Designing an assessment system

Derek Rowntree, Lately Professor of Educational Development, Institute of Educational Technology, Open University, Milton Keynes MK7 6AA, UK.

http://iet.open.ac.uk/pp/D.G.F.Rowntree/Assessment.html

e-mail: **d.g.f.rowntree@open.ac.uk**

Introduction

As any HE teacher soon becomes aware, assessment is a major concern in designing and running a course. It can take up a large part of both teachers' and students' time, cause considerable anxiety and play a major role in determining how and what learners learn. A recent study at Leeds Metropolitan University (Innis, 1996) confirmed one's suspicions that most of students' out-of-class study time is devoted to assessed tasks. Indeed, it has often been observed that assessment is the tail that wags the educational dog -- but is it always wagging as we might wish? The purpose of this essay is to help you develop or improve an assessment strategy that will truly foster and reward your students' learning.

The literature on assessment is huge -- see the journal *Assessment and Evaluation in Higher Education* or the bibliographies in, for example, Brown et al, 1997; Freeman & Lewis, 1998; Knight, 1995; Rowntree, 1987 or point your Web browser at 'assessing students' and prepare to surf the resulting tidal wave of hits. I am making no attempt to review that literature here. Rather, my aim is to present an overview of the main strategic issues which I hope will then provide a backdrop against which you may want to review your own practice.

The key questions?

In developing your assessment system -- and in thinking about how to explain it and justify it to colleagues, students or other 'interested parties' -- you may need to consider the questions set out below. I'll be using these questions to structure this chapter:

- What is assessment?_
- Why might we assess?_
- What might we assess?_
- Who might do the assessing?_
- How might we assess? (which includes When? How often? and Where? as well as By what means?)_
- What might we do as a result of assessment?

As you may imagine, these questions cannot be answered independently of one another. Certain answers to any one of the questions may determine or preclude certain answers to

one or more of the others. Hence, planning an assessment strategy (like planning a course) is often an iterative process. That is, we may need to re-consider aspects we've looked at earlier in the light of our later thoughts about other aspects. The ideal, of course, is to have them all in mind at once, like a juggler spinning a number of plates.

But this is not always easy, especially if there a several 'jugglers' involved, each spinning to a rather different tune! For rarely are we free to answer such questions just as we like. The way we assess may have an impact on the work of other people and it may have to fit in with what is expected (or demanded) of us by others. Hence, arriving at an acceptable assessment strategy may involve consultation, discussion and agreement among a variety of 'stakeholders'. Which of the following might have an interest in your assessment strategy or wish to influence it?:

- colleagues teaching the same course or module_
- colleagues teaching related courses/modules_
- colleagues in other disciplines_
- the management' (at department or institution level)_
- external assessors (e.g. from other universities)_
- outside bodies (e.g. professional or employers' organisations)_
- and, of course, your students.

What is assessment?
It is easy to get fixated on the trappings and outcomes of assessment -- the tests and exams, the questions and marking criteria, the grades and degree results -- and lose sight of what lies at the heart of it all. For me, assessment is essentially an attempt to get to know about the student and find out the nature and quality of his or her learning -- e.g. his or her strengths and weaknesses, interests and aversions, motivations and approaches to learning.

Assessment is not the same thing as giving marks or grades. Certainly, we cannot sensibly mark or grade a student's work without having first assessed it. (What did the work consist of? How did it match up against our criteria? What do I think of it?) But we can, and often do, assess without seeing the need to quantify or attach letter grades to what we have observed. Often the most appropriate response is to comment to the student on what we have seen -- e.g. about how they have improved in various aspects of their work or about ways in which they might build on what they have achieved or might overcome some difficulty.

Nor is assessment confined to the arena of tests, examinations or formal assignments. Those of us who are still fortunate enough to meet our students in seminars, discussions or practicals where we can interact with them as individuals will also be assessing their performance in these situations. The assessment may be informal and unselfconscious but we cannot interact usefully -- indeed we cannot claim to be teaching -- unless we regularly attempt to appraise the nature and quality of the student's thinking or practical

performance. It is an essential element of teaching. Such assessment will probably not be translated into grades or even 'count' towards the student's overall for the course (but it may well influence the way we do mark the student's more formal pieces of work).

Formative vs summative assessment

What I am highlighting in the last two paragraphs is the distinction between:

- 'formative' assessment -- where one is using what one has learned of the student's learning to help him or her develop or improve that learning, usually by giving appropriate oral or written feedback to the student, and_
- 'summative' assessment -- where one is summing up one's judgement of that learning, usually with a mark or letter grade (though it could be done with a written profile of strengths and weaknesses).

Power relations in assessment

While formative assessment is usually for the student's benefit, summative assessment is often for the benefit of other people -- e.g. other teachers or potential employers -- who might use the information you provide to make decisions affecting the student's life-chances. Herein lies a potential conflict of roles for the teacher -- between helper and informer -- and a conflict between formative and summative assessment. Students who most need help may be reluctant to reveal their difficulties, or to choose learning options that are more challenging, for fear of being adversely reported on.

In this respect, summative assessment can be seen to highlight an inescapable power relationship between students and teachers. That is, the teachers are in a position (rather like parents or employers) to reward students according to how far they approve of what they say and do -- whereas the students (unless their institution relates staff rewards to feedback from the 'customers') have no reciprocal powers. Many new teachers nowadays feel uncomfortable about this, especially when their students differ from them in age, sex, religion, social or ethnic background, and so on. The potential for unjust discrimination is obvious. However, so long as we who teach also do the summative assessment, we must learn to work around the conflict and live with the equal opportunities implications.

Why must we assess?

Life for teachers would be a whole lot easier if we didn't have to formally assess our students. We all know how endlessly time-consuming it can be to plan assessment exercises and pore over the students' responses, and then there's all that heart-searching and potential unpleasantness involved in allocating a fair grade to each individual. If only we could assume that because we have 'taught', they must have learned. But we know it's not like that.

By and large, students are strategic beings who allocate their time and effort rationally. That is, they decide what line of action will provide them with the greatest rewards (or minimise hassle). For those students who are studying because they are simply fascinated by their subject, this will mean devoting all the time they can make available to whatever aspects of the subject have most appeal for them. Students who are studying because they are looking for practical applications in their current jobs may similarly immerse

themselves in whatever has most immediate workplace relevance. And students who are anxious to gain certification (or to stay out of trouble) will concentrate on whatever they think is going to be 'on the test'.

In none of the three cases mentioned above can we be sure that students will be exploring the subject as widely or in as balanced and appropriate a manner as they might. This is all right, provided we don't much mind what our students learn so long as they seem satisfied with what they are getting out of their courses. But if we have views (as most of us have) about what is most worth learning, then we need to get our assessment strategy right.

My colleague, Graham Gibbs (1999) provides several examples of how small changes in assessment strategy can dramatically transform the way students learn. He tells, for example, how students on a second year engineering course were finally persuaded to prepare properly for problem classes and actually practice at problem-solving (leading to an improvement in the average exam score from 45% to 75%). This was achieved by the simple expedient of requiring students, on six occasions during the course, to bring their worked problem sheets to class, redistributing them among the group and asking students to mark and write comments on one another's work, using given criteria.

The marks and comments from peers were not recorded, did not count towards the overall course grade and were not vetted by the teachers; yet the results were better than when teachers had been able to work with with much smaller groups and mark the problems themselves in weekly sessions. Graham Gibbs attributes the improvement to the assessment system having: (a) required students to put an adequate amount of time in, (b) encouraged them to apply themselves to appropriate learning activities (not just problem-solving but also applying and internalising relevant assessment criteria), (c) provided timely feedback, and (d) introduced the element the peer-pressure element that meant they were less willing to shirk or skimp the work for fear of losing the respect of their classmates.

In another case, Graham Gibbs records how a major change in students' learning practices was achieved by changing the examination task; in yet another, the breakthrough strategy was to group students into learning groups of four and tell them that they would be taking the exam individually but each would be awarded the average mark of their team. Using assessment strategically to persuade students to learn appropriately may make no extra demands on teachers and, indeed, may make fewer.

The variety of roles for assessment
If assessment is properly designed it can play six major roles in ensuring the teaching and learning is productive; that is, it can:
- remind students that someone cares about their learning_
- enable teachers to clarify what they believe is most worth learning_
- direct students' efforts towards key aspects of the course_
- engage students' in learning activities appropriate to the subject_
- reveal a student's strengths, weaknesses and ways of learning_

- enable students to be given feedback that will help them improve.

Apropos that second to last role, the student's qualities may become clearer not just to the teacher but also to the student; as one told me recently:

"I was surprised when doing the assignment how much knowledge and how many ideas I had developed . . . it was only reflection on the assignment that made me realise how much I had learned."

Notice that those essential functions for assessment do not necessarily involve the giving of marks or grades. Nor do they have anything to do with evaluating the teaching effectiveness of the course. So we might add that assessment has four further roles; it can also help:

- teachers, by assigning marks or grades, to indicate succinctly how the standard reached by each student's work compares with that of other students or with some criterion level (e.g. a pass-mark) or with the student's previous work_
- people other than the teacher (e.g. administrators, employers or other teachers) to appreciate what standard any given student has attained_
- teachers and administrators to evaluate the effectiveness of the course and improve it where necessary_
- other people from outside (e.g. teaching quality assessors) to evaluate the quality of our course provision.

You may need, at some point, to consider whether an assessment exercise designed to meet a role from one of these lists can realistically be expected to meet a role from the other list also. For example, you may devise a collaborative group exercise that generates valuable learning and gives you valuable information about how to help members of the group but find that it doesn't produce information that administrators, evaluators or examination boards will find easy to relate to. A controlled examination, on the other hand, may seem to serve most of the second set of roles with much serving any of those in the first set.

What might we assess?

What will you be looking for in the work of your students? What knowledge, skills and maybe attitudes or dispositions will you be wanting to encourage and reward? You will do doubt already be be familiar with the idea of aims, objectives and learning outcomes (Allan, 1996). These will be what to focus on in developing your assessment strategy. Whatever knowledge, skills and attitudes or dispositions you assess will tend to be those that students attend to and improve in -- provided, of course, you make it plain to students just what it is you will be assessing them on.

Views about knowledge

The learning outcomes you prefer to assess will be largely dependent on your beliefs about what worthwhile knowledge consists of. In this, I imagine a continuum. Towards one extreme on this continuum we find teachers whose first loyalty is to a public but carefully circumscribed body of pre-existing knowledge. This knowledge -- comprised

not just of facts, but also of special ways of looking at the world, special ways of talking about it, and special ways of acting upon it -- is derived from the experience of scholars, practitioners, researchers, all experts in their particular academic discipline.

These teachers act on the assumption that each of their students ought to ingest a sizeable portion of this knowledge if they are to be, for example, a decent physicist, safe doctor, competent statistician -- or even if they are merely to understand the concerns of such specialists. Knowledge, for them, is non-negotiable. Their ideal student is one who knows how to provide the right answers. In this context I remember the head of a geology department who told me: "We are a Marxist department and our students study Marxist geology".

The other extreme is attractive to teachers who distrust generalisations about what every student ought to know. Instead, they are interested in students developing their own knowledge by constantly reflecting on and critiquing their own experience, their own concerns, their own conceptualisations and whatever effective ways they have found of looking at the world and acting upon it. They might even argue that this is what distinguishes between the competent physicist, pilot, statistician etc, and the outstanding one.

For these teachers, knowledge is a personal and (in so far as the knower is influenced by other people) a social construction. Their ideal learner is one who knows how to ask potent questions. Here I remember the literature professor who told me: "I want my students to be able to say something that makes me rethink my own conception of an author or a literary work".

Dissemination vs development
As you will realise, I am opposing two stereotypes here. Few real teachers will dig in at either end of the continuum. The same teacher may even take up different positions with regard to different sections of the same course. But even within a department one may meet a fair range of disagreement about the extent to which there is stuff that everyone must know and the extent to which we can encourage learners to develop their own vision. Richard Boot and Vivien Hodgson (1987) may have had something like this continuum in mind when they distinguish between what they call a "dissemination" model of learning:"knowledge can be conceived of as a (valuable) commodity which exists independently of people and as such can be stored and transmitted (sold)."and the "development" model whose aim they say, is:"the development of the whole person, especially the continuing capacity to make sense of oneself and the world in which one lives."

Clearly, teachers will favour different kinds of learning outcome according to where they locate themselves on the continuum I have suggested above. The 'disseminators' will frame their objectives in terms of the body of knowledge or procedures that must be passed on intact to their learners for them to reproduce. There will be no room for different learners to have individual objectives each according to her or his own personal angle on the subject.

But the 'developers' can also have explicit learning outcomes. These may also be framed in terms of a body of knowledge but are more likely to put the student's perceptions into the frame also&emdash;e.g. not "outline the pros and cons of using objectives in planning HE courses" but "identify the pros and cons of using objectives in planning courses in your discipline".

Student autonomy
Developmentally inclined teachers are also likely to encourage student autonomy. And not just the limited autonomy of being able to choose their own way of working (towards some one else's objectives) but also that of being able to develop and pursue their own objectives. (Postman & Weingartner, 1969, got a lot of us thinking along these line when they challenged us to "prohibit teachers from asking any questions they already know the answers to".) Many courses now include or take the form of a project in which students (with guidance from a tutor) define their own objectives, content and approach which they then pursue with the tutor's continuing guidance.

If it is one of our aims to foster students' autonomy, we may have in mind not simply objectives relating to the content and methodology of our discipline, but also 'self-actualisation' objectives like these suggested by the humanistic psychologist and teacher, Carl Rogers (1961, pp 280-1):

" The person comes to see himself [sic] differently.

He accepts himself and his feelings more fully.

He becomes more self-confident and self-directing.

He becomes more the person he would like to be.

He becomes more flexible, less rigid in his perceptions.

He adopts more realistic goals for himself.

He becomes more acceptant of others.

He becomes more open to the evidence, both of what is going on outside of himself and what is going on inside himself."

A course that embraced learning outcomes like these would clearly demand very different forms of assessment than one whose only aim was to transmit a body of knowledge. In particular, it would almost certainly allow for assessing at least some objectives that students had decided for themselves or negotiated with their teachers.

Spelling out the discourse

Any worthwhile assessment strategy needs to include provision for helping students understand just what we are looking for in their work. This goes beyond showing lists of objectives or learning outcomes to students or even sharing with them our marking criteria. We may need to discuss with them, and on more than one occasion, not just the content or intended outcomes of the course but also the nature of the 'discourse' in our discipline -- what the discipline's practitioners regard as appropriate subject-matter, the form of argument that is expected, the kind of evidence that is acceptable, the criteria for truth or elegance, the language and style of presentation.

It may not be enough to hope that students will absorb the subtler aspects of the discourse by some kind of osmosis. They may well be taking courses in more than one discipline and it can seem contradictory for a student to be told, as one I know was, that her geography essays are "too literary" while her history essays contain "too many headings and lists"! Maybe we need to show and explain examples of good and not so acceptable practice within our discipline?

Who might do the assessing?
Unlike school teachers, teachers in higher education in the UK are normally responsible for all the summative assessment of their students as well as for the formative variety. Certain assessment exercises may be subject to 'second marking', and at some stages there will be input from an 'external examiner'. But this still leaves the individual tutor assessing the work of individuals who are well known to him or her and, ultimately, helping decide what class of degree they are to be awarded.

A variety of assessors
But the teachers who design the courses may not be the only ones who assess -- whether formatively or summatively. Much of the assessment on campus (e.g. the marking of problem sheets or the supervision of practicals) may now be done by graduate teaching assistants. In the Open University and in other universities that have enrolled large distance learning classes, much of the marking will be done not by the course developers but by a small army of part-time tutors or associates. These assessors may be physically distant from the students, from one another and from the course developers. And if students' courses call for them to spend time in work attachments or internship, their managers or mentors in the workplace may also be expected to contribute to assessment.

If people other than those who developed the course are to be involved in assessment (even if colleagues in the same department), they cannot be expected to guess what assessment strategy the developers may have had in mind. They will need some form of briefing, if not training, for their role. And they may need continuing support once they have embarked upon it. So a coherent assessment strategy must include some means of providing what such assessors might need.

They will need printed documentation about the assessment strategy and a chance to discuss it both with the course developers and with one another. They will need similar briefing notes about each different form of assessment they are meant to carry out and about individual assessment exercises -- e.g. what are they to look for in each exercise

and what marking criteria apply. This is not merely an information-giving exercise, for some 'assessors' may find it difficult to understand or agree with the rationale espoused by the course developers and a certain degree of negotiation may be called for! Even with the assessment of mathematical calculations, for example, there may be differences of opinion about whether to give credit for elegance of working (or indeed any working) or just for the correctness of answers. And whatever criteria are agreed, team of assessors may also need to take part in some sort of 'agreement exercise' where they all independently assess the same set of student products or performances as a means of trying to establish common standards.

Peer and self-assessment
Nor are other experts the only others who may be involved in assessing our students. They may well be expected to do some assessment themselves -- either of their colleagues' work (peer-assessment) or of their own work (self-assessment). Much has been said in favour of peer- and self-assessment -- e.g. in helping students internalise appropriate standards and improve their own work by critiquing that of colleagues, in providing feedback of a kind students are more likely to heed and, not least, in taking some of the marking load off hard-pressed tutors (Boud, 1986; Robinson, 1999). Such benefits can be worth having even where your local traditions insist that peer- or self-assessment can be formative only and contribute little or nothing to a student's summative results (as with Graham Gibbs's engineers).

However, we need to face the possibility that such assessment will be making new demands on students and they may initially be resistant to the prospect. So, if this is to be part of your assessment system, you will need a plan for demonstrating to students (not just telling them) that they are likely to benefit from it and for helping them develop the necessary skills. This may take some weeks. Graham Gibbs mentioned that the lerning teams who were to share their average mark had first to be reassured of the positive reults this approach had achieved elsewhere.

External assessors
Nor must we forget the need to negotiate with our 'external examiner'. British universities, as a gesture towards parity of standards between institutions, have long required departments to have some form of input to their assessment procedures from a peer in another institution. This may include commenting on the overall assessment strategy and on major assessment exercises as well as double-marking a sample of students' work.

How far such input can or does help ensure that one institution's standards are much like another is questionable. It may not even be high in the external examiner's list of priorities (Wicker, 1993). There is no doubt, however, that a discerning and resolute external examiner can bring about considerable rethinking of a department's approach to assessment (see Appendix 3) and in the week I write these words a UK university has been censured by the national Quality Assurance Agency, largely on the basis of damning reports from its external examiners.

How might we assess?
Before we plunge into considering methods of assessment, there are two related issues we might bring into play -- when to assess, how often and where?

When shall we assess?
Most HE courses nowadays include both continuous assessment (e.g. weekly problem sheets or fortnightly essays) as well as, or instead of, end-of-course examinations. Most courses always did include regular assessed work, but often it did not count towards the overall course grade and so may not have guided and stimulated students' activities quite so powerfully as continuous assessment for credit can do.

Clearly, your assessment strategy should enable students to demonstrate the knowledge, skills or attitudes they have acquired or improved upon during the course. Some of these will be assessable week by week; some only after several weeks of learning. So which will you be assessing when?

Some learning outcomes will be so dependent upon the impact of the course as a whole that a true assessment will be impossible until near its end. For example, medical students might be assessed immediately after the relevant teaching session on whether they can, for example, list the signs and symptoms of rheumatic fever or the characteristics of presystolic murmur, or whether they can elicit a patient's history in a systematic but sympathetic manner. But whether they can diagnose a patient suffering from rheumatic heart disease may be an outcome that is only assessable after they have had many weeks of experience.

Topic vs course objectives
Much higher education requires of students that they integrate a number of separate objectives; not merely that they have attained them and can demonstrate them separately. So, in planning your assessment strategy, you may find it useful to distinguish between what we might think of as:
- 'topic objectives' -- relating to each particular topic within the course, and_
- 'course objectives' -- where the student may need to synthesise ideas from several topics and perhaps demonstrate more developed skills or attitudes than could have been expected earlier.
-

How often to assess?
So you may need to consider which outcomes you will be assessing at different points in the course. And this raises the issue of how often to assess -- once a week, once a fortnight, once a month? Frequent assessments have the benefit of pacing the students' activities and enabling you to find out in time to do something about it if any are having difficulties. They may also make it easier (especially with large classes) to be sure that the person producing the work is the person who is registered for the course. On the other hand, it is possible to do so much assessment, so frequently, that students wilt under the pressure and lapse into surface level learning or even into cheating.

Budgeting assessment time

Using a variety of forms of assessment, some demanding less time than others -- e.g. brief quizzes as well as lengthy assignments, may help us avoid overloading our students (and ourselves). But we must always ask ourselves, about each assessment exercise we are inclined to set:

- Will this form of assessment encourage appropriate learning?_
- How long is it likely to take our students?_
- Does that seem reasonable in terms of the learning benefits they are likely to get from it?_
- Can they afford the time in the light of other things I am hoping they'll be spending time on right now?_
- Are there likely to be conflicting demands from teachers of other courses the students are taking? (This aspect is often overlooked.)_
- Is it realistic in terms of other (non-academic) events that are likely to be happening on campus or elsewhere in the students' lives around this time?_
- Have I (or some other appropriate person) got time to give students timely and helpful feedback?

Such questions need to be considered before the start of a term or semester and we need to devise a sequence of assessment exercises that will stimulate and recognise the kind of learning that is appropriate to successive phases of the course. (See Appendix 1 for a case study of how one such planned sequence was worked out.)

Where to assess?

There are two apparent issues here -- where is the student to produce whatever it is that is being assessed and where will the assessors be when they are doing the assessing? But there are underlying issues of openness vs control and of fairness vs truth.

Control vs openness

When I was an undergraduate, the only assessment exercises that counted were those done under the eye of the assessor or the assessor's proxy (an invigilator in an examination). Such a location for assessment offers the ultimate in control: the identity of the person producing the work can be verified; they can be required to tackle a previously 'unseen' task within a set time limit; and they can be denied access to other people and resources that might help them. (Though one or more of the strictures can be relaxed, thus lessening the degree of control.) The actual assessment, of course, takes place elsewhere, at a time and place to suit the person who is marking the examination scripts.

Fairness vs truth?

The rigidly controlled examination was once held to be fairest for all students, since all are treated alike. However, it was never as fair as it seemed, since students are not all alike and some find the pressures and constraints of examinations more stultifying than others. Hence, a student's performance under highly artificial exam conditions may bear little resemblance to his or her performance in more relaxed 'real-life' circumstances. In some professions that students may follow after graduating (though not in ours!), it may occasionally be necessary to work under exam-like constraints of time, lack of prior warning and reduced resources; so we may justify giving students some experience of)

this. But few professionals might reckon to produce their best work under such constraints; so how do students perform if the conditions are more open? Unfortunately, there is always a fear that, under such conditions, some students may cheat (which is, of course, unfair to the others) -- so the case for at least some controlled examining continues to be made.

Nevertheless, in recent years -- partly in recognition of the fact that assessment activities are themselves a vehicle for learning rather than merely a device for demonstrating what learning has already taken place -- many assessment items (e.g. essays or problem sets) are produced in a place of the student's own choosing (away from the eye of the assessor or a proctor). Hence they will be able to spend as long as they like over the task and consult whatever resources and colleagues they wish. If teachers have any doubts, however, about the authorship of what is produced, they may temper this openness by engaging students in discussion about the work they offering for assessment, e.g. by means of a viva.

With assessment of practical activities, e.g. laboratory procedures, dance or medical diagnosis, both students and assessors need to be in the same place at the same time, e.g. in a laboratory, a theatre or a surgery. Such demands are clearly more controlling (for the assessor as well as the assessed) but certain abilities can be truly assessed only while they are in action, and even a video of the event cannot be relied on to give the full picture.

By what means?
There are many different methods of assessment to choose from (summarised in the figure below). Broadly speaking, we may either assess something the student has produced (a product) or we may assess the student in action (a performance). These products and performances may be solo efforts or collaborative (group) enterprises. They may take a variety of forms to suit the nature of our discipline and the particular learning outcomes we are looking for.

Assessing the right thing
Hence we need to choose assessment methods that will enable students to demonstrate the most important learning outcomes we have identified -- even if other, related but less important ones would be easier to test. For instance, if we want to know whether someone can repair a bicycle tyre, we won't find out (or encourage the appropriate learning) by asking them to write an essay about the ecological significance of cycling or even about the mending of bicycle tyres. We can only truly do so by asking them to mend a bicycle tyre -- even though we may find this more costly and inconvenient to arrange.

Similarly, the so-called 'objective' or multiple-choice tests -- while easy to present and mark via computer screens (though not to develop worthwhile questions for) -- may be of limited use in relation to the more sophisticated abilities we are seeking to encourage and reward in higher education. (Yet they may nevertheless be useful as a means of checking that the learner is still engaged with the course.) To use the words of Robert MacNamara (when president of the World Bank), 'We must strive to make the important measurable, rather than the measurable important.'

Even when we are do seem to be testing what we think we are testing -- e.g. testing students' understanding of chemistry by asking them to solve chemistry problems -- we may be deceiving ourselves. A review of research on problem solving in chemistry shows that many students manage to solve chemistry problems using algorithms and memorised procedures without understanding the chemical concepts on which the problems are based (Gabel and Bunce, 1994). Hence we would need to refine our assessment by asking ourselves how to pose the problems in such a way that algorithms are of no assistance.

Fitness for purpose

So, our assessment system might well include a combination of multiple-choice questions, essays, practical tests and other forms (open or closely controlled) -- but all used in different ways, for different purposes and in different phases of the course. The key issue here is that of fitness for purpose. What forms of assessment will encourage the kinds of activities we want our students to engage in and help them to develop and display the kinds of ability we hope to recognise and reward?

Thus, if want students, for example, to come to grips with a variety of sources and conflicting viewpoints or contrary data, arrive at a perspective of their own on the topic and present a convincing argument, we cannot rely on multiple-choice questions or short-answer tests. It has long been noticed that students who think they are to be assessed with multiple-choice questions will study by looking for factual, testable items, gobbets of detail or technical terms, while those who believe they are to write an essay or report will be looking for underlying ideas, themes and general principles. Students' ability to choose correctly among other people's answers (as in an objective test) gives us little clue as to the kind of answers they might produce for themselves (or vice versa). As the educator Jacques Barzun once said: 'The mere recognition of what is right in someone else's wording is only the beginning of the awareness of truth'.

Similarly, if we want to encourage collaborative learning among our students, we need to set them tasks in which they can collaborate. And if we want to ensure that they do indeed collaborate, we need to assess the product(s) of their collaboration (or else watch them doing it). The ability to work productively in teams is seemingly much prized by employers but conflicts with the individualist, competitive ethos of most higher education institutions. So any form of assessment intended to encourage and reward group-work may need to be introduced to students with extra attention to explanation and justification.

Validity, practicality and reliability

Throughout this section (and occasionally in earlier ones) I have used the concept (though not the term) of validity -- that is, does the chosen assessment method truly test what you think it is testing (or something else instead)? It is worth noting that validity is often in conflict with practicality. That is, a test that would give valid evidence may sometimes be so costly and inconvenient to arrange -- e.g. setting up a series of lengthy practical sessions with students to be observed one at a time by the assessor -- that it can be tempting to fall back on a simpler (but less valid) procedure instead.

There is also a potential conflict between validity and reliability. The latter concept, which might be better called 'consistency', concerns the extent to which a given piece of student work is likely to be given the same grade by different assessors (or even by the same assessor on different occasions). The assessment of multiple-choice tests is highly reliable, because every marker (or even a machine) would give a student the same score. But we cannot expect the grading of written work, practical performance or even sets of calculations to be so reliable. Even before the beginning of this century (Edgeworth, 1890), researchers were demonstrating that a group of apparently well-qualified assessors may award a wide range of marks to the same essay. Worse than that, they may not even agree as to whether that essay is better or worse than some other essay. More recently, Graham Gibbs (1995) reports that:

"...a study of double marking of psychology assignments at the University of Cambridge found no significant correlations between markers: marks were effectively random. A study of internal and external markers of psychology assignments at the University of Plymouth found marks for a script varying by up to 35% between (internal) markers and external markers to be even less reliable."

Can we improve reliability?
Poor reliability -- or the likelihood that different experts might make different judgements of the same piece of work -- seems to me inevitable when we are assessing products or performances that are meant to reveal more than replication of disseminated knowledge or routines. In higher education we are usually expecting students to go 'beyond the information given' (in the memorable words of the educator Jerome Bruner), and add something of their own. The more they do so, the more likely they are to diverge from some 'right answer' or 'right approach' we may have had in our minds.

Reliability can always be improved -- e.g. by briefing students in such a way that they are more likely to give similar answers or by briefing assessors in such a way that they are more likely to looking for the same things and rewarding them in the same way. This may be more appropriate in some disciplines than others, and in some aspects of a discipline rather than others. But it would be to encourage and reward convergent learning rather than freeing students to be creatively divergent. How much convergence are you looking for in the learning outcomes you are trying to achieve?

Remember that the more guidance you give as to what kind of answer is expected, the less chance your students have of showing they do not need such guidance. Where you draw the line must depend on your objectives. If you give your learners too many hints and clues, you may lead them to demonstrate achievement not of the objectives you initially had in mind but of some lesser, more restricted ones, instead.

As an illustration, look at the sequence of questions below. They all concern the Swedish political system. But each one is testing rather different abilities and insights:

Qu.1 What aspects of the political system of modern Sweden seem to you most worthy of comment?

Qu.2 Comment on the political stability of modern Sweden.

Qu.3 Explain the political stability of modern Sweden.

Qu.4 Identify and discuss three factors that might help explain the political stability of modern Sweden.

Qu.5 Identify and discuss three factors that might help explain the emergence of a stable political system in Sweden despite the massive social and economic changes engendered by processes of modernisation.

Qu.6 Which three of the factors listed below might best help explain the emergence of a stable political system in Sweden despite the massive social and economic changes engendered by processes of modernization?

(a) Affluence
(b) Gradual economic development
(c) Traditional legitimacy
(d) Civic culture
(e) A homogeneous political culture
(f) Equality
(g) Congruent authority patterns
(h) Elite consensus
(i) The habituation of mechanisms of conflict resolution
(j) Political institutionalisation

One by one, the questions become more specific -- leaving the student less and less responsibility for deciding what is significant about the situation referred to, and in what terms to respond to it. For instance, the student answering Question 1 may choose not to discuss political stability. The student answering Question 2 may choose to describe political stability without explaining it. The student answering Question 3 may choose to refer to more or to fewer than three factors. And so on.

Finally, Question 6 takes multiple-choice form. It tests merely whether the student can recognise the three crucial factors when they are presented in a list. It could be followed by further multiple-choice questions taking each of the ten suggested factors in turn and asking 'If this was one of the three factors you chose, which of the statements below best explains the contribution made by that factor'. But such a sequence of multiple-choice questions would not be testing the same abilities as, say, Questions 3 or 4.

By similar means -- that is, by curtailing their scope for exercising independent judgement -- assessors too can be led to greater convergence. We may give them very precise instructions (a marking guide) about what they are to reward and what they are to penalise, possibly indicating how many percentage points they may award (or deduct) for each of a number of desirable (or undesirable) factors. An 'agreement exercise' of the

kind mentioned earlier might further increase convergence by helping assessors agree on how to interpret the marking guide. Such devices may improve the fairness of assessment by making it more likely that all students are treated in the same way, whoever assesses them.

One other potential source of unreliability or inconsistency in assessment is worth mentioning before we move on. That is, even if you are doing all the assessing yourself, how far can you be sure that you would respond similarly to the same piece of student work today as you would tomorrow or yesterday? Or that you would be applying the same criteria and standards to the work of two (or more likely twenty, thirty or more) students' work on any given day?

So, even if you are the only assessor, you may still need to write down your criteria -- for the course assessment in general and for each individual exercise or assignment -- so that you have by you an aide memoire that will reduce the likelihood of your standards drifting over time.

Whether the criteria or marking guides are for you only or for other assessors also, you will need to share them, at least to some degree, with your students. But to what degree? I leave you to consider the ethical issue as to how far we might be justified in what is the common practice of giving students less detailed guidance than we do assessors about what the assessment exercise is really looking for. Is there a potential conflict here between validity and fairness?

What might we do as a result of assessment?
As a result of assessment -- which we may do either by informal observation or by arranging formal exercises -- we gain some sort of information about our students. This information we interpret, with various degrees of validity, encouraging us to feel we know more about what they have learned and what they have not learned; about their strengths and weaknesses; about their original insights and strange blind spots; about their interests and aversions; about their learning styles and prospects for growth.

Two ways of responding
We may, of course, want to temper our confidence by reminding ourselves that what we have seen is a product of the unique situation we have observed or arranged, may not be typical of what the student can or might yet do, and in any case it might be evaluated quite differently by other assessors. Nevertheless, however firmly or tentatively we regard our new-found knowledge, what are we going to do about it? Broadly speaking, as I mentioned at the beginning of this essay, we may use it:
- (formatively) to help the student learn or_
- (summatively) to give a mark or grade or written report that attests to the quality of the student's present abilities -- and perhaps compares it with that of other students.

Informal assessment (e.g. during classroom or laboratory work) rarely results directly in grades or marks, though it may influence the way we view a student's next piece of formal work. (We are all of us liable to the 'halo effect' of responding over-generously or

over-grudgingly to a piece of student's work on the basis of our prior impressions of that student.) Formal assessment (e.g. via exercises like tests, assignments, exhibitions, etc) almost always leads to some kind of summative response, even if this does not always count towards the student's overall course result. However, exercises that give rise to summative assessment may be planned with formative assessment in mind also. That is, the students are to be given a mark or grade, yes, but they also going to be given feedback with a view to helping them improve, either at the assessed task in particular or with wider aspects of the course.

The effectiveness of such an two-handed approach cannot be guaranteed. Much depends on timing and style. As one student told me recently:

"The main interaction students on this course have with tutors is in submitting assignments and receiving them back marked. This interaction is not a useful instructional interaction. The assignment is the full-stop at the end of a section of study, and feedback is generally not taken on board by student, doesn't alter student's views, and is generally seen as a justification of the mark given rather than as a useful interaction."

The formative response
This formative, or teaching response to assessment hinges around giving students useful feedback about what they have been doing or producing. Feedback is the lifeblood of learning. We don't improve at golf unless we get to se where the ball has landed. Learners need to know whether they are getting it right, whether they are improving, how to do better, what other people think of their efforts. Especially, they need a personal response from another human being who has understood how things seem to them and is able to say something that challenges or confirms their understanding, and helps them overcome errors or encourages them towards new insights.

How can you design your assessment system in such a way that each learner will get satisfactory feedback -- especially if you will not be the only one giving feedback to students? How, that is, can you ensure that the feedback is:

- speedy -- i.e. delivered while the student is still likely to care about the assessed work and have it in mind_
- selective -- e.g. not attempting to catalogue all the student's faults (or virtues) but focusing on two or three important ones that the student can do something about it_
- specific -- i.e. not making generalisations about the student's work or how it might be improved without drawing attention to specific examples_
- understandable -- i.e. expressed in language the student will understand, e.g. perhaps not in the sophisticated shorthand of your discipline_
- balanced -- i.e. pointing out some positive aspects of what the student has done (perhaps to start and finish with) besides those aspects that could be improved_
- concise -- i.e. not writing so much (perhaps rivalling the amount written by the student!) that the student is overwhelmed and left feeling inadequate rather than empowered_

- personal -- i.e. referring to what you already know of the student and her or his previous work (rather than commenting as one might to any stranger who had produced the work in question)_
- delivered via the most appropriate medium -- e.g. might some feedback be more acceptable or better understood if given face to face or on the telephone rather than (or as well as) in writing?_
- delivered by the most appropriate person -- e.g. might some feedback be given more speedily, more often or more understandably by other students rather than by a teacher?_
- followed up on -- i.e. who is keeping records (mental or otherwise) so that signs of improvement along the suggested lines can be looked for in the next assignment?

If you are responsible for, or are hoping to influence, the feedback given by other assessors, how might you do so? In much the same way, perhaps, as you might influence their assessing and marking -- that is, by discussing with them both general principles and specific examples of some useful and less-useful feedback. An 'agreement exercise' of the kind mentioned earlier for arriving at common marking standards could be extended to enable colleagues to compare how they would comment on the given examples of students' work.

Finally, it may be worth remembering that, in gaining knowledge about our students' learning, we may have gained knowledge about ourselves also -- especially about where we may have led students astray or been less clear in our teaching than we would wish. Hence we may want to apply the formative response to ourselves too and consider how we might improve our own work.

The summative response
This is the aspect of assessment that many teachers find most irksome, difficult or downright distasteful. As one of my colleagues said:
"I suddenly thought. What am I giving these marks for? What do they mean?"'
You might wonder first, for example, whether the mark is meant to indicate the extent to which the student has attained the learning outcomes for the course or for a particular segment of it. This would be what is called criterion-referenced testing -- where, theoretically, all the students could score 100%. Or should it indicate how the student compares with the other students currently being assessed (or who have been assessed in the past or who might conceivably be assessed in future)? This so-called norm-referenced approach implies a bell-shaped distribution in which few current students can be expected to get anything like full marks, most will be distributed around a middling range of marks and a few may be expected to fail.

Second, you might wonder what your marks would mean to your colleagues -- especially, might they censure you as an unduly lenient or strict marker, or as one whose standards were erratic? How, indeed, does one know what standards to apply? Students often assume that all teachers in a department will have a natural and common understanding of what any piece of work is worth, and new teachers may have similar expectations, but

countless studies of reliability suggest otherwise. Where standards are held in common it is often because of extensive discussions in which the different values and perceptions of colleagues have been set out and debated and consensus painfully constructed.

Third, you might wonder what the marks mean to the student; you might wonder what they mean to other interested parties, e.g. employers or other teachers. Students will no doubt be comparing them with marks they've been awarded on previous assignments or assignments in other subjects, or with the marks of their peers. But the marks won't mean anything specific enough to be actionable unless they are accompanied by the kind of verbal feedback mentioned above. Other people may wonder what the mark is actually telling them: How does the student with 70% really differ from one with 80%, and how is the gap between them different from the gap between students with 70% and 60%? And what do any of them actually know and have the ability to do?

However, even if you have misgivings in this area yourself, there may be little or nothing you can do to change custom and practice in your own institution or even your own department. Like it or not, you'll probably find yourself puzzling over how to award fair letter grades or percentages (or negotiating with other assessors about how to do so (as reported by a teacher in Extract 4). From time to time, you may even find yourself in the uncomfortable position parodied, with uncomfortable accuracy, by Laurie Taylor in the discussion between two assessors below (Taylor, 1980):

'Now candidate 666. 1 found this very jumbled. My notes say "very jumbled -- lacks overall coherence -- little sign of organisation -- no evidence of planning -- somewhat repetitive phrasing." So I went for nothing more than a compensatable pass, 37 to be exact.'

'This is 666?'

'Yes.'

'Well I must say I've taken a rather more charitable view here. I agree about the lack of organisation, but there seemed to be some attempt to be original, some sign of getting away from the standard material. Even a little imagination.'

'What have you got then?'

'Pardon?'

'What mark have you got?'

'Emm ... well I've put down 86 -- although with a question mark after it -- so obviously

I'm prepared to move a bit.'

'Quite a gap. But at least we both seem to agree on a pass.'

The conflation problem
Deciding or agreeing a fair summative score or grade for a single piece of student's work can seem taxing enough. But, in addition, we are often required to combine or conflate a number of such summative scores so as to produce an overall grade for the course or even a programme of courses. This opens further opportunities for conflict between people's differing values and judgements as well as for questionable statistical manipulations.
As an example, consider a student who has earned the following separate grades on a year's course:

Essay 1: C

Essay 2: A

Essay 3: A

Essay 4: A

Multiple-choice test: D

Project: B

Final exam: C

Different teachers might well want to conflate these grades quite differently from one another. Here are just a few of the many possible ways of deciding the student's overall result; they might give the student:

A -- because that is the grade the student gets most frequently

B -- because it 'averages out' the three grades that were higher and the three that were lower

B -- because a project should be a truer guide to the student's worth than any of the other components

C -- because the final exam is what really shows a student's true worth

B+ -- mid-way between project and exam scores (because we can presume the earlier exercises were subsumed by the two main ones).

Of course we can't tell whether any of these rationales might be reasonable without knowing in detail about each of the assessment exercises and the part it played in the overall strategy. And other rationales are possible. The letter grades might be converted to numbers, different weightings might be given to different components and a further

variety of complications and potential injustices might be introduced (see Rowntree, 1985, pp 226-42).

Conflation is clearly not a transparent procedure. It requires a set of rules that have been decided in advance and, preferably, discussed with the students so that they know how to direct their efforts. Indeed, if this is not done in advance it leaves open too many ways of favouring some students over others.

Assessment Board meetings, where colleagues meet to decide how to convert the results of a number of different asssessments into the overall grade for the course or degree -- what has been called the 'all-talking, all-singing, all-dancing, uni-dimensional grade' -- often enable us to see standards shifting before our very eyes. We may see all students' marks for a particular assessment item being changed because the average or spread of marks was so different from that for similar items. We may see the mark boundaries for excellent degrees being increased out of fear that we are producing more excellent degrees than our external evaluators will think feasible. We may see the pass-mark being lowered out of fear that we are producing so many poor ones that our potential customers will enrol elsewhere. You may or may not have a voice in conflation policy (see Appendix 3), but it is as well to be aware of its implications for your students and how you might choose to asses them.

As for whether the result of the conflation -- can seriously be held to convey any worthwhile information, I still echo the scepticism expressed some years ago by Sussex University's Working Party on BA Degree Assessment :

'. . . prospective employers should be warned . . . that the usefulness of the result is limited. First, because the predictive value of a particular class of degree cannot be guaranteed; second, because the actual measurement is extremely rough; third because the classified degree may be measuring qualities which are not necessarily relevant to the purposes an employer may have in mind' (Sussex, 1969)

Employers may choose to act as though a degree classification conveys useful information about the student's qualities or about how they compare with other students. But so much detail is lost (especially when numerous separate assessment grades are bundled together to decide the class of degree) that comparisons are crude in the extreme. What does a student with 2.1 in anthropology have in common (if anything) with students who have a 2.1 in accounting, geology or drama? To what extent can we expect the same kind of qualities from students with the same class of degree in the same subject from two separate universities? Indeed, can we even expect exactly the same body of knowledge, skills and dispositions from two individuals with the same class of degree in the same subject from the same department in the same year?

Many would argue that only a verbal profile -- describing verbally the observed qualities that the letter grade or percentage score so tersely represents -- could even begin to make clear what is special about a given individual.

Final remarks
Whatever doubts we may have about the ways in which assessment is used summatively, we can have no doubts about its formative power. Assessment is a major influence, perhaps the major influence, on what and how our students learn and on how much time they spend studying. Consequently, our best teaching efforts will be wasted unless we are operating an assessment strategy that encourages and rewards appropriate learning rather than inappropriate learning -- and part of that strategy must be to ensure that students truly know what is expected of them (both in assessment generally and in each specific exercise) and are not left to guess at our intentions.

In the accompanying Practice Guide we hope to work with you in developing a strategy that both generates appropriate learning and helps demonstrate to those who will assess us (the Quality Assurance teams) that our teaching and assessment have been effective.

References
Allan, J. (1996) 'Learning outcomes in higher education' pp93-108 in Studies in Higher Education, vol 21, no 1

Boot, R. & Hodgson, V. (1987) 'Open learning: meaning and experience', Chapter 1 in V.

Hodgson, S. Mann, R. Snell (eds) (1987) Beyond Distance Teaching&emdash;Towards Open Learning, SRHE & Open University Press, Buckingham

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

Boud, D. (1986) Implementing Student Self-assessment, Green Guide No.5, Higher Education Research and Development Society of Australasia

Edgeworth, F.Y. (1890) 'The element of chance in competitive examinations', pp 400&endash;75 and 644&endash;63 in Journal of the Royal Statistical Society, vol LIII, September/December 1890

Freeman, R & Lewis, R. (1998) Planning and Implementing Assessment, Kogan Page, London

Gabel, D.L. and Bunce, D.M. (1994) 'Research on problem solving: chemistry', in Gabel, D.L. (ed), Handbook of Research on Science Teaching and Learning, MacMillan, New York (See also: http://nievax.nie.ac.sg:8000/~wwwera/eproc95/p951a-3.htm)

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

Gibbs, G. (1999) 'Using assessment strategically to change the way students learn' in Brown, S. (ed) Assessment Matters in Higher Education; Choosing and Using Diverse

Approaches, Open University Press, Milton Keynes

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

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

Postman, N. & Weingartner, C. (1969) Teaching as a Subversive Activity, Penguin Books, London

Robinson, J. (1999) 'Anonymous peer review for classroom use: results of a pilot project', in proceedings of Teaching and Learning Forum, University of Western Australia, Perth WA

Rowntree, D. (1985) Developing Courses for Students, Kogan Page, London

Rowntree, D. (1987) Assessing Students: How Shall We Know Them?, Kogan Page, London

Sussex (1969) Final Report of BA Degree Assessment Working Party, University of Sussex, Brighton

Taylor, L. (1980) p.34 in The Times Higher Education Supplement, 3 October

Wisker, G. (1993) 'Now you see it, now you don't: external examiners and teaching quality' in Knight, P.T. (ed) (1993) The Audit and Assessment of Teaching Quality, DRHE/Standing Conference on Educational Development, Birmingham

Appendix 1 -- Case study: Planning assessment for a new course

Adapted from pp 212-216 in Developing Courses for Students, Derek Rowntree, Harper & Row London, 1981

In the following paragraphs, my colleague Roger Harrison and I consider how we might assess a new course on National Energy Policy: The course is intended for second year degree students, most of whom will be studying Economics, Business Management, or Planning as their main subject. There will be about fifty such students and the course is meant to occupy them about three hours a week for three terms -- about 90 hours in all.__It is vital to plan assessment with one's whole course in view. Otherwise, the plan may not do justice to the aims and intended outcomes of the course. Certain parts of the course (especially the early parts which are relatively easy, both for the student to learn

and for you to assess) may be over emphasised. So let us get an overview of the proposed Energy course. We could list the topics to be covered in the form of a syllabus. But perhaps you will see the topics in better perspective if they are shown embodied in the proposed aims and learning outcomes or objectives.__Aims__The course aims to explain the concept of energy (using practical demonstrations) and the importance of energy in industrial society, to provide practice in carrying out experiments and surveys relating to energy studies, to engender a sense of concern about the way in which reserves of fossil fuels are being depleted at the present time, and to give guidance in the planning of an energy policy.__Objectives

The student should be able to:
By the end of Term 1
- List the various forms of energy and describe how one form may be converted into another.
- Explain why energy is important in modern industrial society.
- Distinguish between energy and power.
- Perform simple calculations concerning energy transformation and power.
- Decide appropriate methods to determine the power of an energy transforming device.
- Explain what is meant by the conservation of energy and why energy appears to be lost in everyday situations.
- Explain quantitatively why there are limitations to the proportion of thermal energy which may be transformed into useful work.
- List the commercial sources of energy currently and prospectively available and estimate the reserves or potential of each.
- Identify the principal energy flows in a given community (or industry, or factory, etc.).
-

By the end of Term 2
- Identify the deleterious consequences which may result from a given industrial energy transformation proposal.
- Identify the energy inputs into a given product, using data provided.
- Evaluate a technical proposal with regard to energy consumption.
- Suggest ways of reducing energy consumption in particular contexts, and assess their feasibility.
- Identify the economic factors relating to the use of energy.
- Form a judgement as to the weight to be given to economic considerations in evaluating energy proposals.
- Suggest objectives for a national energy policy.
-

By the end of Term 3
- Apply the experience of terms I and 2 to real world situations demanding that the student exercise initiative, selectivity, understanding, judgement, and social skills.

In brief, the first term lays the scientific foundations; the second term brings in the economic and social issues; and the third term requires the student to apply the theories

and concepts he has acquired so far. How are we to assess such a course?

To begin with, shall we assess at the end of the course, during the course, or both? Well, the course is certainly one in which the student's final understanding should be greater than can be assessed by adding up all the partial understandings he or che has attained on the way through. That is, some integrative assessment activity seems to be called for. This could be a long essay towards the end of the third term. Or it could be a final examination -- perhaps a fairly open one (e.g., questions published in advance, reference works available, etc.).

However, perhaps we need assessment during the course as well? For instance, the third term's objective clearly suggests a project. In fact, we can decide that the third term will be devoted to an empirical enquiry related, for example, to objective 9 or 11 (identifying energy flows in a community or energy inputs into a product). Student will negotiate the precise context of their individual enquiries with their tutor, and will carry it out in a local factory, hospital, or whatever. Such a project will give rise to a report, itself a long 'essay' (with graphs, calculations, etc.). Clearly, this ought to 'count' for purposes of overall assessment since it relates to a vital learning outcome -- so vital that a whole term is devoted to it.

At the same time, if the students are concentrating on their projects during the third term they cannot be expected also to produce a long essay integrating what they have learned during the course. Nor can their project reports be expected to perform this integrative role. So, we had better settle on an examination at the end of the third term. Perhaps it will concentrate on objective 16, giving maximum scope for the pulling together of what has been learned both in terms I and 2 and during the term 3 project experience.

So we have decided on a project report towards the end of term 3 and an examination at the end of the course. Is anything more needed? Well, the scientific/experimental background in term I is rather vital and we would be unhappy to see students skimp it. In fact, we do not think students could make real sense of the rest of the course without it. So we will assess their attainment of the term I objectives two or three times during that term. Probably multiple choice questions will suffice; but, since the tests will be marked by tutors, short 'own answer' questions can also be included. While the objectives they will test are vital, we cannot pretend they are very high level. (They will not be tested in the final exam.) It would not be appropriate to let the tests contribute too much to the student's overall grade for the course. Let us regard them as a critical, qualifying tests. That is, students have to get satisfactory scores in order to be allowed to continue with the course, but the score will not count in deciding their overall grades. (The decision on what is a 'satisfactory' score can be left until we have worked out just what level of proficiency is essential for the rest of the course.)

Now, what about the second term -- do we need any assessment there? Well we do perhaps need to help motivate students in reading up on the 'economic/social factors (rather than spending undue effort preparing for the project). So some sort of assessment seems desirable. How about a long essay in the area of objectives 12 15? This, in many

ways, can be seen as preparatory to designing an energy policy which is to be tested in the final examination.

So here is the assesment plan we would include in our course proposal:

Term 1: Objective/short answer, 'qualifying' tests (objectives 1 7).

Term 2: Long essay (objectives 12 15).

Term 3: Project report (largely objectives 8 11). Final exam (largely objective 16).

The only other issue is one of weighting. That is, of the three forms of assessment that are to 'count' towards the student's overall grade for the course, which is to count most and which least? We can see no reason why either the long essay or the project report or the final exam should be given greater importance than either of the other two so let them all count equally. That is, the student who does poorly on the exam can still make good by doing particularly well on the other two components. So too can the student who does poorly on the project or the long essay. She or he does not have to do well on all three in order to do well on the course. But they do have to do well on all three to do well on the course. The criteria by which we are to judge whether students have done well or poorly will have to be argued out with colleagues once they have accepted the overall strategy.

But these are only our first thoughts on assessment for that course. We would expect them to be modified in discussion with colleagues, especially as the coverage and style of the course became more clear. But it is important to throw ideas about assessment into the discussion before the course structure begins to harden. Because, for example, the suggestion that we assess on the basis of a project report in the third term is, indeed, a suggestion also as to the structure of the course. Assessment is a direct expression of the intended style of teaching and learning.

Appendix 2 -- The assessors assessed?

The following comments were sent to me by a reader of my book, Assessing Students: How Shall we Know Them? (Rowntree 1987). For reasons that will become obvious, my correspondent must remain anonymous. She or he was writing from a higher education institution (the 'college') where assessments of students' work were 'moderated' by examiners from a neighbouring university. She or he refers to many points I made in my book and reflects on them in the light of her/his own experience.

I think one thing that will be very valuable is your analysis of the various types of assessment and of the various functions of assessment too. This is crucial since so many of the unintended social effects of assessment schemes seem to occur precisely because assessment items are so ambiguous. Sometimes the ambiguity seems deliberately used -- for example, in my present place of work it is not uncommon to "sell" assessment to students on the grounds of it being diagnostic and helpful -- and then to use the grades to

select those who will proceed to Honours courses, etc.

What makes this ploy particularly unpleasant is that essays designed with diagnosis in mind are often quite unsuitable for discriminating between students (in that they do not provide a wide enough spread). In such circumstances, the final examination (used as well) comes to be all important (since it does spread out students) despite its lesser official weight. My point is that assessment often is not clearly thought out and its function often is switched to a discriminatory one in a crisis (e.g. when "too many" students might qualify). The ambiguous nature of assessment makes it very difficult to spot the switch.

Another issue you discuss concerns the transformation of apparently open ended, divergent tasks (like projects and simulations) into convergent tests at the point of marking. We use both projects and simulations here and in both cases this transformation occurs. In free choice projects like dissertations, for example, a whole series of negotiations about titles and themes takes place, and gradually the work is shifted towards what lecturers want and expect. Again, when marking, our external University markers do seem to have a rather narrow "right" project in mind, for marks are added to (or more usually subtracted from) our internal College marks according to apparently hard and fast rules of procedure. Naturally, none of this is made public and, as a new member of the Examiners' Panel, I was apparently expected to know what was required without the vulgarity that would arise if they actually had to make their rules explicit.

I was pleased to find the bit about "maximum marks only for God" for this is our University's attitude too. In their case, this implicit belief that no student could ever get more than 85%, and that there exists in each case some Platonic "perfect" dissertation, has an important consequence. On paper it seems that students have to get an average of 70% over all their assignments to get a first class degree. However, when the "maximum marks only for God" clause is invoked, it becomes impossible to get more than 85% in any one assignment. In effect, then, the student has to get 70 marks out of an available 85 (i.e. 82% in real terms). Crafty, no? And again, ambiguous enough to fool people and to allow all kinds of ad hoc rules, concepts of "normal numbers of firsts" and so on to be introduced.

My own modest appeal to these wise men to spell out just what they see and how they respond was seen as awfully embarrassing and rather radical (which was the intention, I confess -- once out in the open the criteria would not stand up to the slightest criticism and we all knew it). Instead, I was asked to be pragmatic and silent. Indeed the whole affair was rather nasty, and dis crepancies between my marks and everyone else's were seen as evidence of incompetence or of subversion and I was hastily removed from the Examiners' Panel (and unpromoted for a year!)

Your own cool and sensible words on the inevitability of disagree ment among markers would be well appreciated in this college. Again, I was in a position to see how inter marker agreement is actually obtained. It Is "negotiated" among the individuals who

differ. Nothing is explicit, but a wide divergence is seen as em barrassing and one only makes that mistake once. Next time, one takes every precaution to offer for moderation marks that will be within 5% or so of what the University examiners would award. Experience at "pysching out" the examiners works well at our level too! (So does fear of failure!) When in doubt, we play safe and undermark. Again, a kind of semi secret normal curve system operates -- we just don't offer too many papers over the magic 70%, nor do we fail many. No one will say what the curve should look like, no one employs the routine statisticaI analysis to ensure the curve actually is as expected, no one designs assignments to produce a normal curve of results, no one publicly advocates or defends such a practice.

In any case, the University panel have recently emphasised their own divine right to "adjust" our internally awardcd marks by insisting we mark scripts only in pencil. How's that for a new category or status -- we college hacks are mere markers in pencil!

Doubtless our own little system is no worse than anyone else's, as your examples seem to show! Interestingly, your point about the trend towards making questions explicit is one I am pursuing myself at the moment. My own questions are usually specific and detailed. As you say, such techniques indicate convergence and reward it -- but at least the process is an honest one! Students are awarded marks according to how well they match the "right answer" in my head -- so I explain what it is I want as clearly as possible. Lacking political power to change assessment, my own solution is at least to try and show everyone how con vergent it really is, underneath all the liberal stuff about essays, free choice dissertations, etc.

Many of my colleagues recoil from the thought of using explicit norm referenced techniques -- but they fail to realise that the same logic is used whenever they decide to be "realistic" at examiners' meetings. There is also the desire to be "objective" and "fair" too -- but not to the extent of publicly using multi choice tests, etc. The most paradoxical colleagues for me are the English teachers, who advocate the search for personal (intangible) meanings and who also set and mark exams and grade just as rigidly as everyone else. What interests me is how they manage to explain the contradictions to themselves.

One final point. Your insistence on the validity and relevance of assessment is excellent. Your advocacy of self assessment is very welcome too. However, the proposal to assess students using a profile system has one danger, it seems. The danger is that the search for greater validity will lead to the universality, the omnipresence of assessment. Humans are complex -- but the attempt to capture that complexity in a fully valid profile might lead to a new totalitarianism with even more activities coming under the assessors' gazes. The alternative I prefer (at the moment) is the one you close the book with -- to put a warning on academic grades (and profiles). This, I think, is to make the hidden curriculum explicit; I tell my students about the work of Snyder and Becker, etc., partly so they won't be taken in by the assessment they receive or administer themselves when they become teachers. An A grade then should lead no one to think he is perfect, or more worthy as a person, etc., -- nor should an E grade have the reverse effect. Anyway, Becker did find, I

recall, that cynical (I'd prefer realistic) students did well in tests and tended not to suffer so much psychologically.

Anyway, I must go now -- to mark some essays!

Yours from the backwoods, Dr. Anon

_Return to Derek Rowntree's [Home Page](#)

If you want to send comments, please write to d.g.f.rowntree@open.ac.uk