of the self-assessment, three quantitative measures are used in this study.

1   Actual grades

The actual grades used by students and tutor are averaged across the sample and compared for each dimension. Correlations between tutor's and students' assessments were also conducted. This data is quoted in Tables 1 and 2 below.

2   Degree of match

This is defined as the percentage of the sample whose rating exactly matches the tutor's rating on each dimension. In addition the distribution of the sample with a variance of one and two grades is also computed. Finally, the percentage of the sample whose rating exceeds that of the tutor is also reported. This data is contained in Table 3.

3   Univariate analysis

The analysis here examines both the actual values as well as the variance between tutor grades and student self assessment using statistical tests. Table 4 uses non-parametric univariate tests to compare the students' and tutor grades for each dimension. Table 5 investigates the variance between the student and tutor grades using the same non-parametric tests.

Variable	Minimum	Maximum	Mean	Median	Standard deviation (SD)
Tutor grades					
Presentation	4	6	5.84	6	0.426
Research	2	6	5.31	6	1.045
Knowledge and understanding	3	6	5.37	6	0.883
Analysis	3	6	4.96	5	0.841
Evaluation	2	6	4.10	4	0.984
Overall grade	3	6	4.94	5	0.827

Variable	Minimum	Maximum	Mean	Median	Standard deviation (SD)
Student grades					
Presentation	4	6	5.45	6	0.709
Research	3	6	5.04	5	0.865
Knowledge and understanding	3	6	4.71	5	0.645
Analysis	3	6	4.41	4	0.674
Evaluation	3	6	4.37	4	0.636
Overall grade	3	6	4.61	5	0.571

Table 1 Descriptive statistics of sample[7]
BAAF 4 results (n=48)

Analysis using the first measure outlined above indicates that the students on average consistently under-rated themselves on all of the dimensions with the exception of the evaluation dimension. Here they overestimated their abilities in comparison with tutor ratings. There was also evidence of a greater degree of variability in the tutors assessment of the evidence of research (SD for tutors rating = 1.045; SD for students rating SD = 0.865) and on evaluation than was evident from the students self assessment (SD for tutors rating = 0.984; SD for students rating SD = 0.636).

Dimension	Correlation between tutor grade and student grade
Presentation	0.179 ns
Research	0.470 ***
Knowledge	0.444 ***
Analysis	0.471 ***
Evaluation	0.438 ***
Overall grade (Whole integer data)	0.611 ***

*** Significant at one per cent level

Table 2 Correlation table

Table 2 confirms the significant degree of agreement between the tutor's ratings and the student's self-assessment. The highest level of agreement is on the knowledge and understanding dimension and the lowest level of agreement on presentation. While on the face of it this may seem to contradict the evidence from Table 1, the correlation coefficient gives a more powerful test of the level of agreement since it measures the rank ordering of the grades as well as just simple measures of

---

[7] The grading system use had 0 as its poorest grade and 6 as its highest grade.

agreement and variance. The fact that there is a statistically significant correlation between tutors and student ratings for all the criteria used is indicative evidence of the effectiveness of the self-assessment exercise. However, this is tested more rigorously using non-parametric univariate tests below.

Table 3 gives details on the degree of match in grades on each of the dimensions.

Element	Same grade	One grade variance	Two grade variance	Student's grade exceeds the tutor grade
Presentation	59.2	30.6	10.2	6.1
Research	42.9	46.9	10.2	16.3
Knowledge and understanding	20.4	69.4	10.2	12.2
Analysis	38.8	51.0	10.2	8.2
Evaluation	36.7	55.1	8.2	42.8
Overall	53.1	44.9	2.0	8.2

Table 3 Degree of match in grades[8] (Percentages used throughout) BAAF 4 results

Overall, 53 per cent of students gave exactly the same overall grade as the tutor. A further 45 per cent were plus or minus one grade with only two per cent having a two grade variance. Overall, only eight per cent of the sample over assessed themselves. However when analysis is done on each dimension, we discover that the exact degree of match is poorer for knowledge and understanding and evaluation than for example for presentation. 42.8 per cent of students rated themselves much better on evaluation than the tutor.

The hypotheses

There are a number of hypotheses which require to be tested. The first of these was to test to see whether there were any statistically significant differences in the degree of match in grades awarded by the tutor and the students. Thus the first hypothesis is:

H1: There will be no statistically significant differences in the degree of match in grades awarded by the tutor and the students.

---

[8] The grading system use had 0 as its poorest grade and 6 as its highest grade.

Element	Kruskall-Wallis test	Mann-Whitney test
Presentation	10.087 ***	3.176***
Research	3.859 **	1.964 **
Knowledge and understanding	20.528 ***	4.531 ***
Analysis	12.343 ***	3.513 ***
Evaluation	1.611	1.269
Overall	6.013 **	2.452 **

* - significant at the 10 per cent level
** - significant at the five per cent level
*** - significant at the one per cent level

**Table 4 Univariate analysis - a comparison of tutor and student grades[9]
BAAF 4 results**

Using actual grades awarded for each dimension by tutor and students and applying non-parametric statistical tests indicates that we can reject this hypothesis for all the criteria used in the assessment. There was a statistically significant difference in students' self assessment and tutor assessment on all the dimensions with the exception of evaluation. Contrary to the findings of Larres et al 2003, students in our sample consistently under assessed their abilities.

Element	Kolmogorov-Sminov Z test	One sample T test
Presentation	2.439 ***	3.577 ***
Research	1.621 ***	1.866 *
Knowledge and understanding	2.348 ***	5.505 ***
Analysis	1.716 ***	4.869 ***
Evaluation	1.536 **	-2.046 **
Overall	2.117 ***	3.474 ***

* - significant at the 10 per cent level
** - significant at the five per cent level
*** - significant at the one per cent level

**Table 5 An analysis of the variances between tutor and student grades[10]
BAAF 4 results**

---

[9] This table compares the actual grade awarded by the tutor and the student rather than the difference between them.

[10] This table investigates the variance between tutor and student grades eg a tutor grade of 6 and a student grade of 4 would result in a variance of 2.

Using an analysis of variances between tutors and students' grades also leads to the similar conclusion. There are highly significant variances for all the dimensions using the appropriate statistical tests. Therefore we can confidently reject the null hypothesis. Students consistently under assessed themselves on all the dimensions.

In addition, there are suggestions in the literature that female students will tend to undervalue their performance more than male students (Boud and Falchikov, 1995). Therefore a second hypothesis was tested.

H2: There will be no statistically significant gender differences in the degree of match between tutor and students assessments.

	Gender (Pearson Chi Square)
Presentation	1.584
Research	6.756 ***
Knowledge and understanding	1.197
Analysis	0.063
Evaluation	0.023
Overall	0.000

* - significant at the 10 per cent level
** - significant at the five per cent level
*** - significant at the one per cent level

Table 6 The influence of individual characteristics on self-assessment variances BAAF 4 results

The evidence suggests that this hypothesis can in fact be accepted, with the exception of the research dimension. This suggests that there were gender differences with there being more variance in ratings on the research dimension between female students and male students. Female students were much harder on themselves in rating the amount of research they had carried out than their male counterparts.

It is also sometimes suggested that 'good' students will be better at self-assessment than 'poorer' students (Boud and Falchikov, 1995). In order to test for this, a good student was defined in two ways. Firstly, a broad definition was used. A good student was thus defined as a student whose grade is 5 or more in the final assessment. A narrower definition was also used which defined a good student a one who scored a grade 6. Finally, a third definition was used based on a student's final honours classification. Those students who obtained a First or 2:1 honours degree were classified as good. Thus a third hypothesis is:

H3: There will be no statistically significant differences in the degree of match made by good and poor students and the tutor.

	1 Student ability (Pearson Chi Square)	2 Student ability (Pearson Chi Square)	3 Student ability (Pearson Chi Square)	4 Student ability (Pearson Chi Square)
Presentation	2.019	0.005	0.152	1.642
Research	2.069	3.679*	1.838	9.317***
Knowledge and Understanding	0.206	1.426	0.057	0.611
Analysis	2.581	6.204***	0.348	0.987
Evaluation	0.942	0.166	0.330	0.234
Overall	1.181	12.765***	0.016	0.023

* - significant at the 10 per cent level
** - significant at the five per cent level
*** - significant at the one per cent level

Table 7 The influence of individual characteristics on self-assessment variances BAAF 4 results[11]

The evidence suggests that on the broad definition and also using honours classification, there were no statistically significant differences between good students and weaker students on almost all of the dimensions or in the overall grade. The exception was that good honours students rated themselves lower than the tutors on the research dimension. However, on the narrower definition good students were more accurate in their self-assessment than weaker ones for the research dimension and the analysis dimension and in the overall grade. Thus the very best students were more accurate in their self-assessment on a number but not all of the dimensions.

Qualitative data

In addition, to complete the assessment grid, students were required to complete a self-evaluation pro forma in which they where asked to write about what lessons they had learned as a result of doing this work. They were asked to identify strengths and weaknesses and to comment on how they might have improved the work. Out of a cohort of 53, 48 students (90 per cent) completed all aspects of the proforma.

---

[11] Student ability refers to the overall grade that students were awarded by the tutor. We dichotomise students into those who were awarded the highest grades (5 and 6) in column 1 and grade 6 in column 2 and compared them with all other students. Column 3 splits the sample into those students who achieved a First and column 4 splits the sample by students who achieved a First or Upper Second. Our results are qualitatively the same when we use the Mann-Whitney test for both gender and student ability.

In a significant number of the comments, students used the criteria to report on strengths and weaknesses. For example, students reported as follows.

I feel I have enhanced my ability to perform research and critical analysis through this assignment.

Female student, overall grade 4

I feel the strengths of this report was (sic) the research conducted as well as the knowledge and understanding I gained from this.

Female student, overall grade 5

As a result of this work I have learned....that tax can be interesting! The strengths of this work is (sic) in its presentation, application of knowledge and analysis of the issues identified.

Male student, overall grade 4

From doing this work I have learnt the importance of doing the research early. I felt the way I approached my analysis was successful. It was not something I had done before in this way and it is an approach I would take again.

Female student, overall grade 6

In addition, the tutor was more able to provide feedback using a pre-prepared set of statements for each criterion in a statement bank. Thus it was possible to ensure that comments matched grades for each dimension. This approach helped give more effective feedback which also appears to have enhanced the learning process. As one student commented in the discussion forum after the coursework was returned:

Thank you for the extensive feedback from you regarding my coursework. That type of information will be extremely useful to me in the future. The self-assessment form, although initially seemed annoying, gave me a good insight into my work and I had never reviewed my own work in the way I did until this paper. Being 'forced' to do this let me see that even marking it I could spot places where I fell down and could improve. (Shame I didn't self assess long before the deadline!)

However, I would like to put in writing that this information would have been more useful had it been in use from right back in first year. I would suggest considering using that pro forma feedback forms on all courseworks. I have never received feedback as extensive as yours and feel that getting this back each time will lead to a gradual improvement of grades throughout the years of the course. (I'm sure that this information in earlier years would have lead to improvements in my grades over the years as opposed to getting round about the same percentage every year.) Breaking down the grade into separate sections also helped understand more fully how our papers are marked.

Just wanted to get that in writing to you. I feel if you were to put this issue to the class the general consensus would be that although the self-assessment was a hassle at the time, their accuracy is testimony to their usefulness and that such feedback information such as what (sic) was given would be extremely beneficial starting much earlier than the last round of courseworks!

Male student, overall grade 5

Discussion

Based on a small sample of honours students from one subject area, these results seem to indicate that there is considerable value in using self-assessment. It appears, from this study, that that there is more evidence of under assessment by students than has been reported in the literature. Students were told that the assessment which would 'count' in this exercise was the tutor's assessment. One might have speculated therefore that there might have been an encouragement for students to take the exercise less seriously than if their assessment had been the one to count. On the other hand, the fact that their assessment did not 'count' may have discouraged students from inflating their grades.

It is possible to speculate that there may have been an element of 'game theory' in their strategies. According to game theory, the 'least risk' (maximum) payoff strategy for students would be to rate themselves low. It may have been that students did not want to be seen to be over assessing themselves. On this view it would be less humiliating to be seen to be under assessing performance than to be over assessing where their perception of themselves would be contradicted. However, as the results of the exercise were not made public and as the tutor's assessment was done 'blind' (ie without knowledge of the student's rating) it is difficult to believe that this played a significant part in the self-assessment process.

There is evidence from the qualitative feedback that students, who obtained a higher grade than they gave themselves, received a considerable boost to their morale. In addition, they took a more positive view of their whole learning experience. As one student wrote in an e mail to her tutor:

> Thank you! I can't believe I got that mark (I'm still shaking). I honestly did think this was one of my poorer pieces of coursework, but I'm very glad you didn't agree! Thanks also for such a detailed feedback, it's not often we get this and have found it very useful.
> Female student, Grade 6

One explanation for the evidence of systematic under assessment may have been the use of grade related criteria themselves and explicit statements attached to them. Making explicit statements about the standards and discussing at some length what was meant by each of the criteria may have had an intimidating effect on some students. In consequence, they may have felt that the highest grades were beyond their ability.

It is also clear from the evidence that those dimensions of research and evaluation appeared to be where there was the greatest degree of variation. Thus it may be that the criteria were not as explicit as the tutor imagined or that there was less of a shared understanding than on more straightforward dimensions such as presentation.

Conclusion

The evidence from this research suggests that self-assessment using a university-wide GRC scheme can be a powerful way of enhancing learning.

More controversially, the results of this research appear to suggest that contrary to recent research evidence students did not over assess themselves. There is also

significant evidence that feedback to students is enhanced using statements which are linked to the GRC. With more detailed training and induction of students (including practice in assessing previous cohorts work) using the GRC scheme, it opens up the possibility of a significant change in the role of the tutor. Thus given what we have discovered about the accuracy of self-assessment using GRC and the general tendency for students to under assess themselves, the tutor's role could change from that of the 'primary' assessor to that of a moderator thereby significantly improving both the efficiency and effectiveness of the assessment process.

Future research needs to test these hypotheses on a larger data set with a more heterogeneous population in different subject disciplines. Further research is currently being conducted nationally in Scotland on the effectiveness and efficiency of assessment[12] and it is hoped that the results of our research will help inform this debate.

Acknowledgements

The authors would like to thank the teaching staff in the Department of Accounting and Finance of the Aberdeen Business School at The Robert Gordon University for allowing them access to their students and to all the final year accounting and finance students who took part in the self assessment exercise and who allowed them to quote from their emails and online discussion forum.

References

Bloom B S et al (1956) *Taxonomy of Educational Objectives I: The Cognitive Domain*, Longmans, London

Boud D (1986) *Implementing Student Self Assessment*, HERDSA Green Studies, Vol 5, University of New South Wales, Australia

Boud D and Falchikov N (1995) What Does Research Tell us about Self Assessment? in Boud D (ed) *Enhancing Learning through Self Assessment*, Kogan Page, London

Boud D and Lubin J (1983) *Self Assessment in Professional Education*, University of New South Wales, Australia

Bridges P et al (1999) Discipline-related Marking Behaviour Using Percentages: A Potential Cause of Inequity in Assessment, *Assessment and Evaluation in Higher Education*, 24 (3) 285-300

Brown S and Glasner A (eds) (1999) *Assessment Matters in Higher Education - Choosing and Using Diverse Approaches*, Open University Press, Buckingham

Fitzgerald T J, Gruppen L D, White C and Davis W K (1997) Medical Students' Self Assessment Abilities: Accuracy and Calibration, *Proceedings of the Meeting of the American Educational Research Association*, Chicago

Hornby W (2003) Assessing Using Grade Related Criteria: a Single Currency for Universities? *Assessment and Evaluation in Higher Education*, Vol 28, No 4, 435-454

---

[12] The author is a member of the sector-wide Quality Enhancement Steering Committee for Assessment which is looking at the issues of efficiency and effectiveness of assessment.

Larres P M, Ballantine J A and Whittington M (2003) Evaluating the validity of self-assessment: measuring computer literacy among entry-level undergraduates within accounting degree programmes at two UK universities, *Accounting Education*, 12 (2) 97-112

Mabe P A III and West S G (1983) Validity of Self Evaluation of Ability: A Review and Meta-Analysis, *Journal of Applied Psychology*, 67 (3) 280-296

Rust C, Price M and O'Donovan B (2003) Improving Students' Learning by Developing their Understanding of Assessment Criteria and Processes, *Assessment & Evaluation in Higher Education*, Vol 28, No 2, 147-164

Rowntree D (1989) *Assessing Students: How Shall We Know Them?*, Kogan Page, London

Salmon G (2002) *e-tivities: The Key to Active Online Learning*, Kogan Page, London

Stephani L A J (1994) Self, peer and group assessment procedures, in Sneddon I and Kremer J (eds) *An Enterprising Curriculum: Teaching Innovations in Higher Education*, Kogan Page, London

Yorke M P, Bridges H and Woolf (2000) Mark distributions and marking practices in UK higher education: some challenging issues, *Active Learning in Higher Education*, Vol 1 (1) 7-27

Appendix 1 - The Robert Gordon University grade related criteria

Grade	Definition of grade	Description of grade The characteristics of performance at each grade are given below. It is not expected that all items will be assessed in each individual assessment task but some exposure to each is expected at each level of the course/programme as stated in the level learning outcomes contained in the Definitive Course Document.
6	Excellent: Outstanding performance	Outstanding performance and achievement overall. The work of the candidate has much exceeded the threshold standard. The characteristics of work at this standard are: ● a thorough grasp of the **subject matter** ● a very high ability and originality in applying key **process skills** ● a very high ability in analysis, synthesis, evaluation and problem solving (**higher cognitive skills**) ● very high order ability over the specified range of **subject specific/professional practice skills**.
5	Commendable/ Very good: meritorious performance	A very high standard performance and achievement overall. The work of the candidate is well above the threshold standard. The characteristics of work at this standard are: ● a very good grasp of the **subject matter** ● a high ability and originality in applying **key process skills** ● a high ability in analysis, synthesis, evaluation and problem solving (**higher cognitive skills**) ● high order ability over the specified range of **subject specific/professional practice skills**.
4	Good: Highly competent performance	A highly competent performance and achievement overall. The work of the candidate has exceeded the threshold standard. The characteristics of work at this standard are: ● a good level of knowledge and understanding of the **subject matter** ● highly competent and displaying some originality in applying **key process skills** ● highly competent in analysis, synthesis, evaluation and problem solving (**higher cognitive skills**) ● a highly competent performance over the specified range of **subject specific/professional practice skills**.

3	Satisfactory: Competent performance	A satisfactory performance overall (as specified in the detailed marking/grading schemes for each assessment). The work of the candidate overall is at the threshold standard. The characteristics of work at this standard are: ● a satisfactory knowledge and understanding of the **subject matter** ● competence in applying **key process skills** ● adequacy in analysis, synthesis, evaluation and problem solving (**higher cognitive skills**) ● competence over the specified range of **subject specific/professional practice skills**.
2	Borderline fail: Failure open to compensation	A standard of performance overall which marginally fails to achieve competence. The work of the candidate overall is just below the threshold standard. Work just below the threshold standard is characterised by the candidate demonstrating: ● marginally unsatisfactory knowledge and understanding of the **subject matter** ● near competence in applying **key process skills** ● some evidence of ability in analysis, synthesis, evaluation and problem solving (**higher cognitive skills**) ● competence over most of the specified range of **subject specific/professional practice skills.** NB Safe working practice as appropriate to the discipline must be demonstrated for a student to be awarded this grade or higher grades.
1	Unsatisfactory, fail	The standard of performance demonstrated by the candidate overall is well below the threshold standard. Work in this grade is characterised by achievement of some of: ● very limited knowledge and/or understanding of the **subject matter**, exhibited in a very patchy manner ● limited/ occasional success in the application of **key process skills** ● occasional evidence of some of analysis, synthesis, evaluation and problem solving (**higher cognitive skills**) ● limited competence over the specified range of **subject specific/professional practice skills**.
0	Very unsatisfactory, abject fail	The work presented by the candidate is skeletal and/or irrelevant.

Subject matter: Assessed under headings knowledge, understanding (comprehension), application.

Key process skills: Assessed under headings communication and presentation skills, numeracy, information technology and computing skills, interactive and group skills, research skills.

Higher cognitive skills: Assessed under headings analysis, synthesis, evaluation, problem solving.

Subject specific/professional practice skills: Assessed under headings appropriate to the discipline.

For each assessment, staff should identify the relevance (and weighting) of each of the four areas. Within each area, the relevant source(s) of assessment criteria should be identified and criteria for each grade for each such source written to match the definitions.

Appendix 2 - The standards to be applied

A **top student (Grade 6)** would display a **very high level of presentation** as well as showing considerable evidence of **wide ranging research**, which would be accurately sourced and to which reference would be made in the text. In addition, the student would be able, working independently to display **a comprehensive knowledge and understanding** of the relevant theories, models and frameworks and be able to apply them consistently in a number of **complex** situations. The student would also display **detailed analysis** and be able to correctly **synthesise a large number of different ideas and concepts**. Finally, there would be considerable evidence of **critical analysis** and accurate **evaluation** on the basis of the wide- ranging evidence presented.

A **very good student (Grade 5)** would display some of the skills, knowledge and understanding of a top student and some of the same characteristics of an average or modal student. The above average student would be expected to perform consistently (and in a majority of the dimensions) at **a level above the modal student** in terms of **presentation**, level and **depth of research undertaken**, the comprehensiveness and depth of knowledge and understanding. In addition, the above average student would be expected to be able to **apply the concepts, theories and frameworks consistently in a number of areas largely as a result of independent study** and to display **consistent evidence of critical analysis and accurate evaluation in a significant number of instances**.

An **average or 'modal' student (Grade 4)** would display **good presentation skills** as well as showing **some evidence of research**. In addition the student would display **a good knowledge and understanding** of the relevant theories, models and frameworks and be able to **apply them in complex situations with some tutor guidance.** The student would also display **evidence of analysis** and be able to **synthesise correctly a number** of different ideas and concepts. Finally, there would be **some evidence of critical analysis and accurate evaluation** on the basis of evidence presented.

A **threshold student (Grade 3)** would display **competent presentation skills** as well as showing **some limited evidence of research**. In addition, the student would display **a knowledge and understanding** of the relevant theories, models and frameworks and be able **to apply them in simple situations with tutor guidance**. The student would also **display some evidence of analysis** but **limited evidence of critical analysis and accurate evaluation** on the basis of evidence presented.

A **marginal failed student (Grade 2)** while displaying some of the outcomes of a threshold student would fail to reach the minimum threshold in a number of significant dimensions.

A **failed student (Grade 1)** would display none of the above characteristics.

If there is no evidence on which to make an assessment a student will score zero.

Developing a variety of assessment methods, including self and peer-assessment - Post-workshop report

Professor David Lines, Centre for Enhancement of Learning and Teaching, The Robert Gordon University and Workshop Director

When reflecting on the reports written by the main speakers, as well as the feedback supplied by the breakout groups, it soon becomes apparent that academics in Scottish universities are facing real professional dilemmas and tensions in their teaching and the ways students are assessed. On the one hand, the literature increasingly emphasises the **necessity** to introduce radical approaches to assessment, while on the other a number of factors conspire to produce an environment that is **perceived** to prevent, or at least delay, change.

Readers may wonder why the word 'necessity' is emboldened in the paragraph above. The answer, as Chris Rust, Trudy Banta and Mark Wilson all say in their different ways, is that in order to ensure validity, a variety of assessment techniques must be used. If they are not, curriculum alignment cannot take place.

Certainly, if innovative teaching is being introduced - and there seems to be a lot less fear of 'bold' teaching than there is of 'bold' assessment - then for full alignment to occur assessment must also change. Otherwise, any benefits in terms of 'deep learning' that might be derived from this new style of teaching will be lost because students will respond strategically to the demands of the tests rather than what, and the way they have been taught.

Similarly, if we set new learning outcomes designed better to prepare students for the working world, or for lifelong learning, then the assessment methods associated with them must again change. It would seem self evident that if, for example, a learning outcome is defined as the ability to work in teams, then a two-hour essay simply cannot test such a skill. Yet, as one social science student reported in the recent Quality Assurance Agency for Higher Education (QAA) Scotland consultation exercise in answer to the question of whether the assessment methods they had experienced allowed students to demonstrate what they had learnt, one wrote: 'No. We are assessed by infrequent and stressful essays! There needs to be more formative assessment from the start of your university career' (Critical Thinking, 2004).

The simple logic of the connectivity between learning outcomes, teaching and assessment is hard to deny. The theory is well known and widely accepted. Alongside, there are drivers, such as improved staff development and training, which encourage teachers in higher education to think more deeply about their teaching techniques and how to relate them to learning outcomes. Taken together the theory and the practice would appear to represent an irresistible force, against which the bastions of assessment conservatism could not stand. But that would be to deny the apparent strength of such conservatism, as the workshop revealed, and this is where the professional dilemmas and tensions referred to above come in, for while outcomes and teaching can be altered and updated with relative ease, assessment, it would seem, for a variety of reasons explored below, cannot.

In his paper, Christ Rust offers six barriers to the introduction of innovative assessment techniques. They are:

- not convinced (personally)
- institution/colleagues/students not convinced
- tradition/inertia/no incentive to change
- time and/or other costs
- too difficult (need help/training)
- not allowed (regulations/procedures).

These six were echoed in the breakout groups, though often using different terminology. For example, the impact of the Research Assessment Exercise, the necessity to raise a personal research profile and the lack of kudos associated with teaching relative to research, were quoted as reasons for a lack of innovation. These could fall into Chris Rust's categories of the 'institution not being convinced' or 'time and/or other costs', but they could also be seen as more structural than those.

Certainly, structural problems were identified as barriers to innovation. Such barriers included modularisation, but external examiners, professional and regulatory bodies were also cited. These final three tended to be conservative in approach and often (regrettably this also included external examiners) somewhat ill-informed about the benefits that might be derived from new ideas on assessment. It was hoped that better training and closer working relationships might ease the situation.

The massification of higher education was also seen as a barrier, since it prevents individual lecturers from getting to know their students, and this in turn prevents the use and application of what are perceived as 'high risk' assessment strategies.

There is too, an even deeper structural issue, one which is highly political and which resonates in communities far removed from the education 'industry', and that is the need to hold education to account for its use of taxpayers' money. This I have characterised elsewhere (Lambert and Lines, 2000) as Assessment **of** Education as opposed to Assessment **for** Education. The latter is largely what this and all the other assessment workshops have been about and is characterised by formative assessment used in a timely and creative way to develop deep learning. In contrast, the need for, and use of, league tables emphasises public accountability and promotes the application of large-scale, high stakes summative tests which are seen as reliable (though often in practice, not) and 'safe'.

Yet as Trudi Banta makes clear in her paper, better assessment provides better information for **all** stakeholders. She cites Colorado State University, which has involved employers in the assessment process, and Eastern New Mexico University where every fine arts student has their portfolio or audition assessed each semester by a panel that includes staff, students, community representatives and other professional staff from outside the fine arts department. Such an eclectic engagement is crucial for a better understanding of what and why we assess and is needed to demonstrate to stakeholders, especially policymakers, that education, as in most walks of life, has moved on.

This challenge to perceptions was a theme taken up, not just in the breakout groups, but also by Win Hornby in his case study presentation. Here, he suggests, good (ie valid) assessment can drive both teaching and the teacher. He writes, '...the tutor's role could change from that of the 'primary' assessor to that of a moderator, thereby significantly improving both the efficiency and effectiveness of the assessment process'. The benefits of self and peer-assessment are therefore far wider than the initial impact on the students, though that should not be treated lightly. This was a view shared by the breakout groups and by Christ Rust in his paper.

One of the major benefits of self and peer-assessment is the speed and quality of the feedback, and it was feedback and feedforward that were taken up by Mark Wilson in his presentation. He argued that unless feedback was timely and informative, no one, neither student, nor teacher nor administrator, benefited from the assessment. Indeed, it was even possible that less capable students could end up practicing tasks that were incorrect, thereby reinforcing errors and undermining the foundations of learning.

It was the challenge to orthodoxy that became a recurrent theme at the workshop. Earlier in this paper I referred to the 'bastions of assessment conservatism' as if, given such mighty opposition, there was little a comparatively small band of innovators could do to overcome them. Yet once the opposition was explored in detail in the breakout groups, it proved remarkably vulnerable. For instance, the argument that institutions were not convinced by new assessment processes or would not allow them to be used, was simply untrue. No one could produce convincing evidence that a manager at any university had ever prevented innovation, though it was certainly the case that levels of support in its introduction were highly variable.

The argument that the regulatory bodies opposed change was demonstrably false given that the workshop itself was organised and funded by the Scottish Higher Education Funding Council and QAA Scotland. Professional bodies were also given short shrift by some, although others were less sanguine.

In the end, what it all seemed to come down to was motivation: if people wanted to innovate and use a variety of assessment techniques, there was nothing fundamental to stop them doing so. Barriers certainly exist, but many, if not most are over emphasised, perhaps as a let-out clause for those who know they **should** be doing things differently, but for whatever reason or reasons are unable to do so.

Yet it would be wrong to place the blame entirely on teaching colleagues for failing to change, facing as they are ever-larger classes and ever shrinking units of resource. There are things that can be done to help. For example, holding this and others workshops in the series, is one. The very notion of 'enhancement' sets a new and supportive tone.

At the level of the institution, the research/teaching tension remains and is likely to continue, but there are things that can be done to alleviate the imbalance. Most obviously the next Research Assessment Exercise could place greater weight on education-related research and publications (something the Higher Education Academy is strongly advocating), while promotions to professorships would also raise the profile of teaching and learning. Other incentives might be offered, such as prizes or small bursaries for innovations in teaching and learning, as well as extensions to existing Teaching Fellowship schemes.

One other important barrier exists to innovation in assessment, and that is the student body itself. Though they may not recognise it as such, many students come to university with a risk/reward equation in their minds. They will sacrifice a great deal in the expectation that success at University will return a better long-term reward than if they had not gone into higher education. Though this is a generalisation, they will have experienced, throughout their schooling, a number of high-stakes examinations in which they will have been largely successful. Such examinations, such as Scottish Highers and A levels, tend to be highly conventional, since they are externally set and marked by examination boards or assessment bodies. This style is what students are used to and what they are comparatively good at. As if to confirm this, the majority of students in the QAA Scotland survey answered 'no' to the question as to whether the way they are currently assessed might be improved, though such a view might simply reflect the fact that they have not experienced anything like the range of innovative techniques outlined in the pre-workshop material and so are not in a position to make a judgement.

For students, stepping outside a conventional assessment pattern creates a risk-intensive environment. For them the cost of failure is high and this is why they often resist change. Yet many admit to 'hating exams' and almost all would agree that a variety of assessment tasks would better test their overall knowledge, understanding and skill acquisition. As the survey showed, feedback and formative assessment are recognised by students as crucial to progress in summative settings as well as personal intellectual development and they bemoan the lack of both.

Though the survey analysed by Critical Thinking for QAA Scotland provides only indicative information, it is clear from the report that students are not in any systematic way being admitted into the assessment process. The irony of such exclusion should be lost on no one. In the end, it is the students who are most affected by assessment methods and methodologies and while the concept of the student as 'customer' is a contested one ('partners' is perhaps a better choice), students are without doubt central stakeholders in the higher education process. If the notion of partnership were developed to create what has been called a 'Powerful Learning Environment' (Direick and Dochy, 2001), students and their teachers (perhaps better called 'mentors' or 'facilitators' under this new system) would work together to help define learning outcomes, methods of acquiring the knowledge and skills that would best meet those outcomes and the assessment that would confirm that learning had taken place.

Cynics will argue that such an idea is pie-in-the-sky, but as Mark Wilson so vividly describes it, such a response is part of a 'pass-the-buck-culture' in which failure is seen as entirely the student's fault and has nothing to do with the instructor, the teaching or the methods of assessment used. It is time to move away from such a 'them and us' relationship. Once a dialogue has taken place between staff, students and university management and thence on to employers and the wider community, concerns over using and applying innovative processes in teaching, learning and assessment will evaporate. It is apparent from this workshop that the will exists to try new approaches, and furthermore, that where tried there is a beneficial impact on all involved: it is clearly time to embrace change.

References

Critical Thinking (2004) *Student Perspectives on Assessment. An Analysis of the Responses to the Consultation Exercise*, Quality Assurance Agency for Higher Education, Gloucester

Direick S and Dochy F (2001) New Lines in Edumetrics: New Forms of Assessment Lead to New Assessment Criteria, *Studies in Educational Evaluation, 27* (4) 307-329

Lambert D and Lines D (2000) *Understanding Assessment, Purposes, Perceptions, Practice,* RoutledgeFalmer, London