Chapter 4  Assessment and degree programmes

Prelude

You might feel that I am making heavy weather of assessment issues in higher education. One way of responding to such scepticism is by arguing that the problems I see with higher education practices are familiar in other areas. If I can show that, then it is easier to argue that higher education assessment problems are basic in the sense that they attach to all attempts to appraise complex achievements.

Professor Patricia Broadfoot edits the journal Assessment in Education, which mainly prints papers on the assessment of schoolchildren's learning. In my opinion, research in this area is longer established, more extensive, better funded and, in many ways, of higher quality than research into assessment in HE. She argues (2000) that

... the emphasis on a 'measurement-driven' pursuit of higher standards ... has not been successful in raising overall standards of student achievement [nor] in enhancing the quality of teacher education and professional development more generally ...[t]he prevailing obsession with the pursuit of 'objective' measurement has constrained education to a point where fundamentally important, but necessarily more amorphous forms of learning are being excluded.

(p. 200)

One reason why she is uneasy about the 'assessment society' is her critique of

... the modernist assumption that it is possible to 'measure' quality by applying criteria to the available evidence and, on the basis of this evidence, to form a judgement. It is further typically assumed that it is both appropriate and desirable to express such judgements in the form of grades or marks ... quantitatively, or 'categorically,' rather than descriptively, in order to provide a common basis for comparison.

(p. 207)

This is '... despite the now enormous body of research literature documenting the inherent vagaries of what is, inevitably, a process shot through with human subjectivity' (p. 209). To it she adds the comment that

Existing approaches to assessment are almost exclusively concerned with explicit learning, with what has been consciously learned and reproduced in a formal setting. However, the goals of learning are likely in the future to centre increasingly on the acquisition of attitudes, skills and personal qualities.

(p. 212)

Those sorts of learning outcome resist measurement. Her observations repeat the critique I have made of the idea that assessment = measurement, notably in my treatment of reliability in Section 1.5. I introduce this chapter with this summary of her work because it forces us to address a major assessment issue: if most assessments will not produce valid and reliable measures of achievement, then how is assessment to be managed?
4.1 The structure of this chapter

In this chapter I am elaborating some of the ideas I used in Chapter 1. I shall look more closely at measurement theory, repeat my claim that many of the complex learning outcomes that we value are resistant to affordable, reliable and valid assessment and describe a more differentiated programme assessment system. I begin by returning to learning outcomes. In Chapter 3, I summarised the advice that is routinely given on how to write learning outcomes for local, module-level assessment purposes. When the focus moves to programme-level, new issues arise. This is because there is an expectation that outcomes will be developed across a programme and lead to generalisations about learners' competence in respect of each programme outcome. That assumes that judgements about any outcome that are made in different modules are compatible and can therefore be aggregated to produce some generalisation about achievement across the whole programme. This raises some technical problems, although I have already pointed to some solution strategies, such as using the same grade indicators to help judgement of the same learning outcomes wherever they appear in a programme. Yet, this strategy may be harder than it looks because, as was noted in the last chapter, statements of achievement have to be written for different degrees of proficiency (First class, Lower second, Pass, etc.) and at different levels (first year, final year; certificate, honours degree). I am going to try and pick at some intractable issues in the following way:

1 I want to have a closer look at the idea of learning outcomes.
2 I shall then raise some questions about the idea of competence and its certification.
3 It will then be helpful to return to outcomes and show how hard it is to produce programme-level statements of outcomes-related achievement that are useful for cumulative assessment based on judgements made in several modules.
4 Having deconstructed common-sense ideas about outcomes-based judgements of competence arising from a programme of study, I shall offer an alternative view of programme assessment systems that centres on three ideas:
   - Warrants
   - Claims
   - Process standards.

4.2 Learning outcomes again

I mentioned in Chapter 1 that there is a view that module and programme designers should start with assessment and work backwards to decisions about teaching, learning and curriculum sequence. By that, I meant that the first decision for planners to make would be about what they want to be able to assess at the end of the module or programme — about the assessed learning outcomes. Figure 1.3 embodies this view. Attractive though this is and sensible though it seems, many higher education teachers have found this outcomes-and-assessment approach to be a difficult one to follow. I think that part of the problem has been that they have been operating under the influence of an impoverished view of assessment (namely, that assessment equals measurement), but the idea of learning outcomes has contributed much to their difficulties. The review of learning outcomes in Chapter 3 is complemented here by a short critique of mainstream thinking on learning outcomes, which prepares the way for a fresh look at some core themes in programme-level assessment.
Stenhouse (1975) trenchantly attacked the idea that schooling could be geared to objectives. He pointed to severe problems with the early twentieth century rational curriculum planning view that if schooling was a preparation for adult life, then it could be designed by identifying the important features of being an adult, expressing them as learning objectives ('the student will be able to ...'), teaching to those objectives and assessing students against them. Problematic features were the assumption that complex social practices could be reduced to objectives; the requirement that objectives be stated unambiguously in terms of behaviours; the tendency of objectives to bloat and divide, leading to unwieldy lists of objectives; the assumption that learning is straightforward enough for it to be possible to specify what needs to happen for an objective to be reliably achieved; the belief that all worthwhile learning can be pre-specified and neatly pinned down by objectives; and problems of assessing achievement efficiently, reliably and validly.

Stenhouse's view was influential until the UK government decided to introduce a National Curriculum based on managerialist principles that, unless effects can be measured, a system cannot be controlled and improved. At that time, a group of higher education teachers working with the UK Employment Department was exploring the usefulness of 'learning outcomes' for higher education (Otter, 1992). The difference between learning outcomes and objectives is not clear from the report and it is probably fair to say that 'learning outcomes' is only a more neutral way of talking about learning objectives, since Otter's view was that:

- 'outcomes should be framed in terms which made them capable of assessment' (p. 5).
- '... outcome definitions which are not assessable are not of any practical value' (p. ii).
- outcomes should be the product of an 'analytical process (functional analysis)' and have 'sufficient detail to indicate the level of work' (p. 13).
- outcomes should be written with clarity and precision (p. 29).

This is open to the charges that Stenhouse laid against the learning objectives movement, doubly so if higher education is concerned with complex learning in which certainty is more elusive than it is with school learning. If interpretation and dispute are central features of many disciplines in higher education, then there are interesting problems for those wishing to apply the determinate language of learning outcomes. Furthermore, the Otter report was quite open about the potential that the assessment of learning outcomes has for the management of quality, an implication being that outcomes assessments would provide performance indicators for a new higher education managerialism.

Interestingly, although Otter's report acknowledged that assessment is pivotal in a learning outcomes approach, her handling of assessment issues is not confident. For example, there appears to be a distinction between assessment and grading (p. 51), which may be intended to represent the difference between norm-referenced and criteria-referenced assessment, since she claimed more than once that norm-referenced grading practices were usual. I find this significant because it suggests an unsure grip of assessment theories. Worse, the statement that '... an outcome-led approach requires staff to develop and use methods of assessment which measure achievements directly' (p. ii) assumes that:

- all learning outcomes are assessable
- assessment is measurement
- the development of new assessment methods will be sufficient to fix any assessment problems: 'Development work is required on techniques for
assessing some of the more complex and difficult of the outcomes of higher education' (Otter, 1992, p. iv).

Points such as these are compounded by uncertainty about the distinction between objectives, outcomes, aims and competences, and by disagreements about the inspiration for these constructs. Although more recent work, notably that by Gosling and Moon (2001), clarifies matters in some respects, it complicates them in others. Aims, they say, are descriptions of what the teacher will do or teach, with outcomes describing what students will be able to do. However, they point out that there are programme outcomes, which identify what students ought to be able to do at the end of the programme; module outcomes, which will overlap with programme outcomes; and then assessment criteria, which describe the outcomes that attract grade A as opposed to grade B. Furthermore, module outcomes and grade criteria need to be written at a particular level – first-year undergraduate work is typically level 1, final year work level 3. This is complicated: programme outcomes, module outcomes, levels, and assessment criteria all need to mesh. Designers therefore face real problems in trying to introduce enough fine and valid distinctions to sustain this hierarchy of differences. What is really involved in taking a programme learning outcome, such as ‘critical thinking’, and working it out as module learning outcomes for levels 1, 2 and 3, each of which then yields five (or more) grade criteria? Notice the assumption that there are content-free hierarchies of achievement, so that we can say in general terms how Pass performance critical thinking at level 1 is inferior to Pass performance at level 2 or second-class performance at level 1, and so on. I do not want to go very far into this because it quickly brings us to philosophical and psychological issues that could overwhelm attempts to concentrate on good assessment practices. I do want to say that the assumption is questionable: for example, the view that there are hierarchies of mathematical concepts has been challenged (Hoyle, 1990), as has the related assumption that there are general factors in human performance that can be meaningfully separated from specific factors (Anderson et al., 2000).

Even if these more abstract (but potent) objections are set aside, there is a real danger of outcome/criteria statements proliferating like a plague.

I make one further point, which is that the more complex and difficult outcomes are not susceptible to measurement (see also Broadfoot, 2000, and cited in the introduction to this chapter). The Otter report is one of the most significant in a sequence of advice about assessing learning outcomes that creates intractable problems by implying that assessment, in the sense of measurement, is desirable and possible if only enough effort is combined with enough techniques. This sort of thinking was behind the third phase of assessment shown in Figure 1.1 and has left its mark in the advice prepared by Gosling and Moon (2001). Underlying this sort of stance is the view that the things to be assessed are determinate - fixed and precisely definable. This is contentious. Consider, as colleagues teaching practical and vocational subjects have to, the concept of competence. This concept, certainly as it applies to the sorts of achievement we associate with graduates, is less determinate and more fuzzy than everyday language suggests. If I am right, it is easy to see that ‘measurement’ and ‘competence’ are not kindred concepts because competence is not ‘real’ in the sense that it is not objective and invariant. It follows that no amount of skill at writing learning outcomes will reconcile; ‘measurement’ and ‘competence’, except in those cases where ‘competence’ describes simpler, determinate achievements. But higher education is not primarily in the business of stimulating such achievements. As I suggested earlier, this analysis pushes us to take a differentiated view of assessment, one that can handle the idea that it is hard to conceptualise competence and therefore no easier to assess it.
The problematic notion of competence and its assessment

The more complex the thing we are trying to assess, the more expensive it is to get tolerable levels of reliability without seriously harming validity – if, indeed, we agree that it is at all possible to get reliable judgements. This is a problem that looms large in many subjects, particularly, but not exclusively, those concerned with professional practices.

Reflection 4.1 Competence

Another advance organiser.

1. What do you understand by ‘competence’?

2. How would you know that someone was competent at something?

3. Let me challenge your response to question 2. How would you know that someone was competent at something? What degree of certainty would you want, how much evidence and what sort of evidence would you want?

4. How much would you be prepared to pay, in terms of teacher time and effort, to be tolerably sure about (a) simple competences [typing speed], (b) low-stakes competences [is aware of own value systems], (c) complex competence [skilled at designing buildings to meet client needs and resources], (d) high-stakes competence [is a good mathematics teacher]?

5. Do you trust warrants of competence [for example, bachelor’s degree classifications]?

6. Do you find warrants of competence useful?

7. What are the implications of your answers to questions 5 and 6 for HE assessment practices?

8. Note that we are even less well placed to talk about ability, which is an uncertain inference from judgements of present performance. ‘Ability’ is a word best avoided, I think.

So, I am not interested in simple competence, in determinate achievements that may be a part of higher education – memorisation; routine use of procedures, formulae and tools; transactional writing – but which are not its quintessence. When I refer to competence, I am thinking about regularly showing complex achievements that experts judge to be of appropriate quality. This is rather an ambiguous definition, which is worrying for those who want to measure competence and certify that those with enough of it are fit to practice or warrant that they have robust, transferable skills.

One approach to the definitional problem has been ‘functional analysis’. It became central to the development of national vocational qualifications (NVQs) in the UK and has influenced some professions, notably schoolteaching. The idea is that you define what a competent professional does and then break that down into its component competences – nearly 30 for schoolteaching. The assessment of competence is, in one sense, simplified, because rather than trying to make global, uncertain and difficult decisions about competence, judges have the more manageable job of weighing evidence about the individual components of competence. Criteria can be written to guide the assessment of each element and, once the criteria for each element have been satisfied, competence can be warranted, certified or proclaimed. This is a rational approach, rather like that used in natural sciences where complex wholes are broken into their components for analysis. It fits quite neatly with the rational planning-by-
objectives approach to curriculum design and assessment and produces a step-wise method for making complex judgements. Of course, there is a price to pay for the greater certainty that comes from treating the assessment of competence as the assessment of a set of separate, simpler, sub-operations, namely that this can be a lengthy process for assessors who may have, say, 27 'competences' to judge and for students who have to plan to provide evidence of achievement on 27 fronts.

Another objection is that competence may be something other than a fistful of competences, or elements of competence. Just as 'skills' may not be transferable but we may learn how to use social practices in novel situations, so it may be that the competent person has learned how to orchestrate the separate elements and deploy them appropriately. For example, learning how to brake, steer and change gear as separate operations is not sufficient for driving well, because that needs the elements to be mindfully orchestrated. This reminds me of what medieval scholars called the fifth essence (air, fire, water and earth were the other four), an unseen, undetected animating principle. The point is that, if there is a quintessence (the word means 'fifth essence') of competence, how is it to be recognised? assessed? measured?

Worse, there is a view that competence, like expertise, is 'embodied' and embedded in what we do — largely tacit, if you like. In other words, the competent person might be able to act well without explaining clearly the reasoning behind the actions. This poses a massive assessment problem because it means that direct observation would be necessary to gauge competence and that the assessor would have to intuit correctly whether an action was competent or not. However, it is not necessarily easy to infer competence from behaviour, odd though it may seem to say so. For example, as a teacher I give no sign of hearing a child calling for my attention. Is this evidence of incompetence (because you infer that I am insensitive to what is happening in my classroom) or of massive competence (because you infer that I have heard and made a professional decision that it would be a mistake to respond to inappropriate, calling-out behaviour)? There is a further complication. Experts, it seems, do not always follow the rules. Rule-based solution strategies, the ones that medical texts on diagnosis teach, for example, are for novices. Experts tend to use them only when they face novel problems or ones that have not been cracked by their usual approach, which is to compare the case before them with ones they remember to be similar in some important respect. The implication is that good practice will sometimes amount to showing text-book behaviours and sometimes not. So, however skilled the observers, judging competence is hardly going to be an objective process because it depends on high-inference judgements of behaviours that can be ambiguous.

Nor can we step around this inferential problem by saying that learners taught by the competent teacher get better than expected results, if only because many factors intervene between teacher action and student learning. It does not take long to see that you cannot assess teaching quality on the basis of learners' raw scores, because achievement on exit from a school or programme relates strongly to achievement on entry. One response is to calculate added value scores, that is the difference between exit and input grades. This is much better, but researchers into school effectiveness find that it is not just teaching quality that affects the amount of learning that happens, with the result that learners with similar entry scores are liable to get different exit scores, even if their teachers are equally good.

The point can be generalised to the assessment of other forms of competence. Although we might like to judge on the basis of results, there are so many mediating factors that lie between the professional's action and the clients' well-
being that it is hard, perhaps impossible, to determine how far the professional can be praised for client outcomes. Or excoriated.

The implications for the assessment of competence are clear. They are also unhelpful because there is a widespread view that we should attribute client outcomes to professionals and that good assessment techniques should allow us to measure outcomes with confidence. Managers and policy-makers tend to be impatient with the claim that technical solutions can only reduce uncertainty, not eliminate it.

If we want to make general statements about competence, we must be confident that assessment processes are reliable enough for us to trust the statements. Here it is important to recap a point made in Section 1.5 about the need for multiple judgements if we want to reach reliable conclusions. The more complex the achievement and the more valid the approach to assessing it, the more judgements we need if we want good levels of reliability because valid judgements of complex achievements will each tend to be unreliable. The more we want to reach reliable conclusions about complex achievements, the more individual judgements we need to have to hand. But these high-inference judgements of complex learning are also expensive and can be time monsters, which means that there has to be a programme assessment plan if we are to have enough judgements to allow us to be acceptably confident in our final reasoning about competence and other complex learning achievements.

A further difficulty is elaborated in Sampling the curriculum, below. Assessments, especially exams, are usually samples of what we want students to learn. What are the implications for the assessment of competence of basing warrants and certificates on assessments that only sample the domain?

Sampling the curriculum
From learning outcomes to learning claims

This section is a more theoretical analysis of the themes that were explored in a fairly practice-centred way in Section 3.2. The suggestions in that section are fine as far as good local assessments are concerned but once we say that a programme produces graduates with achievements that the university will warrant, then this good sense founders. Tricky theoretical issues have to be addressed and, in my view, good teachers come to accept the limits of (reliable, affordable, summative) assessment.

A fundamental point is that the nature of things affects what can be known about them (these are points 1 and 2 in Problems with high-stakes assessments of complex achievements, above). Some material things can be known with a great deal of certainty using shared language. They are determinate. Many human constructs can be treated in the same way, as if they were real things like those in the world of matter. Others are far less determinate, being inherently fuzzy, complex, changing and subjective. As Grint (1997, p. 13) reports, the more complex things are, the harder it is to say things about them that are precise and important. The more precise we try to be about fuzzy, complex, changing and subjective achievements, the more we distort them by slicing away their essences – fuzziness, complexity, changeability and subjectivity – to do so.

Remember here that higher education has been encouraged to concentrate more on ‘mode 2’ knowledge, which is messy and contexted, a blend of propositional, ‘know that’ knowledge and procedural, ‘know how’ knowledge (see Section 1.2). These learning outcomes are especially complex, particularly fuzzy and non-determinate. Recall too the USEM (understanding, skills, efficacy beliefs, metacognition) model of a curriculum to enhance students’ claims to employability, which was introduced in Section 1.5. It says that an employability-enhancing curriculum promotes understanding of subject matter, social and academic practices (‘skills’ if you prefer), efficacy beliefs and metacognition. Professional competence can easily be understood as a conjunction of these elements. Much, perhaps all of the SEM (skills, efficacy beliefs and metacognition) is not real and determinate. Indeed, what counts as efficacy, critical thinking or reflection is disputable: what I call critical thinking you call metaphysical twaddle; what you call efficacy, I call following orders; and I deny that any old thinking counts as reflection but I’m not sure how I would define ‘reflection’.

So, even assuming that it is legitimate to encourage and assess SEM, there are obvious difficulties in seeing how we might fairly assess such fuzzy outcomes. And, if we try to do so, there is a grave danger that in writing learning outcomes to be reliably measured, we will transform the complex essences that the outcomes are supposed to describe, especially if we try to write outcomes at many levels, as described in Section 4.2, above.
There are three options:

(a) Go down the precision/reliability route at the price of damaging what you claim to assess.

(b) Reject the learning outcomes approach, but at the price of having nothing to say about the assessment of authentic higher education achievements.

(c) Develop a new theory of the assessment of non-determinate human achievements.

The stance taken affects the way in which outcomes are written. Stance (a) pushes writers to specify what the learner will know or be able to do at the end of a successful learning sequence. The language is often behavioural, leading to outcome statements such as, ‘At the end of two class periods, the student will find inverses to Laplace transforms with non-repeated linear and quadratic equations’ (Heywood, 2001, p. 36).

Plainly, this specificity implies that a lot of outcome statements will be needed. Since it would be unwieldy to try to report at this level of specificity, we would need to construct some more general categories, which would mean producing rules for aggregating data for lower-level outcome statements. These rules assume that it is permissible to combine data from different sources to get a category score (although that is rather like adding numbers of lemons and numbers of red currants to get numbers of pieces of fruit). Marks are often weighted as well, although decisions about whether data from coursework are to count as much, less, or more than examination data will favour some students and work against others.

The Student Assessment and Classification Working Group (SACWG) has produced a series of papers showing that grading processes in English universities do not produce reliable, comparable and useful assessment data. Diversity is rampant. Criteria are interpreted differently in different departments (Woolf and Cooper, 1999). Marks that get one classification in one university could be weighted, combined and processed in a different way in another university, leading to a different class of award. See Yorke, Bridges and Woolf (2000) for a recent example of SACWG’s work and read Combining marks below for an amplification of this point.

Combining marks

Even if you have marks that are reliable and valid, and can make sense of them, more problems arise when you add together marks from more than one assignment or test. Extract 4.1 examines several of the difficulties relating to allocating and interpreting marks, and in particular the difficulty of combining marks from tests or assignments which are different in nature. If one of your assignments allocates marks for students’ ability to write and another allocates marks for the number of correct answers to largely factual questions in a test, what does the combination of marks from these two assessments mean? To make this problem explicit, imagine that Student A scored 40 per cent and 60 per cent respectively on the two assessments and that student B scored 80 per cent and 20 per cent respectively. What would both their average scores of 50 per cent mean? It is a little like trying to add apples and pears. Without a ‘profile’ statement showing the different abilities of these two students, the raw score of 50 per cent means rather little. If you had weighted these marks differentially the students’ averages would have been different without them actually performing any differently. For example, if you had allocated a quarter of
your course marks to the writing assignment and three-quarters of your marks to the factual test, then the two students would have had overall course marks of 55 per cent and 35 per cent respectively.

If you had reversed these weightings the students’ overall course marks would have been 45 per cent and 65 per cent respectively (see Table 4.1). Notice that if you change these weightings not only do the students’ averages change dramatically and their rank order reverse, but the average marks for the course (for these two students combined) also changes, from 45 per cent to 55 per cent. Suddenly your course looks more successful, solely because you have weighted the assessment components differently! None of these changes in averages had anything to do with the standards the two students achieved on the two assessment components. When exam boards look across courses and add up marks to produce degree classifications they are usually unaware of the nature of the assessment components which produced the marks.

Table 4.1 The effect of combining marks for assignments with different weightings

<table>
<thead>
<tr>
<th>Marks on writing assignment</th>
<th>Marks on factual test</th>
<th>Average if weighted 50/50</th>
<th>Average if weighted 25/75</th>
<th>Average if weighted 75/25</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student A</td>
<td>40%</td>
<td>60%</td>
<td>50%</td>
<td>55%</td>
</tr>
<tr>
<td>Student B</td>
<td>80%</td>
<td>20%</td>
<td>50%</td>
<td>35%</td>
</tr>
<tr>
<td>Average (A+B)</td>
<td>60%</td>
<td>40%</td>
<td>50%</td>
<td>45%</td>
</tr>
</tbody>
</table>

Once you add scores from different courses with different aims and outcomes, it becomes even more confusing. For example, what could the combined marks from a history dissertation and a French language exam tell you about the students who have studied both courses?

Even if you could make sense of such a combined score, the statistical problems you face are overwhelming. In the worked example in Extract 4.1 that follows, the combined scores students get after adding results from different exams, and their overall rank order in their class, are determined by characteristics of the distribution of scores in each exam. By simply including an exam with a wider range of marks the students’ overall ranking in this example is actually reversed and every student ends up with a different degree result! Students suffer all kinds of arbitrary consequences and unfairness not as a result of arbitrary marking but simply because the average marks and the marking scales used for individual assignments and on individual courses differ.

In the book from which Extract 4.1 is taken, Brown et al. go on to demonstrate that there are simple statistical solutions to most of these problems, but as these are seldom if ever applied we have omitted them. Some institutions have started using the same marking scale across all courses and assignments with descriptive general criteria for each point on the scale, and this will help avoid some of the scaling problems.
The only sensible conclusion is that you should not be too confident about what marks, and especially combined marks, actually mean. You should use your common sense in interpreting marks and especially when making important assessment decisions based on them – for example, decisions about whether a student has passed a course and can progress to the next one ...
... Figure 15.2 shows the effects on the total marks awarded when one assessor uses a narrow range of marks and another uses a wide range. The point is not merely theoretical. If a student takes a module that has a different range of marks from other modules then his or her degree class can be affected. The problem usually occurs when an optional module is taken outside of the main programme.

<table>
<thead>
<tr>
<th>Course A</th>
<th>Course B</th>
<th>Course C</th>
<th>Total</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>58</td>
<td>62</td>
<td>40</td>
<td>160</td>
</tr>
<tr>
<td>2</td>
<td>68</td>
<td>42</td>
<td>45</td>
<td>155</td>
</tr>
<tr>
<td>3</td>
<td>60</td>
<td>40</td>
<td>50</td>
<td>150</td>
</tr>
<tr>
<td>4</td>
<td>31</td>
<td>59</td>
<td>55</td>
<td>145</td>
</tr>
<tr>
<td>5</td>
<td>48</td>
<td>32</td>
<td>60</td>
<td>140</td>
</tr>
</tbody>
</table>

In course C the examiner has awarded marks ranging from 40 to 60. Now see what happens if the examiner had used a mark range from 20 to 80:

<table>
<thead>
<tr>
<th>Course A</th>
<th>Course B</th>
<th>Course C</th>
<th>Total</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>58</td>
<td>62</td>
<td>20</td>
<td>140</td>
</tr>
<tr>
<td>2</td>
<td>68</td>
<td>42</td>
<td>35</td>
<td>145</td>
</tr>
<tr>
<td>3</td>
<td>60</td>
<td>40</td>
<td>50</td>
<td>150</td>
</tr>
<tr>
<td>4</td>
<td>31</td>
<td>59</td>
<td>65</td>
<td>155</td>
</tr>
<tr>
<td>5</td>
<td>48</td>
<td>32</td>
<td>80</td>
<td>160</td>
</tr>
</tbody>
</table>

As one further example of the differential between course work and exam assessment, some researchers into boys’ increasing underachievement at school have suggested that it is because boys do not take to coursework but are able to redeem themselves through examinations. Modern school assessment practices demand lots of coursework and favour girls, who like it better. In the same vein, Lewis Elton (1998) has suggested that any grade inflation in higher education may be partly attributable to the greater importance coursework now has. He says that students tend to do better in coursework assessments than in examinations.

Interpreting marks
Section 4.3 sketches an approach to programme assessment that is intended to be responsive to this line of analysis by introducing the idea of claims to learning as complements to warrants of achievement.

4.3 Warrants, claims and process standards

Formal education, particularly higher education, is concerned with abstraction and generalisation. We do not want to know that a graduate can solve a particular problem but to know the likelihood that she or he can recognise and solve similar and not-so-similar problems. However, to repeat a point, when it comes to complex learning outcomes, such as identifying and working on novel problems, it is hard, expensive or even impossible to make high-reliability statements about the likely quality of future performances. Certainty is elusive when judging something as complex and plastic as competence. The simpler the outcomes, the more feasible it is to warrant or certify them. The more complex, the harder, more expensive and, perhaps, impossible it becomes to warrant the quality of future performances.

This would seem to leave many of the things that higher education is trying to develop – ‘mode 2’ knowledge, many ‘skills’, metacognition, and self-theories – unassessed, which is worrying for those who believe that students take seriously only that which is assessed. I don’t think these outcomes of learning have to be unassessed, though. There is no reason why local grade indicators cannot be developed and used, especially for low-stakes purposes, as when tasks are set in order to generate formative feedback. The mistake would be to assume that these local, provisional judgements provided a secure basis for warranting achievement. Instead, they provide an excellent basis for assessment conversations whose main aim is to stimulate further learning. These processes also put students in a good position to develop claims to achievement in respect of those complex learning achievements that HEIs cannot (or cannot afford to) warrant.

As I said in Chapter 1 (Section 1.6), it is increasingly recognised that formative assessment, which is low-stakes and uses local grade indicators, is an attractive choice when learning outcomes resist attempts to describe them reliably, at an affordable cost and without trivialisation. It provides feedback on how to improve performance and, when organised well, it should stimulate learners to reflect on their own achievements and identify directions for development: it should stimulate metacognition. Formative self-assessment has particular potential here, especially as a prelude to assessment conversations with others.

But how does this contribute to claimsmaking? In the UK, the Quality Assurance Agency endorses the idea that students should graduate with progress files, which will comprise an official transcript showing their certified achievements and personal development planning (PDP) documents, which will usually include a portfolio. The portfolio will contain claims to achievement and it would make a lot of sense for those claims to be concentrated upon learning outcomes that the university is not able to warrant. The claims would be backed up with evidence, which could come from activities related to the degree programme, out-of-class activities and work experiences. (For more information about PDP, visit http://www.ltsn.ac.uk/genericcentre/projects/pdp/intro.asp or http://www.
Formative assessment should make a good contribution to PDP because it is all about identifying achievements and thinking about what to improve next and how. In the process, portfolio claims are produced, tested and documented.

APL and APEL

This, then, is how learning outcomes that are not summatively assessed and warranted can be brought into a programme assessment system. Some achievements are assessed pretty reliably and certified by the HEI. Others are claimed by students and supported by evidence that has been largely generated and tested through formative assessment and which is supported by evidence of the programme’s process standards. I flesh out this differentiated approach to programme assessment planning in the next section, taking developments in Engineering programmes as a suitably complex example.
4.4 Programme-level assessment planning: an illustration

This section illustrates a general claim by describing one attempt to bring good thinking about assessment to bear on a typical, complex assessment problem: namely, how to set a standard for engineering degrees that are authentic expressions of engineering processes and then devise assessment procedures that touch all of the outcomes without making them into contorted skeletons. Engineers have been grappling with the problem of maintaining authenticity in the face of a need to assess authentic achievements.

In 2000, The Engineering Professors' Council published an interim report (EPC, 2000) of its work on the development of an output standard for undergraduate engineering degrees. The standard can be regarded as a set of threshold criteria, written as 26 'Ability to' statements – statements saying that an engineering graduate will have the 'ability to ...':

These statements came from a functional analysis of the process of engineering. Interestingly, the EPC decided to value complex learning outcomes, including ones to do with values, more than with information retention and command of standard problem-solution formulae. Because they are statements about achievements at the end of engineering programmes, they have to be stimulated, developed and assessed across whole programmes.

An assessment working group was then established to advise on the assessment of these 'Ability to' statements in engineering degree programmes. Its report (EPC, 2000) tried to offer the engineering community an approach to programme-level assessment that did justice to their ambitious threshold criteria without embroiling them in expensive attempts to measure things that cannot be measured – certainly not with validity at a cost that university departments can bear. I present Extracts from drafts for the Assessment Working Group of the Engineering Professors' Council's report below. They are intended to show how different assessment approaches have been integrated to offer a programme-wide approach to assessing complex learning outcomes.
The assessment of student learning: some distinctions

Most assessment in higher education is summative. It warrants or certifies students' achievements. There is a well-established distinction between assessment that has summative purposes, which means that it is a high-stakes, graded judgement of achievement, and that which has formative purposes. The aim of formative assessment is to help students to identify their own level of performance and how they might improve their future performances. With formative assessments the stakes are perceived to be lower; less is visibly at risk if there is error in the judgement.

Any learning achievement can be the subject of low-stakes, formative assessment, even complex ones relating to ill-defined or 'soft' skills. In such circumstances it would be hard to claim that the assessor's judgement would be as reliable as, say, a score on a set of multiple choice questions (MCQs), but that need not matter. The purpose is conversational, the anticipated outcome is learning and learning often involves dialogue. Seen like that, the assessor's judgement is a starting point in a learning conversation. It is not a final judgement and, although it should obviously be a fair judgement, it does not have to be reliable in the same way as summative assessments.

When the purposes of assessment are summative, reliability is at a premium. Some achievements can easily be reliably assessed. These assessments are called 'low inference' assessments and are typified by MCQ tests of information retention. Low-inference assessments may be reliable but they only work with determinate achievements where there is little ambiguity about the correct answer. The Engineering Professors' Council's (EPC's) output standards put considerable emphasis on achievements that are far more complex, where credit could be given for a range of solutions and for the means by which the solutions were developed. Although there is a temptation to use low-inference measures, such as MCQ tests of information retained, as proxies for such complex achievements, their reliability is bought by reducing complexity to simple proxy measures. In other words, there are sharp questions to be asked about their validity or worth. Where complex learning achievements are in question, there is a tension between the demands of reliable assessment and the requirements of valid assessment. In a paper prepared for the Assessment Working Group (AWG), Hamer and Macintosh (2000: 3) said that accuracy, which can be taken to be a facet of reliability, is not enough. They added that: 'Assessment practices that have as their main goal the chimera of precision will fail to meet both the needs of the individuals at whom they are directed and of the society of which they are a part.'

Can all 'Ability to' statements be assessed?

It will be apparent by the end of this report that all the 'Ability to' statements can be assessed in some way. However, that does not mean that all can be summatively (reliably) assessed, let alone within the resources available to most departments. Unfortunately, high validity and high reliability only go together when simple, determinate achievements are being assessed. In fact reliability itself is costly, can be difficult and is often to be bought by using artificial techniques that may be poor predictors of life-like performances. Complex processes are required to judge complex abilities and the more complex the abilities which the performance is supposed to show, the more samples are needed and the more complex is the assessment process. The process can be simplified but only by simplifying that which is to be assessed ... This may make for more reliable assessment but in the process the abilities in question have become simplified: routine problem-solving has been substituted for complex problem-working. If validity is to be preserved, reliability costs soar.

The EPC has set an output standard that authentically reflects engineers' work processes. Inevitably, some parts resist reliable assessment and others are only open to
tolerably reliable assessment if resources are invested in well-trained graders using good grade indicators to judge many pieces of work providing evidence of 'Ability to' achievements. Exactly which learning outcomes can be warranted depends partly on:

- The nature of the outcome (there is no great problem with the reliable and cheap assessment of information retention).
- How assessors decide to treat the outcome (any complex achievement can be simplified to make it easier to assess: it is a professional decision whether that loss of validity matters).
- How much cunning, time and money are invested in measures to increase reliability (authentic assessments of complex performances tend to be unreliable but acceptable reliability levels can be achieved at a price).

It follows that the EPC may wish to advise the engineering community about

- The outcomes that are best suited to summative assessment and best practice in summative assessment.
- The outcomes that are well-suited to formative assessment and best practice in formative assessment.
- How students might learn to make claims to achievement in relation to output standards that are not summatively assessed. Portfolios may be an answer.

... Criteria, threshold standards and grade indicators

... As the Interim Report (EPC, 2000) recognises, departments will need to develop grade indicators for levels above threshold level because students expect summative, high-stakes assessments to be reliable across the mark range, not just at the threshold level. Grade indicators help to reduce the area of disagreement amongst markers (although continued conversations about shared practices are needed to give the indicators life) and help students to understand better what those markers want and will reward (although they too will need help to understand the meanings behind the wording of the indicators). It is hard to see how assessment practices can lay any claim to reliability in the absence of such clear, understood and used indicators.

There is evidence, mentioned in Section 4 [of the Report], that engineering teachers find it difficult to develop grade indicators. (Teachers of other subjects also find this difficult.) Again, then, reliability may be better secured with assistance from the Assessment Working Group (AWG) or the Learning and Teaching Support Network’s Engineering Subject Centre (LTSN Eng).

It is also important here to note that the articulation of clear ability statements, criteria, threshold descriptors, and grade indicators alone does not lead to valid or reliable judgements. It is crucial for the academic community, and for students, to get together and discuss the meanings of these statements. The discussion needs to be benchmarked with examples of student work and discussions need to continue until there is an acceptable level of agreement on the judgements made.

These suggestions for improving the reliability of assessment practice are not a miracle cure for reliability problems. On the one hand attempts to produce better criteria or benchmark statements almost always lead to amplification and proliferation, so that simple benchmark statements, such as EPC has produced, accrete hosts of sub-criteria, clarifications and new statements designed to fill gaps that emerge. On the other hand, criteria, benchmarks and rules always have to be interpreted in contexts ...

The place of formative assessments of complex achievements

There is no requirement that all learning outcomes be warranted, that they be summatively assessed. The requirement is that they all be assessed in some suitable way because there is a belief that what is not assessed is not valued and a view that assessment can be a powerful aid to learning. On both grounds, then, it is necessary to have assessment arrangements for outcomes that are, in practice, beyond the practical reach of reliable, summative procedures. Formative, low-stakes assessment can be considerably cheaper, which means that resources for higher-reliability assessments can be released by not wasting effort trying to assess reliably
complex achievements that tend to resist reliable measurement.

What is envisaged is a formative assessment system in which:

- Many outcomes/abilities/achievements are formatively assessed. This assessment would be low-stakes, designed to give learners useful feedback on how to improve performance against programme-wide criteria. It would be embedded in the learning activities. Student participation in formative assessment would be a requirement for progress through the programme.

- Feedback should be fast, focused, relevant to the assessment criteria, developmental and personal to the student. Reliability would come second to plausibility of judgement, because if a learner felt that a judgement was wrong, then it would be important in the interests of learning for there to be open dialogue about that. This could help to reduce the incidence of the undesirable ‘final language’ of assessment and generally to reduce the negative emotions associated with the assessment of learning.

- Authentic assessments become easier to manage. The bugbear of authentic assessments has been getting reliability levels that are good enough for high-stakes purposes. Reliability is not such an issue when assessments are low-stakes and the main intention is to promote learning dialogues that inform future work.

- Each programme learning outcome should be complemented by grade indicators, including threshold descriptors, which would give teachers and students a better idea of what would be rewarded.

- Students should have the programme criteria from the first, regularly use them, share them, and practise applying them.

- Peer- and self-assessment should be embedded in programmes. Both save teachers time (which can then be used on high-stakes assessment) and help learners to become familiar with programme grade indicators. There have been heroic attempts to devise summative self- and peer-assessment systems but the position here is that they are best kept for formative purposes.

- Information and communications technology can support on-demand self-assessment that can provide feedback and even coaching on points of difficulty.

The value of this formative approach to assessment can best be shown by reference to pages 11 to 14 of the Interim Report (EPC, 2000). The Civil Engineering ‘Ability to’ statements say graduates should have experience in relation to ten statements and awareness in relation to six. Expressed in these terms, these are ‘Ability to’ statements that resist summative assessment. Students, though, should benefit from plenty of opportunities for formative feedback on work related to these 16 statements. Both teachers and students should benefit from using fuzzy learning criteria or indicators to organise their assessment conversations. Plainly departments could not warrant student achievement in respect of ‘Ability to’ statements that were mainly subject to formative assessment. However, these formative assessment arrangements, combined with a careers/employability support programme, should enable students to lay powerful claims to achievement which they could substantiate with material drawn from the learning portfolios they would keep. (This meshes with the Quality Assurance Agency’s recommendations on progress files.) Where reliable summative assessments allow departments to warrant achievement, valid formative assessment helps students to lay claim to achievement.

... Process standards

Whatever balance is struck between formative and summative assessment, or between coursework and examinations, there would be a need to describe a department’s process standards as well. Carter (2001) explains why. Talking of conventional closed-book examinations he remarks that

... even here the nature of the test is determined by how closely the questions set match those which the student may have seen before as exercises or worked examples ... That makes it almost impossible for external examiners to make comparisons between the standards in different institutions.

(p. 3)

Professor Carter is making a point about the degree to which the programme process
standards require students to solve novel problems without a high degree of scaffolding. Two similar levels of test performance might reflect entirely different process standards and therefore show two quite different achievements. The one associated with novel test items and little scaffolding would be the one that best fits the EPC output standard. The other would be a better fit with Year 1 work. Process standard statements would also be important because some abilities will not be summatively assessed and, consequently, departments need ways to assure graduate schools and employers that students are engaged by activities that are likely to lead to the un-warranted achievements to which students will lay claim. These standards could be verified by external examiners or other peer reviewers from other departments who should consider whether learning engagements across the complete programme are fit for the purpose of promoting progression. This should ensure that ingredients and processes that are likely to lead to good outcomes are in place and lead us to hope that, on balance, appropriate learning will follow.

... Social measurement theory (Campbell and Russo, 2001) has underpinned the argument that many complex abilities do not lend themselves to affordable and/or reliable assessment. They should nevertheless be subject to appraisal but formative assessment is more appropriate. Some achievements will be fairly reliably assessed and will be warranted or certified by the institution. Others cannot be warranted but good formative assessment arrangements will support student learning and their claimsmaking. Out of consideration for those who need to weigh student claims, departments are encouraged to provide statements about their process standards to complement the academic transcripts they will be providing to all new graduates. This is all summarised in the following figure.

Figure 4.1  An overview of ‘Ability to’ assessment arrangements (later published in EPC 2002)
Reflection 4.2 Responding to Figure 4.1

Are you persuaded by the assessment model summarised in Figure 4.1? Is it generalisable and useful?

If you think it has value, how would you try to convince colleagues of its worth?

If you see problems but have some sympathy with the overall approach, consider whether the model can be modified to ease those problems. Or is it the case that the problems are either ones that doom the model or ones we have to accept on the basis that they are not as bad as the limitations of current practices?

If you are hostile to the model, can you describe something better? That might take the form of a defence of present practice on the grounds that it meets the needs of your programme.

In each case the greatest benefit is to be had by sharing your ideas, which are likely to be concise, with a critical friend in your programme team or a fellow participant in any teaching and learning programme with which you are engaged.

4.5 Objections to this differentiated approach to programme assessment plans

Attracted though I am by this differentiated view of assessment, I need to explore three difficulties with my reliance on claims-making for the assessment of complex learning outcomes, which summative assessment practices find it hard to handle.

First, there is a need to programme time for students to reflect on achievements across courses (self-assessment), to test out their emerging claims by talking with each other (peer-assessment), and lay plans for improvement (often in conversation with teachers and other advisory staff). This cannot really be done in modules. Even when modules are, say, Lancaster-size (one module equals one-eighth of a degree), students really need to review their achievements across all modules, which helps them to think about modules where they are under-performing or where there are special opportunities for developing new claims to achievement. So, if this formative, claims-making approach to the assessment of many complex achievements is to work, programmes need to have scheduled reflect-and-plan sessions designed into them. There are obvious practical problems here for many established programmes, especially for ‘wide-choice’ ones, where students may feel little allegiance to the programme.

This hints at a second problem, which is student resistance. Although portfolios have attracted many educationists, students have often been much less keen. This is understandable because Personal Development Planning (PDP) has often been an add-on and an optional one at that. There are several reasons why students could be unclear about the use value of Personal Development Planning. Many will have kept records of achievement in schools, only to find employer and university indifference to them. They may also wonder how PDP will improve their degree class, especially if it focuses on achievements that the university does not certify, which means that they do not count towards the all-important degree classification. In short, it looks like a lot of effort for no bankable return. I think that there are good responses to these very legitimate fears but, unless programmes have been carefully designed to allay these concerns, then my
E-screening uses software to reject applications that do not use key words such as ‘flexible’, ‘self-reliant’ or ‘team player’.

One reason why employers like ‘A’ level scores is because they have decent levels of reliability built in, unlike university degree classifications, which do not.

Formative assessment, claimsmaking and portfolios approach will founder on student indifference.

The third problem concerns those who will hear these claims. Why should graduate schools or employers take these claims seriously? And how might they test the claims’ goodness, except by expensive interviews? Taking the second point first. It is likely that claims will not be tested unless a student gets to a short list. Initial screening is likely to continue to be based on degree class, e-screening of online applications and ‘A’ level and other entry scores. Once a short list has been constructed employers expect to spend time and money assessing who is the best candidate for the job. You could argue that portfolios make it easier and cheaper to select at this second stage, especially when compared to very expensive selection procedures such as assessment centres. (And they may not work as well as some imagine – Sternberg, 1997.)

This still does not explain why employers and graduate schools take seriously claims they cannot test. Of course, they can probe claims at interview but I want to suggest that if HEIs were more forthcoming about their programmes, then stakeholders could judge student claims better and identify the programmes most likely to produce the sort of graduates they wanted. Employers and others need to know whether student claims are upon learning processes that are likely to have stimulated the claimed learning. They need to know about the process standards that have operated throughout the programme – for example, whether in their degree programmes learners worked in different ways in different settings so that they are likely to be as flexible, good at problem identification and working, and skilled at working in teams as they claim to be. If student claims to achievement are to be taken seriously, HEIs will have to publicise their programme process standards, that is the parts of their programme specifications that show how each learning outcome is supported by the learning, teaching and assessment arrangements. I wonder whether the module mark and degree classification transcripts that HEIs are going to issue might be accompanied by a transcript of the programme process standards.

These standards should be assessed for goodness of fit with the learning outcomes, possibly by external examiners who are likely to become appraisers of programme quality. For example, external examiners might judge that a claim to skill at group work should involve experience of: working in small and larger groups; working in self-chosen and no-choice groups; sustained as well as short-term group work; group leadership; taking on a range of group roles; reflecting on and responding to feedback on one’s own performances in groups. They would have reservations about student claims not based on such process standards, as would other educated consumers. Similarly, I would mistrust student claims to be used to working autonomously unless the process standards showed that students had to identify for themselves topics to study, design methods of inquiry, carry it out and report on their conclusions. And I would want to know what experiences students have of processing and interpreting numerical data before taking seriously any Humanities student’s claim to numeracy.

4.6 Communication and networks of shared practices

Underlying this is the idea that good assessment practices are based on transparency, in the sense of making it clear what is happening and what the intended learning is. This is essential as we come to see assessment as something far more complicated than the precise measurement of achievement. Once it is understood that some achievements will not be warranted by HEIs, that it is not clear what grades mean, and that the achievements that HEIs certify with the greatest confidence will tend to be those that matter least to employers and
graduate schools, then common-sense notions of assessment need radical revision. And precisely because this differentiated understanding of assessment runs counter to common sense, it is important to explain what we do when we assess particular achievements and why.

Briefing students about the task and criteria
Good assessment is about good communication as an act of courtesy to graduate schools and employers. It matters far more to another group, namely students, as is made clear above in *Briefing students about the task and criteria*. I am returning to my ‘knowing students’ theme when I say that they are disadvantaged as players, achievers and claimsmakers if they do not know what the rules of the game are, what the favoured academic practices are and why these rules and practices are valued. I illustrate this with Figure 4.2. The terminology is likely to be unfamiliar but I have used it because it connects with what is known as ‘actor-network theory’, which deals with the formation of interest groups and networks. The underlying idea is that it takes time to join networks, activity systems or communities properly. Figure 4.2 suggests that we can understand it as having four ‘moments’ or overlapping stages. Students coming into a learning community will generally take time to learn and internalise its practices, which, if it is a good community, are likely to differ from those that bought success at lower, simpler levels. Figure 4.2 indicates some of the things that designers might think about when they work out ways of helping students to become full members of this activity system, community or network. They – the designers – need to have that concern in mind when they are constructing a programme learning environment that stimulates an assessment culture that is conducive to complex outcomes. Without struggling to explain or understand the terminology, I think the point is plain: we need to be considerate if we want students to play our assessment games well. The new practices need to be clearly shown to students and their logic needs to be explained. Notice too that Figure 4.2 indicates that teachers need to do more than explain the rules of the assessment game. They need to make sure that module and programme design provide plenty of opportunities or affordances to support intéressement and mobilisation. Mentkowski *et al.*, (2000) reports of practice at Alverno College are interesting pointers to what this involves.

Good communication also matters to teachers, whether they are module teachers or programme leaders, because, if they are not clear about why students are assessed in these differentiated ways, then there is little chance of them explaining the reasoning to students or acting it out in their own work. For example, if teachers have been used to seeing assessment as a high-stakes affair in which the priority is generating reliable marks, then they are likely to subvert, perhaps unwittingly, an assessment culture intended to stimulate formative conversations. Just as students need to be enrolled into the network and understand the assessment activity system and its sub-systems, so too with teachers.

So much of the material in this pack has emphasised that learners can easily default to unhelpful approaches and operate on the basis of misconceptions if they are not:

- clear about the rules of the game (or games in differentiated systems, where the formative assessment game can be described by Figure 1.6 but summative games look very different – see Figure 1.5);

- aware of the reasoning behind those assessment practices – if they don’t understand why they are good practices;
sure about what assessment values, rules and grade indicators mean in practice;

able to trust that teachers’ actions will square with their words; that if they say that multi-structural answers are going to be rewarded, they don't then go and give the best marks to information-dense descriptions.

This makes communication a major assessment priority: forming shared understandings with students and colleagues of the activity systems that describe formative (Figure 1.6) and summative (Figure 1.5) assessment activities; appreciating the ways in which these understandings tend to be developed and mobilised (Figure 4.2); and communicating as best we can with postgraduate schools and employers who want to know how to tell what students have achieved and are likely to bring to the workplace or postgraduate programme.

<table>
<thead>
<tr>
<th>Moment (or stage)</th>
<th>Explanation</th>
<th>Illustration</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Problematisation</strong></td>
<td>Learners realise that a programme's assessment goals – its rules of the game – are different from those with which they are familiar, which means adopting the new goals if they are to play well.</td>
<td>First-year students find that essays are not now judged on how much information they contain. Third-year students begin to see that high-scoring writing deals with all sides of an argument. Students notice that formative assessment, which they have treated as a 'skive', is highly valued in their degree programme. In each case, there is a dissonance between what they are used to and their emerging understandings of the programme's assessment goals.</td>
</tr>
<tr>
<td><strong>Intérressement</strong></td>
<td>Learners begin to engage with the problem – mastering new assessment practices. They are beginning to break with old practices: to 'unfreeze'.</td>
<td>They begin to think about what it would mean to put well established practices to one side and follow the new assessment practices: writing essays that centre on arguments; writing pieces that look at all sides of an argument; creating and responding to feedback.</td>
</tr>
<tr>
<td><strong>Enrolment</strong></td>
<td>Learners 'sign up' to mastering the problematic assessment practices and make use of the affordances – tools, rules and community of practice – that can help them.</td>
<td>They try out new methods, making use of any tools available (course handbooks, advice on good essay writing, skill-building sessions), referring to assessment rules (grade indicators), and drawing on the community of practice (tutors, other students). Ideally, they become skilled in the academic practices they need in order to play this assessment game well.</td>
</tr>
<tr>
<td><strong>Mobilisation</strong></td>
<td>New assessment practices are internalised by the learners on the course or programme.</td>
<td>As individuals, learners buy into the new practices and become skilled in them. As a group – as a learning community – they buy into the new way as the right way. This can be clearly seen in the case of students at Alverno College, Milwaukee (Mentkowski et al., 2001) who often spend their first semester disorientated by Alverno's distinctive 'assessment-as-learning' practices. Most go on to be strong advocates of them.</td>
</tr>
</tbody>
</table>

Figure 4.2  Students learning to succeed in a new assessment game
**Reflection 4.3 Communication**

You may already know of the Quality Assurance Agency’s plans for Progress Files to improve the quality of communication about student achievement. If not, you could visit [http://www.qaa.ac.uk/currwork/progspec/progspec0600.pdf](http://www.qaa.ac.uk/currwork/progspec/progspec0600.pdf) to get details of a development which, if it is not derailed, will affect those of us teaching in higher education in the UK.

I invite you to consider the extent to which these proposals are (1) feasible; (2) fit for the purpose of improving communication with employers and postgraduate schools about student achievement.

If you are not happy with their fitness for purpose, you might consider how we could better communicate student achievement, always bearing in mind that all innovations have costs — your suggestions need to be feasible.

---

**Summary and endnote**

**Reflection 4.4 Key points revisited**

I began this pack by naming seven points I wanted to make.

1. To challenge the assumption that assessment is measurement (although measurement is one approach to assessment which is fit for some purposes)

2. To challenge the idea that better assessment depends on using a wider range of better techniques. Many assessment problems are not technical problems but are reflections of the way things are. Better techniques will not allow us to assess critical thinking any the better because the concept itself is intrinsically fuzzy and non-determinate.

3. To argue that higher education is, internationally, expected to promote complex learning outcomes. That implies complex thinking about assessment.

4. To insist that it is important to think well about local, module-level assessments of learning and to think about programme assessment systems as well.

5. To stimulate differentiated thinking about assessment, that is to encourage you to plan to assess different sorts of learning in quite different ways.

6. Overall, to encourage the conviction that well-conceived assessment stimulates good learning.

7. To conclude that assessment is about communication that helps us to choose what to do, appreciate why it is valuable and think about how to do better.

I invite you to reflect on this pack by:

- Assuring yourself that you follow the arguments associated with each of those seven points

- Reviewing your reactions to each point. Is it sensible or not? Practicable or not? Relevant to what you are doing or not?

- Making sure that any other points that have impressed you – positively or negatively – are recorded in some prominent way.
If you have the stamina, you might imagine that you have to write a one-side report for colleagues in your module team about the implications of this professional material for your programme practices. What would you say?

And if you do that task, you might follow through by approaching someone, perhaps in your university’s educational development unit, to see what support could be provided for beginning to work in your module on one or two of those implications.

Endnote

Professor Sir Paul Black, one of the UK’s most eminent assessment experts, recently wrote:

... researchers are faced with the difficult task of changing understanding of assessment issues, both amongst the general public and amongst policy makers.

(Black, 2000, p. 407)

His University of London colleague, Professor Harvey Goldstein, has world-class expertise in measurement theory, especially as applied to school effectiveness research. Reviewing a substantial literature (Goldstein and Woodhouse, 2000), he argued that public policy is misshapen because of an indifference to theoretical issues, notably what effectiveness is, whether it can be measured and, if so, by what means.

So, too, I have argued, with higher education.

Better theory is necessary if we are to have better assessment practice. That has been an underlying purpose pervading this pack and my seven declared priorities.
References


Garrick, J. and Clegg, S. (2000) ‘Organizational gothic: translating vitality and transforming the corporate body through work-based learning’ in


Murphy, R. J. L. (1979) ‘Removing the marks from examination scripts before remarking them. Does it make any difference?’, *British Journal of Educational Psychology*, 49, pp. 73–8.


Quality Assurance Agency (2000) *Subject benchmarks*


## Acknowledgements

### Attributions

Materials have been included from previous versions of H850. The sources of these are:

**H852 Teaching in Higher Education: theory and evidence**
Chapter 2 Specifying Aims and Learning Objectives  Ranald Macdonald
Chapter 3 Designing Assessment Graham Gibbs and Derek Rowntree
Chapter 4 Marking and Giving Feedback Graham Gibbs
Chapter 5 Teaching and Assessing with C&IT Agnes Kukulska-Hulme

**H852 Practice Guides**

Practice Guide 2 Specifying Aims and Learning Outcomes David Baume
Practice Guide 3 Designing Assessment Graham Gibbs and Derek Rowntree
Practice Guide 5 Teaching and Assessing with C&T Agnes Kukulska-Hulme

**H851 Practice Guide 4 Making and Giving Feedback  David Baume**

Details of these sources appear below:

<table>
<thead>
<tr>
<th>Section</th>
<th>Identification in this text</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Assessment and your students</td>
<td>Chap. 4, pp. 26-7</td>
</tr>
<tr>
<td>1.2</td>
<td>Clinical legal skills module box</td>
<td>PG3, pp. 43-4</td>
</tr>
<tr>
<td>1.2</td>
<td>Thirty-two assessment methods</td>
<td>PG3, pp. 34-6</td>
</tr>
<tr>
<td>1.2</td>
<td>Exams</td>
<td>PG3, pp. 37-40</td>
</tr>
<tr>
<td>1.3</td>
<td>Using assessment to improve student learning</td>
<td>Chap. 3, pp. 14-21</td>
</tr>
<tr>
<td>1.5</td>
<td>Validity; Reliability; Improving Reliability (1) and (2)</td>
<td>Chap. 3, pp. 6, 9-11, 15-17, 24-7</td>
</tr>
<tr>
<td>1.6</td>
<td>Using journals and logs</td>
<td>Chap. 3, pp. 32-4</td>
</tr>
<tr>
<td>1.6</td>
<td>Case study of David</td>
<td>PG4, p.19</td>
</tr>
<tr>
<td>1.6</td>
<td>Allocating marks to assessment components</td>
<td>PG3, pp. 19-20</td>
</tr>
<tr>
<td>1.6</td>
<td>Formative self-assessment by students; Using CMC for peer and teacher feedback</td>
<td>PG5, pp.22-4</td>
</tr>
<tr>
<td>2.1</td>
<td>Creating grade indicators</td>
<td>PG 4, pp. 11-17</td>
</tr>
<tr>
<td>2.1</td>
<td>Setting criteria and standards</td>
<td>PG 3, pp. 20-2</td>
</tr>
<tr>
<td>2.2</td>
<td>How can I prepare students for assessment?</td>
<td>PG4, pp.29-31</td>
</tr>
<tr>
<td>2.2</td>
<td>What effect does assessment have on your students?</td>
<td>Chap. 4, pp. 26-33</td>
</tr>
<tr>
<td>2.2</td>
<td>On the edge of plagiarism</td>
<td>PG4, pp. 25, 27, 29</td>
</tr>
<tr>
<td>2.2</td>
<td>Some advice on doing the marking</td>
<td>PG4, p. 17</td>
</tr>
<tr>
<td>2.2</td>
<td>Giving feedback</td>
<td>PG4, pp.,20-6</td>
</tr>
</tbody>
</table>
2.2 Internalising feedback Chap. 4, p. 18
2.2 Assessing groups PG3, pp. 45-7
3.1 Submission and marking of assignments PG5, pp. 24-5
3.1 Learning about and from your assessment PG4, pp. 32-5
3.1 Maximising fairness PG3, pp. 29-30
3.2 The language of learning outcomes PG2, p. 23
3.2 Deriving learning outcomes from aims PG2, pp. 16-18
3.2 Implications for assessment strategies and methods PG2, pp. 31-3
4.2 Sampling the curriculum PG3, pp. 17, 18
4.2 Interpreting marks Chap.3, p. 6
4.3 APL and APEL PG3, p. 42
4.6 Briefing students about the task and criteria PG3, pp. 18-19

Section 2.1 Extracts from a social science department's grade indicators P.Knight Unpublished at time of assembling this file.

Section 2.2 The specification for a summatively-assessed assignment P.Knight Unpublished at time of assembling this file.

Section 3.3 A programme assessment plan and some academic and social practices P.Knight Unpublished at time of assembling this file.

Section 4.2 Problems with high-stakes assessments of complex achievements P.Knight Based on the paper later published as Knight, 2002b