Assessment Series No.3

Assessment: A Guide for Lecturers

George Brown

LTSN Generic Centre
Assessment: A Guide for Lecturers
November 2001

George Brown is a retired professor from the University of Nottingham. He has been actively engaged in research in higher and medical education for the past 25 years. He has also worked with professional engineers, dentists and nurses on communication issues. He was the National Co-ordinator of Academic Staff Development for the Committee of Vice-Chancellors and Principals (CVCP, now Universities UK) and he contributed to the design of the training programmes for Subject Reviewers for the Higher Education Funding Council for England (HEFCE) and subsequently, the Quality Assurance Agency (QAA). His most recent text is *Assessing Student Learning in Higher Education* (Routledge, 1997; with Joanna Bull and Malcolm Pendlebury). His most recent article is *Refreshing Lecturing* (Medical Teacher, 2001) and he has recently completed a chapter on 'Culture, Collegiality and Collaborative Learning' for a text to be published by Kogan Page.

Contents

Generic Centre Guides and Briefings

Welcome to the Learning and Teaching Support Network Generic Centre's series of Assessment Guides and Briefings. They aim to provide a series of overviews of important issues and practices in the field of assessment for the higher education community.

The Assessment Guides are intended for colleagues with particular roles and for students, as their titles suggest. The Briefings are primarily intended for lecturers and other staff involved in supporting learning.

The Assessment Series is a snapshot of a field in which development is likely to be rapid, and will be supplemented by specific case studies produced by the LTSN Subject Centres.

The series was developed by Brenda Smith and Richard Blackwell of the LTSN Generic Centre with the support of Professor Mantz Yorke. Experts in the field were commissioned for each title to ensure that the series would be authoritative. Authors were invited to approach the issue in their own way and no attempt was made to impose a uniform template.

The series editors are grateful to colleagues in LTSN Subject Centres and other senior colleagues who refereed the series, and of course to the authors for enabling its publication.

We hope that you will enjoy the Assessment Series and find it interesting and thought-provoking. We welcome your feedback and any suggestions you may have for future work in the area of assessment.

Professor Brenda Smith
Head, LTSN Generic Centre

Richard Blackwell, Senior Adviser,
LTSN Generic Centre

Professor Mantz Yorke,
Liverpool John Moores University

November 2001

Summary

1.
This guide is designed to help lecturers to review, refresh and refine approaches to assessment. It is based upon the theme of alignment between intended learning outcomes, assessment tasks, criteria, marking and providing feedback. Alignment provides the basis for an effective assessment system and for the new approach to Academic Review by the Quality Assurance Agency.

2 .
The purposes of assessment are to give a licence to proceed to the next stage or to graduation; to classify the performance of students in rank order; to improve their learning. Differences in purpose should influence the choice of assessment task.

3.
The key principles of effective assessment and the common weaknesses of assessment systems are primarily concerned with linkages between outcomes, the design of assessment tasks, criteria, marking procedures and feedback. Hints and suggestions on these components and links are provided.

4.
There is a rich diversity of methods of assessment and increasingly postgraduates, work supervisors, peer groups and students themselves are used as sources of assessment. There is a wide variety of criteria. The better ones are simple and easy to understand and use. All sources should participate in training based on the criteria being used.

5.
Suggestions for improving the efficiency of marking and moderating, providing feedback to students and assessing more students are provided together with hints on minimizing plagiarism and on assessing students with learning difficulties and disabilities.

6.
The pros and cons of examinations and course work are discussed. It is suggested that there should be a balance of different types of assessment over a degree programme. The point of balance should be determined primarily by the learning outcomes and the resources available.

7.
The role of the external examiner is an extension of the role of the lecturer, as assessor. It consists of guardian of standards, protector of students and agent of quality. Training and briefing in all aspects of this role are recommended.

Introduction

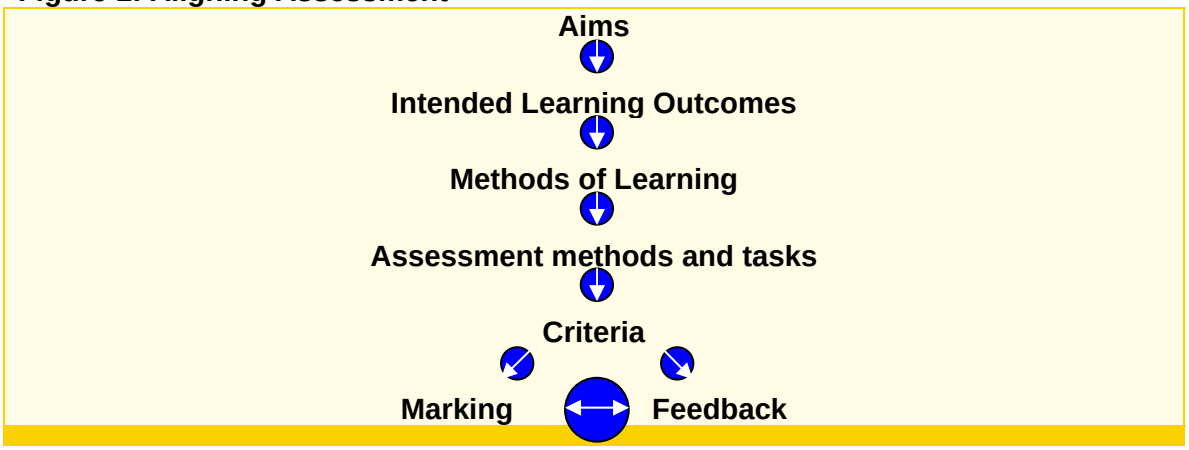
There is no doubt about the importance of assessment. Assessment defines what students regard as important, how they spend their time and how they come to see themselves as students and then as graduates. It is a major concern of those who learn, those who teach and those who are responsible for the development and accreditation of courses. Hence this guide which, together with the LTSN Generic Centre's companion guides for students and managers, provides a broad overview of assessment.

The guide is designed to help you to review, refresh and refine your approaches to assessment. Its main theme is alignment, that is the importance of alignment between assessment methods, assessment tasks, learning opportunities and intended learning outcomes (learning objectives) and between feedback, criteria and assessment tasks. The guide outlines the path of assessment from its purposes and principles to the role of external examiners. Along the way it considers methods, sources and instruments, marking, providing feedback and the problems of plagiarism. It contains hints, suggestions and opportunities to reflect upon approaches to assessment but it does not review the relevant research on assessment nor does it provide specific procedures for designing specific assessment tasks. For those aspects of assessment, the reader is directed to the brief annotated bibliography.

Why is alignment important?

The alignment of assessment with other features of a course is the basis of course design and central to effective assessment. The model shown in Figure 1 summarises the relationships. If the aims are unclear then the system crumbles. Clear and realistic outcomes provide students with a good guide to what has to be learnt and lecturers with a guide on how to teach and what learning opportunities to provide. Relevant learning opportunities help students to achieve the learning outcomes. Effective assessment methods and tasks are related to the learning outcomes and the methods of learning. If the criteria are too fuzzy or unduly specific then it is difficult for assessors to ensure consistency of judgement and for students to fulfil the demands of the assessment task. Without close links between feedback, criteria and the assessment tasks, lecturers cannot help students to achieve the learning outcomes of a course or a programme.

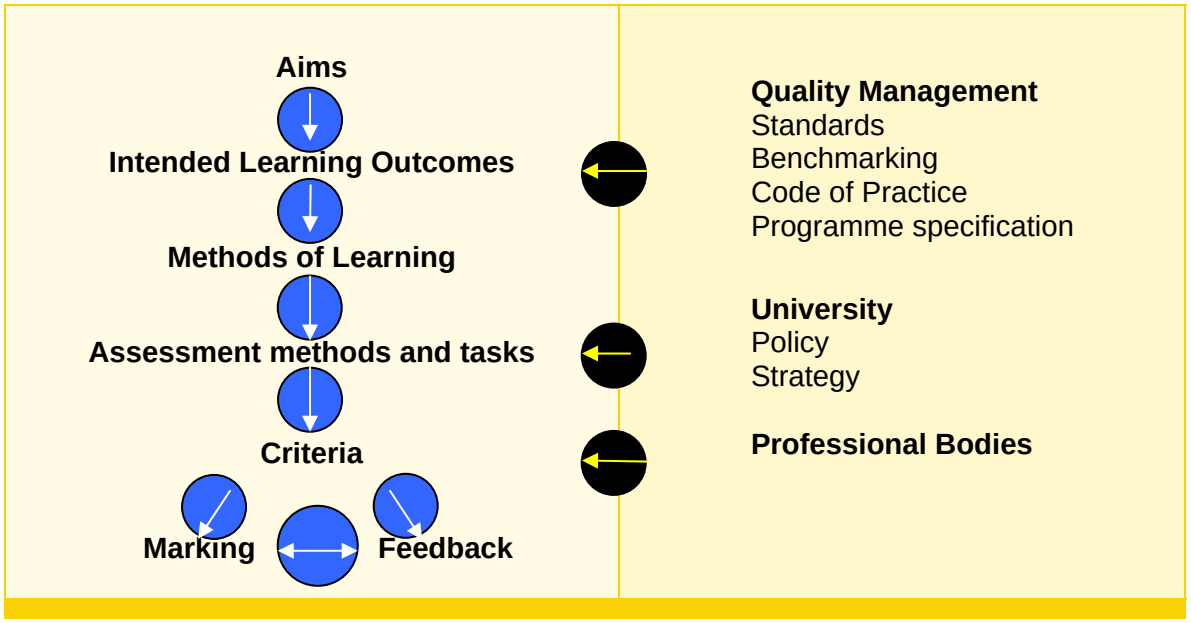
Figure 1: Aligning Assessment



However, the model is not as simple as it appears. It leaves open the questions of whether the outcomes that are being assessed should be those of the degree programme or those of the module (or short course) and whether every learning outcome should be assessed. There are arguments for and against the use of programme or module outcomes as the basis for assessment tasks. If one opts for assessing every outcome of every module then one runs the risk of over-assessing students – and UK students are probably the most overassessed students in Europe. If one opts for programme outcomes one risks not assessing essential knowledge and skills but one has a framework for estimating student progression and achievement. On balance, the better strategy is to ensure that within each module, teaching and learning opportunities are provided which move the students closer to the programme outcomes and that some programme outcomes are assessed in some of the modules so that by the end of the programme all the outcomes have been assessed at each level on at least two occasions. This approach ensures that one has repeated and therefore probably more reliable measures of achievement, and a realistic, not unduly burdensome approach to assessment. A matrix of the learning outcomes of a programme and the assessments used in different modules helps one to identify the links between programme outcomes and their assessment.

An expanded version of the model is shown in Figure 2. It shows the connections between the model and the Quality Assurance Agency (QAA) approach to evaluation. Whilst some of us may have reservations about the micromanagement techniques of the QAA, the model does facilitate the fulfilment of its requirements. The approach of the QAA is authoritarian: it is about X doing things to Y and Z judging them. But within its straitjacket, it is possible to build in some freedom to learn.

Figure 2: External influences on assessment



In summary, the model provides a basis for effective assessment, it demonstrates that course design and effective assessment are inseparable

The purposes of assessment

The three main purposes of assessment are:

- to give a licence to proceed to the next stage or to graduation;
- to classify the performance of students in rank order;
- to improve their learning.

These purposes may overlap or conflict. A common error is to use an assessment task for one set of purposes and then assume that the results from it are appropriate for other purposes. For example, reflective diaries or records of achievement can be useful during the process of learning but if they are used to make judgements or recommendations for employment then students may be reluctant to report honestly their thoughts and feelings. When 'licence to proceed' is a key concern, the assessment tasks should be based on core knowledge and skills and the pass/fail threshold should be absolutely secure. If the primary purpose of assessment is to place the students in rank order, the assessment tasks should be designed to differentiate the capabilities of a wide range of students.

Purposes are related to summative and formative assessment. Summative assessment contributes to the marks for a module, level or degree. Formative assessment provides feedback to students during the course so they have opportunities to improve. Clearly formative assessment overlaps with feedback in learning. Course work assessment is usually both summative and formative. It is supposed to give feedback and count towards the final profile of marks. When this happens, the in-course assessment becomes part of a multiple point summative assessment. Heavy assessment loads and other commitments may tempt us to provide perfunctory feedback.

Some principles of assessment

The key principles of assessment may be derived from the model in Figure 1. They are:

- assessment shapes learning so if you want to change learning then change the assessment method;
- match the assessment tasks to the learning outcomes;
- match the criteria to the task and learning outcomes;
- keep the criteria simple;
- be fair, reliable and valid in your marking;
- provide meaningful, timely feedback.

Do you apply these principles?

Common weaknesses

Common weaknesses in assessment are:

• the tasks do not match the stated outcomes;

• the criteria do not match the tasks or outcomes;

• the criteria are not known to students;

• students do not understand the criteria;

• overuse of one mode of assessment such as written examinations, essays, or closed problems;

• overload of students and staff;

• insufficient time for students to do the assignments;

• too many assignments with the same deadline;

• insufficient time for staff to mark the assignments or examinations;

• absence of well defined criteria so consistency is difficult to achieve;

• unduly specific criteria which create a straitjacket for students and make marking burdensome for lecturers;

• inadequate or superficial feedback provided to students;

• wide variations in marking between modules and assessors and within assessors (selfconsistency);

• variations in assessment demands of different modules.

By implication, more effective assessment systems are relatively free of these weaknesses. *Are yours?*

Designing assessments

Seven questions that lecturers might ask when designing an assignment or written paper are:

1. What are the outcomes to be assessed?
2. What are the capabilities/skills (implicit or explicit) in the outcomes?
3. Is the method of assessment chosen consonant with the outcomes and skills?
4. Is the method relatively efficient in terms of student time and staff time?
5. What alternatives are there? What are their advantages and disadvantages?
6. Does the specific assessment task match the outcomes and skills?
7. Are the marking schemes or criteria appropriate?

The above questions naturally lead in to the design of a specific assessment task. Here there are several pitfalls. Notable amongst these are *unintentional* ambiguities in a question or assignment, under-estimation of time and resources required to do the assignment or to mark it, and neglect of a suitable set of criteria or a marking scheme.

The design of *effective* assessment tasks can be time-consuming. A useful starting point is to look through and note examination questions and assignments that are set in comparable courses in other institutions. Then look at questions set in other subjects in your own institution. You will almost certainly find questions that are intriguing and surprising and you will also find forms of questions that may be transferred into your own subject. Occasionally, you may find the content as well as the form of question is relevant to your own subject. Whilst reading and noting the questions, try to work out what the assessor was trying to assess. (Some examples of examination questions in different subjects are given in Brown, Bull and Pendlebury, 1997.) Return now to the questions in your own subject. Consider what kinds of learning you are trying to develop and assess. Look at the outcomes of your course and think about what kinds of things you want your students to learn and to be able to do.

Alternatively, you can develop some different approaches, new forms of questions and tasks and discuss them with colleagues. Be prepared to provide a rationale for their use. Some people find brainstorming with colleagues is a useful way of generating a series of questions on a topic. Others keep a note-book of possible questions, quotes and problems. Figure 5 offers some approaches that you may not have used.

It is usually better to include new types of questions in course work rather than examinations and new methods of assessment with level one students. Be sure that the wording of the questions is clear. Eliminate any unwanted ambiguities in the questions. Bear in mind that the more open or ambiguous a question is, the harder it is to devise a marking scheme for it, but one can often use broad criteria. If you run a pilot on the assessment tasks, skim read the answers and then devise criteria or a marking scheme or adapt an existing approach. Then test the criteria by marking a few assignments.

Cognitive demands

A knowledge of different types of cognitive demand is an essential ingredient of designing and marking assignments and examination scripts. A useful approach is provided in Figure 3. Paradoxically, the lower three levels are related to 'surface' learning and the higher levels to 'deep' learning. However, one has to use the guide in the context of the course and the capabilities of the students. What may require synthesis and evaluation by a level one student may be routine recall for a final-year student.

Figure 3: Hierarchy of the Cognitive Domain

6	**Evaluation**	Ability to make a judgement of the worth of something
5	**Synthesis**	Ability to combine separate elements into a whole
4	**Analysis**	Ability to break a problem into its constituent parts and establish the relationships between each one.
3	**Application**	Ability to apply rephrased knowledge in novel situation
2	**Manipulation**	Ability to rephrase knowledge
1	**Knowledge**	That which can be recalled

Based on Bloom's Taxonomy of Educational Objectives (Bloom, 1956). Higher levels of the taxonomy subsume lower levels.

An alternative approach is provided by Biggs (Biggs, 1997; Biggs, 1999). He describes how his taxonomy, SOLO (Structure of the Observed Learning Outcome), may be used for analysing learning tasks, hierarchies of objectives and for assessing students' work.

The five levels of Biggs' schema are:

1. *Pre-structural:* The task is not attacked appropriately. The student has not understood the point.

2. *Uni-structural:* One or a few aspects of the task picked up or used but understanding is nominal.

3. *Multi-structural:* Several aspects of the task are learnt but they are treated separately.

4. *Relational:* The components are integrated into a coherent whole with each part contributing to the overall meaning.

5. *Extended abstract:* The components are integrated and re-conceptualised, thereby creating an individual perspective.

The above taxonomies, again context specific, are more useful for classifying cognitive demands in assignments based on discourse. Other taxonomies on problem-solving, practical work and values are described in Brown, Bull and Pendlebury (1997) and Heywood (2000).

Methods, sources and instruments of assessment

At the outset, it may be useful to distinguish between the three terms, methods of assessment, sources and instruments. Method refers to the approach used to assess learning such as essays, problems, multiple choice tests and so on. Source refers to the person such as the lecturer, student, employer or peer group of students. Instrument refers to the marking scheme, the explicit criteria or implicit criteria. Strictly speaking, the instrument of assessment is the source *in conjunction* with the instrument. Figure 4 outlines a range of methods, sources and instruments. Some of the controversy surrounding new approaches to assessment are due to confusions about sources and instruments.

Figure 4: Methods of Assessment, Sources and Instruments

Examples of Methods	Sources	Instruments
Essays	Lecturer i/c*	*Implicit criteria*
Problem sheets	Other tutors	Global
Unseen proses	Postgraduate tutors	*Explicit criteria*
Lab reports	Demonstrators	Criteria reference grading
Presentations	Student self	Specific criteria
Projects	Student peers	Marking scheme
Group projects	Employers	Dimensions
Posters	Mentors	Rating schedules
Work-based learning		Checklists

**i/c refers to lecturer in charge of module or course*

Self assessment and peer assessment are not methods of assessment but sources of assessment which could be used with any method or instrument. Thus peer and self assessment of a group project may be a useful way of encouraging students to reflect upon their team-work and the product, the project report. Explicit criteria should be used to ensure consistency and fairness. On the other hand, final-year marks based solely on self assessment and implicit criteria would, for most of us, seem a totally inappropriate practice. The decisions one has to take are whether the combination of source, instrument and method are appropriate for the outcomes and whether the costs, in terms of time and other resources, merit the approach.

Methods of assessment

A summary of assessment methods is set out in Figure 5. A brief comment on their uses, ease of use, potential validity and reliability is provided. Of these methods, essays, problemsolving, reports on practicals and projects or dissertations are the most common. Methods based on peers, portfolios and other methods of assessing reflective learning are increasingly used. The closer a method is to recall of knowledge and/or well-defined solutions, the more reliable is the method but not necessarily the more valid. One has, as usual, to manage the conflicting demands of validity and reliability, of effectiveness (whether it is a good approach to assessing the learning outcomes) and efficiency (the time taken for students to do the assessment task and for assessors to mark it). The effectiveness of the assessment method depends upon the learning outcomes being assessed and the particular assignment set – not on the method of assessment *per se*.

Within each method, there can be various approaches and cognitive demands. Instead of conventional essays, one can set tasks such as writing book reviews, critiques of research papers or articles for a broadsheet. Critical thinking can be assessed by using multiple choice questions (MCQs) or short answer questions based on a research paper or case.

Ideally, students should experience a wide range of methods of assessment in a degree programme. As indicated earlier, a matrix of programme outcomes versus assessment methods used in each module will reveal the pattern and any lacunae. Programme outcomes need not be assessed in every module but it is sensible to ensure that over the whole programme all outcomes are taught, practised and assessed. Further details of the methods of assessment described in Figure 5 may be found in the annotated bibliography.

Figure 5: Which methods of assessment do you use?

Cases and open problems	Have potential for measuring application of knowledge, analysis, problem-solving and evaluative skills. Short cases are relatively easy to design and mark. Design of more complex cases and their marking schemes are more challenging to design and develop. Marking for grading and feedback are about as fast as essay marking.
Computer-based assessment	Much talked about. Usually software such as Question Mark will be used to format multiple choice questions, mark and analyse results. Wider range of graphics and simulations can be used. Optical Mark readers can be used - but allow for some students not marking the items clearly. Time consuming to set but marking very fast. Reliability is high but validity (match with outcomes) needs careful attention.
Direct Observation	Useful for immediate feedback, for developmental purposes and for estimating performance -providing a simple, structured system is used. The presence of the observer can change the performance so the method should be handled sensitively. Impressionistic observation can be useful if supported by constructive feedback. Can be used by a group of peers to provide feedback as well as assessment. Intensive, lengthy training is required for high reliability if detailed checklists are used. Reliability, validity and manageability are fairly high when structured observation is used.
Essays	A standard method. There are several types of essays that test different styles of writing types of thinking. Measures understanding, synthesis and evaluation, providing you ask the right questions. Relatively easy to set. Marking for grading based on impressionistic marking is fast. Marking for feedback can be time-consuming. Keep the criteria simple. Variations between assessors can be high - and so can variations of the Assessor.
Learning logs/ diaries	Wide variety of formats ranging from an unstructured account of each day to a structured form based on tasks. Some training in reflection recommended. Time-consuming for students. Requires a high level of trust between assessors and students. Measuring reliability is difficult. May have high validity if structure matches learning outcomes.
Mini-practicals	A series of mini-practicals undertaken under timed conditions. Potential for sampling wide range of practical, analytical and interpretative skills. Initial design is time-consuming. Some if not all of the marking can be done on the spot so it is fast. Feedback to students is fast. Reliable but training of assessors is necessary.

Modified Essay Questions (MEQs)	A sequence of questions based on a case study. After students have answered one question, further information and a question are given. The procedure continues, usually for about one hour. Relatively easy to set. May be used in teaching or assessment for developmental or judgmental purposes. Can be computer - or paper-based. Can encourage reflection and analysis. Potentially high reliability, validity and manageability.
Multiple Choice Questions (MCQs)	A standard method. Can sample a wide range of knowledge quickly. Has potential for measuring understanding, analysis, problem solving skills and evaluative skills. Wide variety of formats from true/false to reasonassertion. More complex formats not recommended: they confuse students unnecessarily and they are time-consuming to design. More demanding MCQs require more time to set. Better ones are based on case studies or research papers. Easy to mark and analyse results. Useful for self assessment and screening. Potentially high reliability, validity and manageability. Feedback to students is fast. Danger of testing only trivial knowledge. To save time, look for banks of items on the Net or in US text books. A team of assessors, working to the same learning outcomes, can brainstorm and produce several questions in an afternoon.
Orals	Tests communication, understanding, capacity to think quickly under pressure and knowledge of procedures. Feedback potential. Marking for grading can be fast but some standardisation of interview procedure is needed to ensure reliability and validity.
Objective Structured Clinical Examinations (OSCEs)	Initially used in medicine but can be used in business, legal practice, management, psychology, science courses and social work. Particularly useful for assessing quickly practical and communication skills. Fairly hard to design and organise, easy to score and provide feedback. Could be used in induction phase to estimate key practical skills. Group OSCEs useful for teaching, feedback and developmental purposes. OSCEs can be used towards the end of a course to provide feedback or to test performance against outcomes. Reliability, validity and manageability are potentially fairly high. Probably less labour intensive than other forms of marking but several assessors required at one time. Initially, they are timeconsuming to design - but worth the effort.
Portfolios	Wide variety of types from a collection of assignments to reflection upon critical incidents. The latter are probably the most useful for developmental purposes. May be the basis for orals. Rich potential for developing reflective learning if students trained in these techniques. Require a high level of trust between assessors and students. Measuring reliability is difficult. May be high on validity if structure matches objectives of training.

Poster sessions	Tests capacity to present findings and interpretations succinctly and attractively. Danger of focusing unduly on presentation methods can be avoided by the use of simple criteria. Feedback potential: from tutor, self and peers. Marking for grading is fast. Use of criteria reduces variability.
Presentations	Tests preparation, understanding, knowledge, capacity to structure, information and oral communication skills. Feedback potential: from tutor, self and peers. Marking for grading based on simple criteria is fast and potentially reliable. Measures of ability to respond to questions and manage discussion could be included.
Problems	A standard method. Has potential for measuring application, analysis and problem solving strategies. Complex problems and their marking schemes can be difficult to design. Marking for grading of easy problems is fast. Marking of complex problems can be slow. Marking for feedback can be slow. Variation between markers is fairly low when based on model answers or marking schemes Allow for creative, valid solutions by bright students.
Projects, Group Projects and Dissertations	Good all-roundability testing. Potential for sampling wide range of practical, analytical and interpretative skills. Wider application of knowledge, understanding and skills to real/simulated situations. Provides a measure of project and time management. Group projects can provide a measure of teamwork skills and leadership. Motivation & teamwork *can* be high. Marking for grading can be time-consuming. Marking for feedback can be reduced through peer and self-assessment and presentations. Learning gains can be high particularly if reflective learning is part of the criteria. Tests methods and processes as well as end results. Variations between markers possible. Use of criteria reduces variability but variations of challenge of project or dissertation can affect reliability.
Questionnaires and report forms	A general method including a wide variety of types. Structured questionnaires get the information you want but semi or open-ended questionnaires may give you the information that you need. A mixture of structured and open-ended questions is recommended. Criterion reference grading recommended for judgmental purposes. Broad criteria are more reliable and valid than highly detailed criteria. Detailed criteria tempt users to react negatively or disdainfully.
Reflective Practice Assignments	Measures capacity to analyse and evaluate experience in the light of theories and research evidence. Relatively easy to set. Feedback potential from peers, self and tutors. Marking for feedback can be slow. Marking for grading is about the same for essays. Use of criteria reduces variability.

Reports on Practicals	A standard method. Have potential for measuring knowledge of experimental procedures, analysis and interpretation of results. Measure know how of practical skills but *not* the skills themselves. Marking for grading using impressions or simple structured forms is relatively fast. Marking for feedback with simple structured forms is faster than without them. Variations between markers, without structured forms, can be high. Method is often over-used. To reduce student workload and the assessment load, different foci of assessment for different experiments recommended.
Self-assessed questions based on open learning(distance learning materials and computer-based approaches)	Strictly speaking, a method of learning not of assessment. But could be used more widely. *Self assessed questions* could form an integral part of Open Learning. These could be based on checklists, MCQs, short answer questions, MEQs and other methods. Their primary purpose is to provide feedback and guidance to the users. They can be used to integrate open learning and work-based learning when students are on placement. Reliability and validity is probably moderately high and manageability is high, in the long term, but low initially.
Short answer questions	A standard method. Has potential for measuring analysis, application of knowledge, problem-solving and evaluative skills. Easier to design than complex MCQs but still relatively slow. Marking to model answers is relatively fast compared with marking problems but not compared with MCQs. Marking for feedback can be relatively fast .
Simulated interviews	Useful for assessing oral communication skills and for developing ways of giving and receiving feedback on performance. Video-recorded sessions take more time but are more useful for feedback and assessment. Peer and self assessment can be used. Sensitive oral feedback on performance is advisable. Assessment by simple rating schedule or checklist is potentially reliable if assessors, including students, are trained.
Single Essay Examination	Three hours on prepared topic. Relatively easy to set but attention to criteria needed. Wider range of ability tested including capacity to draw on a wide range of knowledge, to synthesise and identify recurrent themes. Marking for feedback is relatively slow. Marking for grading is relatively fast providing the criteria are simple.
Work based Assessment	Variety of methods possible including learning logs, portfolios, projects, structured reports from supervisors or mentors. Important to provide supervisors and mentors training in the use of criteria. Work experiences can be variable so reliability can be low. Validity, as usual, is dependent upon clear learning outcomes.

Sources of assessment

The main source of assessment is the lecturer although, increasingly, other sources are being used. Hounsell and Thompson (1995) provide practical suggestions for postgraduate tutors and demonstrators. Ashworth and Saxton (1992) provide examples of guidance and assessment of work-based learning and the use of mentors (supervisors) and Brennan and Little (1996) review theories and practices of workbased learning and the role of assessment. Whatever the source of assessment, it is important to ensure consistency between assessors and within assessors (selfconsistency). Brief training sessions, based on the criteria to be used, do improve consistency but do not expect perfection.

Instruments of assessment

Instruments of assessment vary from the use of an holistic approach to very detailed checklists. The main types of criteria are shown in Figure 6: The use of criteria is not as easy as it appears. Not all criteria can be articulated or communicated clearly to novices and the meanings of criteria are context-dependent.

Figure 6: Types of criteria

Intuitive: implicit criteria. Hidden from other markers and students.
Global: based on key features such as organisation, evidence of reading. The assessment leads directly to a single mark. Marking can be fast and reliability high. Feedback to students can be slow.
Criterion reference grading: general criteria for grading students work. Usually based on degree categories. Marking can be fast and feedback to students is fairly fast.
Broad criteria: based on ratings or marks. Used to assess qualities that permeate the whole of an assessment task such as fluency of style or organisation. Usually reliable and feedback can be fast if based on the criteria.
Specific criteria: more detailed than broad criteria. Often overlap and meanings unclear. e.g. What is the difference between structure and organisation? Can be burdensome to use, variations between markers on specific criteria can be low, feedback to students can be fast but not necessarily useful. Numbers on a scale do not tell a student how to improve.
Marking schemes: often used for linear marking such as specific subject content, operations or procedures such as the application of a standard integral, the correct use of the past pluperfect or an accurate translation of a paragraph in a passage of prose. Can be slow if errors have consequential effects so choose, if possible, problems that have only a few pathways. Usually reliable and valid.
Checklists: can be used to assess sequential tasks and simple design specifications. Timeconsuming for assessing complex tasks. Can be reliable.
Detailed checklists: burdensome to use. Not necessarily helpful to students. Intensive training required to ensure reliability.
Detailed criteria: probably the least reliable method and most time-consuming instrument of assessment.

‘Well structured’ may have different meanings in Biological Sciences, English Literature and History and what counts as ‘well argued’ may have different meanings for students at different levels in a programme. Marks on rating scales do not necessarily convey meaning and the written feedback based on the criteria may be interpreted differently by students of different levels of ability. Criteria or marking schemes can change a student’s approach to learning, to completing assignments, to revising and to their approach in examinations. If the published marking scheme or criteria are concerned only with specific knowledge the students may neglect structure, style and syntax. These then can operate as hidden criteria. Multiple Choice Questions (MCQs) based on negative marking penalise students, usually females, who may be unwilling to take risks. Beware of criteria that can unnecessarily exacerbate gender differences.

All of the above points suggest that it is useful to have discussions with students and colleagues about different types of criteria and marking schemes, what they might mean and how they should be used. These guidelines may help your discussions.

The key principles are:

• decide on the essential criteria;

• make the criteria or checklist simple to use;

• allow for brief global impressions;

• give the criteria to the students before they do the assignment;

• if possible, involve them in the design of the criteria and checklist;

• encourage students to use the criteria. The essentials of good criteria are that they:

• match the assessment task and learning outcome;

• enable consistency of marking;

• can pinpoint areas of disagreement between assessors;

• help students to achieve the learning outcomes;

• be used to provide useful feedback to students.

Do your criteria have these characteristics?

Marking and moderating

At the root of the process of marking and moderating are questions of validity, feasibility and reliability. Validity, in educational assessment, is a matter of judgement: do the criteria match the assessment task and the intended learning outcomes? Feasibility is concerned with what is practical, given the time and resources available.

Reliability is concerned with consistency of marking against the criteria provided. It ensures that students are marked fairly. For educational assessment, self-consistency is more important than inter-assessor consistency. If self-consistency is high, one can adjust marks. If it is low than no amount of statistical manipulation can improve reliability. Self consistency is challenging. One needs to know one's values, prejudices and the criteria, to mark by question, not script, and, here is the rub: re-mark if one is

not consistent. The moderator's task is to check the selfconsistency of the assessor. This is more important than whether first or second markers are in agreement.
The more complex or open an assessment task is, the more difficult it is to obtain close agreement between assessors. This well known finding is cold comfort to assessors, external examiners and the QAA. Fortunately, inter-assessor consistency can be improved by using simple criteria and trial marking by the module or course team. The use of broader grades increases agreement on marks – but not necessarily on reasons. Other methods of ensuring consistency are more demanding and not necessarily useful. Double blind marking doubles the administrative and assessment load yet evidence from studies at secondary level (e.g. Murphy, 1979; Newton, 1996) indicate that it is no more reliable than single marking and moderating based on borderline and central samples for each grade. The moderating process should be concerned with verification, not with re-marking. When two markers disagree, it is customary to use a third marker. The third marker may be tempted to compromise rather than apply the criteria strictly.

Recently the notion of returning marked examination scripts to students so they can obtain feedback and improve has been discussed. This will be a very time-consuming activity. It assumes that students will benefit from the feedback and their learning will improve. Before embarking on this path, it would be prudent to run a series of well controlled experiments to check the costs and benefits of the procedure for lecturers and students.

Anonymous marking of scripts and assignments is increasingly fashionable. The approach may reduce some of the biases attributable to the assessor's perception of a student but will not reduce biases due to differences in values. Anonymity is difficult to maintain in small departments and it may weaken the overall effect of assessment on improving students' learning and motivation. Personal contact still has an important role in higher education. How consistent is your marking and that of your colleagues? How do you know?

Feedback to students

The purposes of feedback are to motivate students, to inform them how well they have done and how to improve. The improvements might be to the knowledge-base, understanding, problem-solving strategies and writing skills. These purposes should be set in the context of the learning outcomes, what students are expected to do in the assignment and the criteria that are being used.

Approaches to providing feedback vary from ‘proof reader’ to Grendel’s Mother. Nowadays these excesses are curtailed by the use of feedback forms, although some forms are cumbersome to use and they are not always helpful to students. *What are yours like?* Feedback has been shown to be most effective when it is timely, perceived as relevant, meaningful, encouraging and offers suggestions for improvement that are within a student’s grasp. All the practical hints on providing feedback can be deduced from these findings and there are implications underlying these findings for the management of modules, one’s own time management and approach to providing feedback.

Unfortunately, the feedback provided is not always read, let alone used by students. So it may be necessary to teach students how to use feedback and to check that they have used feedback from their previous assignments. Sometimes the feedback is not available when students need it. Some departments and schools keep all assignments lest the QAA subject or academic reviewers wish to inspect them. This approach defeats the purpose of feedback to students. The costs and benefits of alternative strategies should be considered by senior management and the QAA.

Assessing large numbers of students

The greater part of the time for assessing students is spent on assessing for feedback or for QAA requirements. The strategies that are available for managing and saving time, yet providing useful feedback to students, are essentially:

- reduce the assessment load;
- streamline the feedback procedure;
- delegate the marking;
- review the modular system.

Reduce the assessment load

You can set fewer or shorter assignments. Ensure the workload is uniform across the set of modules offered by the department. A check on the total amount of course work required from students can be revealing. You can set tests or assignments to be completed and/or marked during class time. Capitalise on Information Communications Technology (C&IT) by using MCQs, requiring assignments to be word-processed and for processing, collating and recording marks.

Streamline the feedback procedure

To ease time management, the lecturer can at the beginning of the module set the date for returning marked assignments to students. At the same time, allocate days or parts of days for marking in your diary. Keep to these allocations unless there is a dire emergency. Say 'no' to colleagues and students who wish to encroach on this time. A most useful strategy for improving learning is to mark and comment on the draft of an assignment. Keep a copy of your comments. Grade the final version and add a brief global comment. This approach is more useful for learning than a thorough marking of the final version of an assignment. It demonstrates to students that they can improve. Once an assignment has been marked, it tends to be archived by the students and often there is little transfer of learning to the next assignment.

Use tutorials only for students most in need. Use criteria or checklists plus a personal comment or use a standard pro-forma. Keep them simple otherwise you will spend more, not less, time fitting your comments to them. Alternatively do a global report based on the students' assignments. 'The best assignments had these characteristics ... good assignments ... weaker assignments ...' The method can be used for problems, cases and essays and it can form the basis of a large group teaching session or tutorial.

Delegate the marking

Delegate some of the marking to a team of postgraduate tutors or mentors from more senior years. Provide postgraduate tutors with training, criteria or marking schemes and examples. This approach also provides useful experience for postgraduates in appraising other people's work and in providing feedback. Use peer assessment for one assignment. Provide, or better still, generate criteria with the whole class and provide them with some training in marking based on marking schemes or criteria. Divide the students into groups of four or five. Provide each group with the criteria and some assignments to assess. This task can be built in to the timetabled tutorial programme. This approach not only reduces the teaching and marking load, it also teaches students how to evaluate their own work, how to give and receive feedback and how to work in independent groups. The marks and comments are handed in for moderation by you. You might assign a small proportion of marks for a module for the group work on essay marking.

Review the modular system

The more modules there are, the greater is the assessment and administrative load for students and staff. So do an appraisal of the current modular system and look for ways of reducing assessment requirements. In the longer term this will save you time for other activities.

There are limits to the power of these suggestions. Realistically, if there are more students, then the assessment load will be higher. Time is not elastic so something has to give.

The problem of plagiarism

The pressures on students, whether, academic or financial, can lead to temptation to plagiarise. Three broad strategies can be used to combat it: prevention, monitoring and punishment. Perhaps the most important preventive measure is to teach students what plagiarism is and is not, how to paraphrase and summarise and how to use quotations. Do not confuse the issue of plagiarism with other forms of academic misconduct or cheating. (See the LTSN Generic Centre Assessment Briefing on Plagiarism, 2001.) If students are asked to write essays, model the process (including referencing). This way you can ensure that you give adequate information and instruction on the conventions in your discipline for acknowledgement and citation. One way is to set assignments that limit the opportunities for plagiarism. For example, critically analyzing different recent articles or using last week's newspaper articles may help. Some students seem willing to waste time searching the Web for answers rather than thinking. Spend sufficient time informing students about the precise nature of assessment criteria that will be used. Are you rewarding students for merely finding information or for analysing and evaluating it? If you do not reward for merely finding information but for doing something significant with the information, your chances of minimising plagiarism are quite high.

Involve students more in the design of assessment tasks and in the setting of assessment criteria. There are many resources now for different subjects showing how students can be more involved. If they feel some level of inclusion and ownership, this promotes a more positive attitude to learning.

Do not insult your students' intelligence by giving out the same assessment topics year after year. Students will respond to assessment more positively if they feel it has some meaning. Churning out the same assessments every year says to your students that you do not really care – so why should they?

Monitoring the styles of writing or the presentation of solutions in an assignment is relatively easy. You can benefit from technology and use search engines such as 'Google' to check for passages obtained from the Web and the sites that provide written assignments (see Evans, 2000). Punishment, or the threat of punishment, can inhibit plagiaristic behaviour but, in the long term, once that threat is removed, the behaviour may re-emerge so it is better to work on attitudes and values. Prevention is better than punishment.

Unseen examinations or course work?

The rapid increase in student numbers over the past decade and the consequent rise in plagiarism has led many lecturers to reconsider the pros and cons of examinations and course work. 'Unseen' examinations are easier to design and mark but they narrow the range of outcomes that can be measured. They eliminate plagiarism but they provide only a one-shot sample of students' capabilities, they encourage rote learning and they favour those who have the capacity to withstand acute stress. Course work increases the assessment load for academic and administrative staff and students but it provides a more reliable estimate of a student's capabilities and it can help students to improve. It also measures indirectly the capacity to manage time and withstand chronic stress.

One can reduce the acute stress of examinations by providing students with the topics to be examined or by allowing students to take in their texts and notes. Evidence suggests that these approaches do not change the rank order of students' performance. If the primary purpose of assessment is rank ordering students then one probably only requires four or five assignments per degree programme to make an accurate estimate of a student's capabilities. (If you do not believe this assertion, check the pattern of marks across assignments in different modules.) If, however, one wishes to match performance against programme outcomes then one needs a variety of assessment tasks based on unseen examinations and assignments. Within a degree programme, using only unseen examinations is probably as bad as not using any unseen examinations.

Assessing students with learning difficulties and disabilities

Both examinations and course work present special problems for disabled students. Some of their difficulties are access to learning resources, the physical environment, their rates of information processing and their capacity to communicate their learning. Many universities produce guides for assisting students with learning difficulties and disabilities. The website of the University of Bath (http://www.bath.ac.uk/Students/learningsupport/webb/d33.html) currently contains useful information on assessing students with learning difficulties and disabilities and the DISinHE Centre (www. disinhe.ac.uk) is producing guidance on assessing students with learning difficulties and disabilities. The nub of the issue is keeping a sensitive balance between providing for the needs of students with learning difficulties and disabilities and judgement of their academic performance.

Changing roles: from lecturer to external examiner

The role of an external examiner is guardian of standards, protector of students and agent of quality. It is an extension of the role of the lecturer. It requires an understanding of the subject, course design, opportunities to learn, design of assessment tasks, criteria and sampling together with a knowledge of the appropriate benchmarks and the assessment strategy and system in use in the school. Last but not least, it requires the capacity to make and relay judgements on the overall quality of students' work and the processes of assessment and learning.

The essence of external examining is: know the programme that you are assessing, know the assessment system and know your role. Lecturers who are new to the role of external examining will find useful the text by Partington, Brown and Gordon (1993), the QAA Code of Practice for External Examiners (QAA, 1999) and the video package on Improving Examiners' Meetings produced by Partington et al (1995) for UCoSDA (now the Higher Education Staff Development Agency). The checklist in Partington et al (pp 64-5) is a useful summary of the tasks of the external examiner. Some universities offer induction days for newly appointed external examiners and others offer workshops on external examining and Wisker (1997) provides a set of papers concerned, directly or indirectly, with making the most of external examiners. The core message is: brief them clearly on their role, provide a structured form for their feedback and respond to their comments. Involve them in the design of assignments as well as the examining and moderating procedures and consult them about new developments. Bear in mind that not all external examiners are sympathetic to changes and some request changes that are not viable.

A compulsory examination for all assessors

Answer all questions. This examination is un-timed. Consultation with others (including students) and reference to texts and other sources is RECOMMENDED.

1. What intended learning outcomes do you assess? How well does your approach to assessment align with these outcomes?
2. Justify and criticise your choice of assessment methods and tasks used to assess the outcomes in question 1.
3. Refer to relevant research on assessment in your answer.
4. Describe, justify and criticise your use of criteria, methods of grading and/or marking.
5. Outline and justify your approach to providing feedback to students. Refer to relevant research in your answer.
6. With reference to research findings, describe, justify and criticise your marking techniques to overcome the following:
   a) variations in standards on a single occasion;
   b) variations in standards on different occasions;
   c) variations between assessors;
   d) differences in students' handwriting.
7. How do you ensure that your standards are similar to standards adopted in comparable assessments and examinations?
8. What values underlie your approach to assessment in higher education? How are they manifest in your practice?

Evaluate your answers to questions 1–7.

References

Ashworth, P. and Saxton, J. (1992) *Managing Work Experience.* London: Routledge.

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

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

Bloom. B.S. (1965) *A Taxonomy of Educational Objectives Handbook I : Cognitive Domain.* 2nd ed. New York: McKay 2nd Edition.

Brennan, J. and Little, B. (1996) *A Review of Work Based Learning.* London: Department for Education and Employment.

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

Heywood, J. (2000) *Assessment in Higher Education: Student learning, Teaching, Programmes and Institutions.* London: Jessica Kingsley.

Murphy, R.J.H. (1979) Removing the marks from examination scripts before remarking them. *British Journal of Educational Psychology.* **49,** 73-8

Newton, P.E. (1996) The reliability of marking of General Certificate of Secondary Education scripts: Mathematics and English. *British Educational Research Journal.* **22,** 405-20.

Partington, J., Brown, G. and Gordon, G. (1993) *Handbook for External Examiners in Higher Education.* Sheffield: CVCP UCoSDA.

Partington, J. *et. al.* (1995) Improving Examiners Meetings: A Video-training Package. Sheffield: CVCP UCoSDA.

QAA. (2000) Section 4: External examining. *Code of practice for the assurance of academic quality and standards in higher education.* Gloucester: QAA. Available from: www.qaa.ac.uk

QAA (2000) *Handbook for academic review.* Gloucester: QAA. Available from: www.qaa.ac.uk

Wisker, G. (ed.) (1997) *Making the Most of the External Examiner.* Birmingham: SEDA.

Annotated bibliography

Texts on assessment have different purposes. Some focus predominantly on reviews of research and others are more concerned with providing practical hints. As well as general texts on assessment there are often chapters on assessment in texts on teaching and learning. Web book sites such as Amazon.com and Blackwell.com are worth exploring for texts. There is an enormous amount of material on assessment on the Web. A search based on Google or other good search engines will overwhelm you - so use texts, or wait for someone to produce a guide to useful assessment material on the Web.

Texts on assessment

Atkins, M., Beattie, J. and Dockrell, B. (1993) *Assessment Issues in Higher Education.* Sheffield: Employment Department.
A substantial and scholarly commentary on assessment and the purposes of higher education.

Brown, G., Bull, J., and Pendlebury, M. (1997) *Assessing Student Learning in Higher Education.* London: Routledge.
This text describes in detail all the topics in this guide and it reviews the relevant research. Several workshop activities are included. Useful for lecturers and staff developers.

Brown, S. and Knight, P. (1994) *Assessing Learners in Higher Education.* London: Kogan Page
A lively discussion of assessment replete with hints and persuasive arguments.

Brown, S., Race, P. and Smith, B. (1996) 500 Tips on Assessment. London: Kogan Page
A rich compendium of suggestions on different methods and procedures of assessment.

Gibbs, G., Habeshaw, S., & Habeshaw, T. (1988) *53 Interesting Ways of Assessing Your Students.* 2nd ed. Bristol: Technical and Educational Services.
A useful set of hints on different approaches to assessing student learning.

Gibbs, G. (1992) *Assessing more students.* London: PCFC (now HEFCE)
A useful summary of approaches to assessment that focuses upon the problems of assessing larger numbers of students.

Heywood, J. (2000) *Assessment in Higher Education: Student learning, Teaching, Programmes and Institutions.* London: Jessica Kingsley
For a detailed review of the literature on assessment. Considers Australian and US findings as well as assessment of competencies and examinations in secondary education. Invaluable for researchers.

Miller, A., Imrie, B. and Cox, K. 1998 *Student Assessment in Higher Education.* London: Kogan Page.
A practical guide to assessment supported by research findings.

Rowntree, D. (1987) *Assessing Students - how shall we know them?*. 2nd ed. London: Harper & Row
A seminal work. A stimulating discussion of the ideas and assumptions underlying assessment.

Specific topics of assessment

Assessing students with learning difficulties and disabilities.

Disability in Higher Education (www.disinhe.ac.uk)
http://www.bath.ac.uk/Students/learningsupport/webb/d33.htm
Both these sites contain articles and links relevant to the assessment of learning of disabled students.

Assessing Essays

Hounsell, D. (1995) Marking and Commenting on Essays. In F. Forster, D. Hounsell. and S. Thompson (eds.) *Tutoring and Demonstrating: A Handbook.* Sheffield: HESDA
A useful practical guide on marking essays for tutors.

Computer- based assessment

Brown, S., Bull, J. and Race, P. (eds) (1999)
Computer-assisted assessment for students. A set of articles on different ways of using computers for assessment purposes.

External Examining

Partington, J.,Brown, G. and Gordon, G. (1993) *Handbook for External Examiners in Higher Education.* Sheffield: HESDA (formerly UCoSDA)
Covers all aspects of External Examining including the examination of theses

Laboratory Teaching

Hegarty-Hazel, E. (ed.) *The student laboratory and the science curriculum.* London: Routledge
A useful collection of readings on teaching and assessing laboratory work.

Multiple Choice Questions

Gronlund,N.E. (1988) *How to Construct Achievement tests.* Englewood Cliffs: Prentice Hall.
A comprehensive text on designing assessments that focuses upon MCQ's and allied methods.

Peer and Self assessment

Brown,S. and Dove,P. (eds) (1991) Self and Peer Assessment. Birmingham: SCED Paper 63.
A collection of articles that provide hints and some research on self and peer assessment.

Falchikov, N. (2001) *Peer assessment and peer tutoring.* Buckingham: Open University Press (in press).
A comprehensive text on peer assessment and tutoring which includes theory as well as reports of practical approaches to peer assessment.

Plagiarism

Evans, J.A. (2000) *The New Plagiarism in Higher Education: From Selection to Reflection* [online]. Available from:
http://www.warwick.ac.uk/ETS/interactions/vol4 no2/evans.htm
A well referenced, thought provoking discussion of plagiarism.

Problem solving

Schoenfeld, A.H. (1985) *Mathematical Problemsolving.* New York: Academic Press.
A text with a strong mathematical application.

Work-based learning

Ashworth, P. and Saxton, J. (1992) Managing Work Experience. London: Routledge.
It contains some useful suggestions about assessments and arguments in favour of workbased learning.

Higher Education Research & Development. **17,** 2 June 1998.
The whole of this issue is devoted to work-based learning. Assessment issues are scattered throughout the issue.

The Learning and Teaching Support Network Generic Centre

The Learning and Teaching Support Network (LTSN) is a network of 24 Subject Centres, based in higher education institutions throughout the UK, and a Generic Centre, based in York, offering generic information and expertise on learning and teaching issues that cross subject boundaries. It aims to promote high quality learning and teaching through the development and transfer of good practice in all subject disciplines, and to provide a 'one-stop shop' of learning and teaching resources for the HE community.

The Generic Centre, in partnership with other organisations, will broker information and knowledge to facilitate a more co-ordinated approach to enhancing learning and teaching. It will:

- Work with the Subject Centres to maximize the potential of the network;
- Work in partnership to identify and respond to key priorities within the HE community;
- Facilitate access to the development of information, expertise and resources to develop new understandings about learning and teaching.

The LTSN Generic Centre Assessment Series Guides for:

Senior Managers
Heads of Department
Lecturers
Students

Briefings:

Assessment issues arising from key skills
Assessment of portfolios
Key concepts: formative and summative, criterion and norm-referenced assessment
Assessing disabled students
Self, peer and group assessment
Plagiarism
Work-based learning
Assessment of large groups

Published by
Learning and Teaching Support Network (LTSN)

For more information, contact the Generic Centre at:
The Network Centre, Innovation Close, York Science Park, Heslington, York, YO10 5ZF
Tel: 01904 754555 Fax: 01904 754599
Email: gcenquiries@ltsn.ac.uk
www.ltsn.ac.uk/genericcentre