H850 Postgraduate Certificate in Teaching and Learning in Higher Education

Pack 3

Assessment for learning: practices and programmes

Prepared for the Course Team by Peter Knight, Centre for Outcomes-Based Education, with Jo Tait
### Contents

**Introduction to Pack 3**

Jo Tait

**Contexts and purposes**

Peter Knight

<table>
<thead>
<tr>
<th>Chapter 1 Assessing students taking your module: design issues</th>
<th>13</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1 Assessment and the hidden curriculum</td>
<td>13</td>
</tr>
<tr>
<td>1.2 A palette of assessment tasks</td>
<td>15</td>
</tr>
<tr>
<td>1.3 Being clear about academic practices: knowing the rules of</td>
<td>24</td>
</tr>
<tr>
<td>the assessment game</td>
<td></td>
</tr>
<tr>
<td>Extract 1.1 Essay writing and the quality of feedback (Hounsell)</td>
<td>25</td>
</tr>
<tr>
<td>Extract 1.2 Students' expectations of courses in higher education (Ramsden)</td>
<td>29</td>
</tr>
<tr>
<td>Extract 1.3 Using assessment strategically to change the way students learn (Gibbs)</td>
<td>32</td>
</tr>
<tr>
<td>1.4 Making links</td>
<td>39</td>
</tr>
<tr>
<td>1.5 Complex learning, validity and reliability</td>
<td>41</td>
</tr>
<tr>
<td>Extract 1.4 Validity (Brown, Bull and Pendlebury)</td>
<td>44</td>
</tr>
<tr>
<td>Extract 1.5 Reliability (Brown, Bull and Pendlebury)</td>
<td>47</td>
</tr>
<tr>
<td>Extract 1.6 Improving reliability (Newstead)</td>
<td>50</td>
</tr>
<tr>
<td>1.6 Formative and summative assessment purposes</td>
<td>56</td>
</tr>
<tr>
<td>Extract 1.7 Course context (Gibbs and Griffiths)</td>
<td>59</td>
</tr>
<tr>
<td>1.7 Planning module-level assessment arrangements — problems and issues</td>
<td>65</td>
</tr>
<tr>
<td>Extract 1.8 Planning assessment for a new course (Rowntree)</td>
<td>69</td>
</tr>
<tr>
<td>Extract 1.9 Differences in conceptions of assessment tasks (Hounsell)</td>
<td>74</td>
</tr>
<tr>
<td>1.8 Summary</td>
<td>80</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chapter 2 Assessing students taking your module: practice issues</th>
<th>81</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1 Grading work in your module</td>
<td>82</td>
</tr>
<tr>
<td>2.2 Good module-level assessment — necessary features</td>
<td>93</td>
</tr>
<tr>
<td>Extract 2.1 Cue consciousness and hinting (Miller and Parlett)</td>
<td>98</td>
</tr>
<tr>
<td>Extract 2.2 The hidden curriculum (Snyder)</td>
<td>101</td>
</tr>
<tr>
<td>Extract 2.3 Are assessment practices immune from fraud? (Newstead)</td>
<td>106</td>
</tr>
<tr>
<td>Extract 2.4 The psychology of student assessment (Newstead)</td>
<td>110</td>
</tr>
<tr>
<td>Extract 2.5 Written feedback on student writing (Leo)</td>
<td>120</td>
</tr>
<tr>
<td>2.3 Good module-level assessment — desirable features</td>
<td>125</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chapter 3 From module to programme</th>
<th>133</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1 Good module-level assessment practices</td>
<td>133</td>
</tr>
<tr>
<td>3.2 Learning outcomes: first time around</td>
<td>134</td>
</tr>
<tr>
<td>Extract 3.1 Aims and objectives (Ramsden)</td>
<td>136</td>
</tr>
<tr>
<td>Extract 3.2 Learning outcomes (Allan)</td>
<td>138</td>
</tr>
</tbody>
</table>
Chapter 4 Assessment and degree programmes

Prelude

4.1 The structure of this chapter

4.2 Learning outcomes again

Extract 4.1 Combining marks (Brown, Bull and Pendlebury)

4.3 Warrants, claims and process standards

4.4 Programme-level assessment planning. An illustration

Extract 4.2 Extracts from drafts for the Assessment Working Group of the Engineering Professors' Council's Report

4.5 Objections to this differentiated approach to programme assessment plans

4.6 Communication and networks of shared practices

References

Acknowledgements
Introduction to Pack 3

Jo Tait

This introductory overview of the pack, Assessment for Learning, aims to provide a context for your reading, suggests some ways to engage with the ideas and reflective activities contained within, and describes the way it has been put together.

This single pack forms one part of the extensive materials designed, primarily, for use by participants in the Postgraduate Certificate in Teaching and Learning in Higher Education programme. You can find out more about the programme at http://iet.open.ac.uk/coursesonline/. If you have received this pack as part of that programme, you will have received a full set of these materials: you will also have access to web-based materials and interactive support from peers and tutors. If you are reading this pack as part of your study for that programme, you have probably been following an ordered sequence of study. In this case, you will certainly recognise that Pack 3 is written in a different style and formatted differently from all the others. This introduction explains that difference and provides an overview of the whole pack that outlines the structure and provides some sense of continuity to your study.

The other packs in the set will focus on:

- Professional capability
- Students learning
- Design for learning
- Teaching methods and approaches.

There are inevitable overlaps between the different perspectives addressed by each pack. We have designed every pack as a stand alone resource to support and inform your understanding of students’ learning in higher education (and in other adult learning contexts), whether or not you are working towards a formal qualification.

Our readers

There is a broad range of readers who may find this pack useful or rewarding. You might be:

- a lecturer or tutor with a particular disciplinary background
- relatively experienced or fairly new to your role
- a supervisor in charge of students in work-based practice
- a librarian or IT resources manager with responsibility for supporting students who need to manage information
- a guidance professional with a generic interest in student learning, or
- you may have another role that we have not yet imagined.

Working with this pack

The papers in this pack have been structured by a single academic author with a clear purpose in mind. This is one of the main differences between Pack 3 and all the other packs in the series. You will find, however, that we still provide a range of diverse perspectives in the form of extracts from previously published materials: these sometimes support and sometimes
challenge Peter Knight's authorial voice. You are encouraged to join that
dialogue and assert your own position alongside (and in opposition to)
any of the voices and arguments you encounter. To help you locate
yourself in relation to these materials, as you read, I suggest that you
might ask yourself questions such as:

- If the purposes of higher education have changed so much in the past
ten years, how is my disciplinary or academic community or my
professional body working with and responding to those changes?
- Where do policy decisions at national and institutional level affect my
teaching and how far should (or can) I change my assessment and
design practices to align my work with such policies?
- How do the arguments about assessing the outcomes of learning with
validity and reliability resonate with my experience of how students
learn and how I teach? In what ways do I recognise these issues from
my practice in marking and giving feedback to students?

Decide, now, what you hope to gain from the pack at this reading and
make a note of your own particular questions about assessment. As you
read through the pack, you will find reflective activities and questions that
urge you to stop and think about what you have just read – or are about
to read. These prompts urge you to develop your own positions on
assessment issues, in the light of your practice, your context and your
experience. With such an authoritative voice carrying you forward quite
persuasively, it is important that you check whether these assertions,
however well argued and evidenced, really match up with your own
understanding. Note your comments and responses; make them part of the
pack, in some way, so that you hold your own in this complex discussion.

You may be looking through this pack for quite strategic purposes. If you
are studying towards an assessed course yourself, there will be structures
embedded in the assessment process of your programme of study that will
influence what you look for and how you work with what you find. As
you will read within these pages, the context for study and the assessed
outcomes often carry the learning process. We hope that your own
intentions and expectations will drive your reading, but that the
assessment strategies for the study you are undertaking can help you
prioritise your reading so that you make the most effective use of your
time and attention.

This pack has been constructed in a linear or progressive way, with an
argument that is followed through each chapter. For this reason, we
recommend that you take the time to read it from beginning to end, even if
this has to be a fairly light reading. To make sense of the argument, you
will certainly need to read Peter Knight's own introduction to the pack
because here he develops a view of higher education in 2002 by showing
how its espoused aims and purposes have changed in the past forty
years. His view of policy and some anticipated developments in higher
education form the basis for the following chapters and frame his
purposes in writing.

Chapter 1 considers the issues associated with designing assessment for
individual modules, questioning some theories of measurement and
drawing on the literature about how students learn. Chapter 2 is even
more practice-oriented and, because of the way that the author encourages
you to reflect on and challenge some of the ideas, you may find yourself
reading backwards and forwards to develop your own synthesis of the
varied suggestions for practice. Chapter 3 begins to introduce the challenge of thinking about assessment across whole programmes of study, and suggests that learning outcomes (or outcomes of learning) need to be assessed in as many ways as possible across a programme – that assessment of ‘soft’ and complex outcomes cannot be accomplished within the design of one module. Chapter 4 revisits some of the problems with learning outcomes and the assessment of competence in higher education, and concludes by outlining some broad approaches to assessment that could enable effective programme-level assessment systems.

The continuing construction of the materials

This pack has been constructed differently from the other four packs. Its structure is organised by the critical voice of Peter Knight, urging the reader to reflect on the changing environment for teaching and learning in 2002. The pack does, however, draw on many extracts from chapters that were written in 1998 and 1999 by members of the original course team working together.

We have tried to maintain a clear distinction between the different voices (old and new?) in the way we have presented the texts.

Peter's narrative voice will always appear in Sans Serif text, like this.

The original text, and extracts from journals and research papers, will appear in the standard format for this pack and for the other four packs.

In making our amendments to the structure of the programme and its assessment, we have listened and responded to evaluations and feedback from students, tutors and examiners. And, in the wider environment, we recognise that changes to higher education and the accreditation of teaching in higher education have influenced the needs and expectations of participants. All these factors have affected the pack you now hold in your hands. Its current format allows it to continue to improve in response to feedback and changing needs, so we hope the course will become even more participative, with your engagement and contributions. Please send your comments to iet-pgctlhe@open.ac.uk

Contexts and purposes

Peter Knight

Assessment is a major concern for all who teach, if only because marking takes so much of their time. As class sizes grow and teachers are expected to show that all course intentions are associated with some assessment activity, assessment becomes a sharp cause for concern.

It is also a concern for those who want higher education (HE) to help learners make convincing claims to a range of achievements that are frequently identified with national well-being. With higher education now being seen as ‘part of the productive sector of the knowledge economy’ (as Sir Howard Newby, Chief

---

1 Most people on the Open University programme are based in the UK. This is reflected in the materials, although you will see that I do refer to work done outside the UK. Although there are some differences among higher education arrangements in Wales, Northern Ireland, Scotland and England, they do not systematically affect the ways in which we think about assessment, so I will largely ignore them in this pack. Distinctive regulatory features in the UK mean that some of the things I write about, such as programme specifications and degree classifications, make little sense to people working elsewhere. However, the ideas and principles do. There is also an international concern to widen participation, encourage a broader range of skills and think about learning, teaching, assessment and curriculum on a programme-wide basis.
Executive of HEFCE – the English Funding Council – recently asserted), there is a concern that graduates should be able to contribute to national wealth creation. That implies that higher education is concerned with more than the development of subject-matter understanding. Research keeps finding that employers want graduates with complex achievements such as self-management skills; coping with uncertainty; transfer skills; self-confidence; team-working; and managing others. This policy priority raises some very awkward assessment issues because it is by no means easy to see how self-management, coping with uncertainty, self-confidence and team-working can be assessed.

Another government policy, its widening participation policy, complicates matters because there is little point in getting 50% of 18 year olds into higher education if they face assessment practices that promptly fail lots of them.

Now, it might be said that these policy priorities are no concern of teachers in higher education. Their job is to get on with the day-to-day business of testing understanding in manageable ways. Once, perhaps, that was true – in the days of phases 1 to 3, as Figure 1.1 puts it. Much less so now when, as Figure 1.2 suggests, the policy priority of widening participation demands systemic thinking by higher education institutions (HEIs) and those stimulating change within them.

A programme is a sequence of modules leading to an award, such as a BSc.

A programme is made up of modules. Modules are sometimes called ‘units’.

It is best to avoid the word ‘course’ because it gets used as a synonym for ‘module’ and as a synonym for ‘programme’.

**Figure 1.1** Four phases of thinking about LTAC improvement in higher education
Policy imperatives apart, in the mid-1990s the UK Quality Assurance Agency was beginning to emphasise programmes more than modules. The mid-to-late 1990s can be seen as a time of transition from the view that better learning, teaching, assessment and curriculum (LTAC) depended on widening the range of techniques in daily use to a systemic concern with programme-level arrangements, particularly with coherence, continuity and progression. In this view, better LTAC comes from better programme design. It is being increasingly appreciated that, once we move from asking about student understanding of particular pieces of content and ask instead about the development of critical thinking, reflective practice, good judgement and self-management, we need to be thinking about learning across a whole programme. The focus moves from the module to the programme. The implication is that we need to think about how modules are aligned, to use Biggs’ term (1999), to stimulate coherent learning throughout the undergraduate years. This clearly implies that teachers should be clear about the ways in which their modules contribute to programme learning goals. And, given the proposition that learning, teaching, assessment and curriculum are four elements of the same system, it follows that assessment arrangements need to be planned with the programme in mind. The message is that if programme learning outcomes are to be taken seriously, they ought to be related to a programme assessment plan and each module in a programme should contribute to that plan in some clear way.

This new thinking, differing from the taken-for-granted of the assessment phases 1 to 3, invokes some theoretical and research perspectives that are not to be found in most books on assessment. (You can see this by looking at my earlier
work: Brown and Knight, 1994.) For example, I ask how certain we can be when judging human achievements and what sort of arrangements allow us to be more confident about such judgements. These questions might scarcely trouble the teacher of a free-standing module but it is a different matter when judgements made in a module contribute to an award attesting that a student is, for example, good at group work, independent or open-minded. Such questions are particularly acute for those leading vocational and applied programmes – at undergraduate and master’s levels, within higher education and outside it – who need to develop evidence-informed systems for the assessment of professional competence and professional learning. My argument is that a systemic perspective is necessary and that it is beneficial because it pushes us to become clear about what we judge, how, and why. In the process we come to see that we can become quite confident about the assessment of some learning outcomes, while other outcomes just defy firm and objective judgement. So, a systemic approach to assessment also implies a differentiated approach – it means recognising that different sorts of learning outcome imply different assessment approaches.

That said, the systemic approach is not prominent in Chapters 1 and 2 because they concentrate on our work as module teachers. However, I argue in Chapter 3 that mainstream teachers need to be skilled in programme assessment techniques and appreciative of the assessment issues that should underpin programme design, and I review some of the issues this raises in Chapter 4. I am implying that good teachers think about programme-level issues as part of thinking about the design of their own modules and that in doing so they contribute to good learning, teaching and assessment systems.

By now it will have become clear that I do not see teachers as mere deliverers of other people’s plans. For me, teachers are professionals who make and have to make decisions about how best to stimulate good student learning. I am writing, therefore, about teachers who have a say in how the students they teach are assessed. However, there are part-time teachers who are contracted to follow a script written by others. New teachers and those working as members of module teams may also find that they have no choice but to implement practices decided by others. It is a short step from there to the conclusion that what I have to say may be fine and dandy but detached from your reality as a new teacher, part-timer, or team member. I have a lot of sympathy with that view and with the resentment that accompanies it but I maintain that the position I’m going to develop is relevant because it is something to which to aspire. Agreed, some teachers may be little more than wage slaves but I think that is all the more reason for insisting that good learning, teaching and assessment are favoured by working conditions that stimulate teacher professionalism and the creativity, innovativeness, care and flexibility that it implies. What follows is written in that belief.

My key points

There are seven things I particularly want to say in this pack:

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. Two reasons are:
   
   - Many complex learning outcomes develop over years, not days. This implies programme-wide planning to develop and assess them.
   
   - If we wish to make reliable statements about learning outcomes that are not simple outcomes, then we need to base them on plenty of assessments, not just on judgements made in one or two modules.

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. Which, as Linn concluded, is not easy:

   As someone who has spent his entire career doing research writing and thinking about educational testing and assessment issues, I would like to conclude by summarising a compelling case showing that the major uses of tests for student and school accountability over the past 50 years have improved education and student learning in dramatic ways. Unfortunately, that is not my conclusion.

   (Linn, 2000, p.14)

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.

My approach

I see myself writing a commentary on assessment issues that are bearing upon what we do as teachers in higher education or which are going to affect us a lot more in the next decade. Where I am reviewing established module-level practices, I can draw on a great deal of good published material. As a result, Chapters 1 and 2 are mainly carried by extracts from the work of authorities in this field. You can find attributions to these authors in the Acknowledgements at the end of the pack. In these chapters, I see myself trying to cement these pieces together to provide an authoritative but polyvocal account of the assessment issues that are closest to teachers' practices.

Programme-level issues are not as well covered in the literature so I have more to say in Chapters 3 and 4. Clearly, I am developing my view that we need to pay a lot more attention to programme issues and questions about assessment in relation to complex learning and 'soft' skills. I am arguing that, even if our immediate need is to find ways of reducing the marking load, there is also another sort of need, which is to try to create coherence between our modules and our programmes. You might not be persuaded, but if you work in the UK, you will certainly be repeatedly confronted by this view. If you work overseas, you can also expect to hear more about it.

You might feel that there is an imbalance in the examples I use, that there should be more on, say, the assessment of practical competence, perhaps in the workplace. I agree, adding that I am writing to a word limit and simply do not have the space to do justice to some very complex problems in programme design: the assessment of performances and, by extension, of competence is one
of them. None the less, I encourage you to consider the notion of competence and its assessment (Reflection 4.1). I suggest that the issues that need to be considered in more straightforward assessment situations (the assessment of academic thinking through essays) are all the more formidable when it comes to assessing competence. That said, the examples I have used are a fair reflection of the mainstream assessment literature. Going beyond them would have needed more space than is available to me.

Notice something that I hope will come through later in this pack, that I am framing a question which is often pitched as a problem for an individual teacher as a programme design question. No one can assess any complex competence reliably in a single module. You may be able to afford to get tolerably reliable local judgements of aspects of competence in a module but, if you are trying to judge something complex, you are not going to do it reliably in a single module. In other words, one of the things I am trying to do is to encourage colleagues to consider the idea that many things that they treat as assessment problems that they, as individual teachers, ought to be able to solve are really system design problems for programme teams and educational consultants. From time to time, a Reflection box will invite you to reflect on some of the ramifications of what I am saying.

Lastly, I have been quite sparing in my use of references. I have provided them when scholarship really demands it or when they will help you to find out more if you want. If this gives the impression that assessing well is a matter of common sense, then I apologise. Assessing well depends on good views about what is (ontology), how we might know about it (epistemology), as well as upon a great deal of psychological evidence about learning, transfer, measurement and development. Good assessment may involve having a bag of good assessment techniques, but good decisions about which ones to put in the bag, how to combine them, use them and draw conclusions relate to good thinking about psychology and philosophy.
Chapter 1 Assessing students taking your module: design issues

I am going to begin by establishing that assessment has a powerful hold over student learning, hinting that good assessment can make for good learning. (The reverse is just as true.) In Section 1.2, I say that a simple way of improving assessment practices is to widen the range of tasks we use so that we can reach a wider range of achievements in a more interesting way. I follow this, in Section 1.3, with a point that is sometimes overlooked, namely that students need to understand what we are trying to do, otherwise there is a danger that they will assimilate our novel and varied assessment practices into their existing notions of what assessment is for and what counts as real assessment. The problem is that these notions are often skewed to the idea that assessment rewards evidence of diligence and the amount of information you can display, whereas higher education is concerned with deeper and more complex learning. Unless students are 'knowing' students, who understand our 'rules of the game' there is a danger that good assessment practices will be not have the intended effects. This prepares the way for the idea, in Section 1.4, that assessment needs to be integrated with module and programme learning goals, which means that it needs to be differentiated: different methods are fit for different learning purposes. The next two sections introduce two dimensions on which we might consider assessment. In Section 1.5, I argue, somewhat controversially, that our common wish to have 'objective' and reliable measurements of achievement is often at loggerheads with our wish to stimulate complex outcomes of learning. Valid assessments of complex learning tend not to be reliable. Reliable ones tend not to be very valid. One way of handling this is to concentrate on reliability (at the cost of simplification) when the stakes are high and the purposes are summative, and to respect validity when the stakes are lower and the purposes formative. Section 1.6 examines the purposes of assessment, along a continuum between summative and formative. Section 1.7 invites you to consider and critique the implications of applying differentiated approaches to the design of assessment for learning in individual modules.

1.1 Assessment and the hidden curriculum

The following summary of Snyder's work in America, with the connections it draws with the UK context, makes it clear that the ways in which we assess do contribute to student learning, although it may not always be the learning we would wish to happen. This theme is revisited again in Section 2.2 and you can read an extract from Snyder's work at that point.

Assessment and your students
More seriously, if the aim of a module or programme is to promote complex learning, then assessment arrangements must be taken very seriously because assessment has a tendency to simplification.

I need to introduce an artificial distinction here in order to explain this point. If our aim is to add to students' knowledge or to provide opportunities to practise and refine existing achievements — if our learning outcomes are essentially additive — then assessment need only nudge students to do things which are within their zone of comfort. If, on the other hand, we want to stimulate qualitatively different learning — to stimulate fresh understandings or to introduce new ways of doing things — then we need assessment to push students to change their existing understandings and practices. To put it another way, when complex learning is our goal, we need assessment arrangements that will draw students away from their comfort zone and into the more disconcerting zone of proximal development (ZPD for short). Unless assessment procedures reward attempts to move into this developmental zone, then our goal of stimulating complex learning is likely to founder on students' tendency to assimilate what we are trying to do into their existing tried-and-tested practices and understandings. Rather than accommodating those understandings to the complex goals we have in mind, learners tend to assimilate our complex ambitions to their existing, simpler understandings and practices.

When complex learning is the goal (I explore this in more depth in Section 1.5), then good teaching can be defined as everything we do to try and support it. It follows that designing good assessment arrangements is an important part of good teaching. To design good assessment sequences, whether at module or programme level, teachers need at least an understanding of: how learning happens; undergraduate programme goals; the design of learning environments; and teaching practices associated with complex learning.

Teachers also need to understand the range of assessment methods available and to appreciate that any method suits some purposes more easily than others. This points clearly to the need to develop a differentiated view of assessment — an understanding that assessment is a family of practices, not a homogenous process of objective measurement. I develop this differentiated view in Section 1.2, where I list 32 ways of assessing learning.

### 1.2 A palette of assessment tasks

One of the great achievements of phase 3 thinking about assessment (see Figure 1.1) was to make teachers more aware of the range of tasks that could be set to stimulate student learning, provide feedback for further learning, and provide some indication of achievement. Given the view that the tasks we set and the activities students do are central to their learning (more so than our direct teaching), the choice of assessment/learning tasks is a very important part of module design.

Ohlsson (1995), reviewing the literature on how we learn (a) procedural, 'know how' or skills knowledge and (b) declarative, propositional 'know that' knowledge, suggested that effective learning involved joining in a set of 'epistemic tasks' or games:

- Describing
- Explaining (events or states)
- Predicting
- Arguing
- Critiquing (evaluating)
- Explicating (concepts)
- Defining.
This leads him to the tentative definition of, '... the study of higher-order learning as the study of how people learn to perform epistemic tasks' (p. 52).

Be that as it may, an implication is that good assessment practice will involve selecting tasks that engage students with a good range of epistemic games. This is illustrated by the example of a clinical legal skills module described in the text box below. The list of 32 assessment methods that follows it shows some of the possibilities. I have published a similar list of 50 (Knight, 2002a) and Hounsell et al. have produced a helpful compendium of tasks used in Scottish universities (Hounsell et al., 1996). See the LTSN Generic Centre website (http://www.ltsn.ac.uk/genericcentre/projects/assessment) for more. The discussion of the various ways in which examinations can be used below reminds us that there are often many variations on any assessment method, not just variations between assessment methods.

Thirty-two assessment methods

The list below contains 32 different types of assignment, in alphabetical order, with an example of each. We do not have space to explore all the possible forms of assignment, let alone explain how they can best be used.
Exams

Copyright material removed

OpenLearn
Copyright material removed
OpenLearn
Further examples of forms of question can be found in Habeshaw et al. (1993).
One of the most important characteristics of these questions is the effect they can have on the way students study and revise. For example, if students have no idea what they might be asked or how, their revision might be poorly focused and concentrate on memorising anything which might come up. If they know they will have a design task the sensible thing to do is practise designing. The ‘case of the philosopher of education’ in the series of case studies that inform Extract 1.3 provides a good example of exploiting this principle in the design of exam questions.

Open-ended questions clearly involve a degree of subjectivity in marking, particularly those forms where what is required is not explicit, where the level of outcomes being tested is unclear, and where markers may have their own hidden agendas about what the ‘right’ answer is.

Purposes of assessment

I have implied that one purpose of assessment is to structure, support and stimulate student learning and will finish this pack with the claim that assessment is learning. (Which leads to the claim that since assessment is learning, it is in everybody’s interests to make sure that assessment is the best learning we can imagine.) I want to open up here a theme that I am going to make a lot of: if assessment is valuable in proportion to the quality of the learning it stimulates, then we should be concerned, first and foremost, about the validity of our assessment arrangements – about how stimulating, supporting and structuring they turn out to be. We can worry a lot less about how reliable – objective or accurate – the assessment procedures are because our primary intention is to evoke learning, not to measure it. Of course, we want to make better judgements rather than worse ones, but ‘better’ is, on this view, judged more in terms of usefulness than in terms of accuracy of assessment (whatever ‘accuracy’ might mean).

If that were all there is to assessment, life would be easier and this pack would be a lot shorter. However, there is a strong demand for measures of achievement so that professional competence can be warranted or certified, so that employers and graduate schools can select the best, and to provide performance indicators to (mis)use in the management of teachers, departments and higher education institutions (HEIs). These are high-stakes judgements, which means that we want them to be as accurate, objective and as reliable as possible. I shall shortly argue that reliability and validity disrupt each other, with the result that the learning purpose of assessment disturbs the measurement purpose. A good part of this pack is about trying to cope with this central tension between assessment functions and purposes.

Reflection 1.1 Designing for diversity

The main reasons for using a range of assessment methods in a module are not to provide variety and enhance interest, important though they are, but to engage with a range of learning outcomes.

I suggest starting by considering the learning outcomes of a module you teach. If they look very ‘samey’ and are all to do with content mastery, you might go to the Quality Assurance Agency’s subject benchmarks website (http://www.qaa.ac.uk/crntwork/benchmark/benchmarking.htm) and notice the variety of learning outcomes suggested for the subject area in which you work.
Then review the list of assessment methods (refer to the set of 32 given earlier in Section 1.2 or use another of your choice), identifying those that seem best suited to the assessment of different learning outcomes.

This leads to the conclusion that differentiated learning outcomes imply differentiated assessment methods.

You are likely to identify far more assessment possibilities than you have room for in your module, which could suggest:

1 you need to be very selective about the choice of assessment methods
2 you might reduce the number of learning outcomes associated with your module so that you have a better chance to assess each one well
3 we need to think about programme-wide assessment plans if a wide range of learning outcomes is to be well assessed. (This is the subject of Chapters 3 and 4.)

1.3 Being clear about academic practices: knowing the rules of the assessment game

If assessment can have different functions, if there are many different assessment tasks available and if they can be used for different epistemic games, it follows that it is very easy for learners to approach an assessment task with completely the wrong idea in mind: to treat it as an invitation to display information where the teacher intended it to be an opportunity to take risks with critical thinking. I illustrate this in with the following extract Extract 1.1, entitled Essay writing and the quality of feedback, showing how students can understand the same sorts of task in quite different ways. The extract is a section from a longer piece and will reappear at the end of this chapter as part of Extract 1.9 Differences in conceptions of assessment tasks.
Two groups of second-year university students took part in the empirical investigation: seventeen whose main subject was history and sixteen whose main subject was psychology ...

The main findings stem from analysis of the interview accounts. Within each of the two subject groups, a fundamental difference was identified between the students' conceptions of what an essay was and what essay-writing involved. The full set of categories of description derived (Hounsell, 1984) cannot be presented here for reasons of space, but Table 1 provides a summary of the main differences in conception. Some history students conceived of essay-writing as a question of argument, coherently presented and well-substantiated; others saw it as concerned with the arrangement of facts and ideas. And amongst the psychology students, essay writing was seen by some as a matter of cogency, where substantive discussion was rooted in a solid and coherent core of empirical findings, and by others as relevance, in the sense of an ordered presentation of material pertaining to a topic or problem.

Table 1 summarizes the differences in students' conceptions of what was required of them in history and psychology essays respectively.

In the first group are the two interpretative conceptions of argument (history) and cogency (psychology). What these two conceptions share is a concern with the making of meaning: an essay is seen as a mode of discourse through which one makes sense of a topic or problem in a way which is individually distinctive ... The student is thus the arbiter of the content and form of the essay, making personal choices about the selection and organization of material, whether in the light of an explicit interpretative stance which had crystallized at an early stage or of a steadily growing sense of the dominant theme. Introductions and conclusions are not simply conventions but devices through which a unity of form and content can be facilitated.

The second group of conceptions comprises the two non-interpretative ones of arrangement (history) and relevance (psychology). In each, an orientation towards the establishment of meaning is absent, and each is correspondingly less explicit than its interpretative counterpart. Although there is an acknowledgement that essays are a medium through which one's own ideas, thoughts and opinions can be conveyed, these tend to have an almost incidental status or are seen as 'value added' rather than as the essay's main justification ... In consequence, students' descriptions of their essay-writing procedures appear flat and mechanical, a series of steps which seem uninformed or undirected by considerations of meaning. And the students' accounts of the content of their essays typically do not include a substantive reference to a conclusion.

### Table 1 Conceptions of essay writing

<table>
<thead>
<tr>
<th>History</th>
<th>Psychology</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ARRANGEMENT</strong></td>
<td><strong>ARGUMENT</strong></td>
</tr>
<tr>
<td>Students define an essay as the ordered presentation embracing facts and ideas</td>
<td>Students define an essay as an ordered presentation of an argument well supported by evidence</td>
</tr>
<tr>
<td><strong>RELEVANCE</strong></td>
<td><strong>COGENCY</strong></td>
</tr>
<tr>
<td>Students define an essay as an ordered discussion of relevant material on a topic or problem</td>
<td>Students define an essay as a well-integrated and firmly grounded discussion of a topic or problem</td>
</tr>
</tbody>
</table>
No less importantly, such differences in conception seem to have a more general association with students' grasp of a discipline. Students ascribed to an interpretative conception tended to see the craft of essay-writing as mirroring the practice of the discipline (in history) or (in psychology) as an activity both fostering and demonstrating one's mastery of the accumulated knowledge of the discipline. The non-interpretative conceptions, by contrast, do not have these associations. There are instead abundant indications of vagueness and uncertainty about the nature of essays and essay writing and a perceived gap between aspiration and what is achieved in the finished essay.

In the past twenty years there has been a strong *phenomenographic* research tradition in higher education. These researchers explore the differences in understandings of experience, in this case the differences in students' ideas about higher education: the differences between student understandings of the same thing (such as the nature of learning); differences between understandings of different things (differences between understandings of learning and of studying); and differences in the same student's understanding of similar things in different times and settings (the limits to their willingness to transfer understandings from one time and context to another). One of the most memorable ideas that these phenomenographic researchers have developed is that students differ in the approaches they tend to take to learning tasks. Some incline to 'surface' approaches, which resemble the Arrangement and Relevance categories in Extract 1.1 above. 'Surface' approaches are to do with acquiring and organising material but they are essentially confined to arranging it, to adding to the learner's store of information.

'Deep' approaches commit the learner to trying to construct a personal understanding of the material; to interrogating it, making connections with existing understandings and forming new understandings. They are close to the Argument and Cogency categories.

There are students who for various reasons take 'apathetic' approaches. The term rather speaks for itself, although students may be 'apathetic' because they have 'learned helplessness': failures have taught them that no matter how hard they work, failure will result. Quite reasonably they conclude that it is better to preserve some self-esteem by not trying (and failing) than by trying (and failing).

The fourth set of approaches has been labelled 'strategic'. This means decoding all available cues to try and work out what the teacher wants. The teacher may declare that she wants understanding but the students, interpreting other cues, may judge that what she really wants is evidence of a lot of reading and they use 'surface' approaches to give it to her. Alternatively, they may see that the task of describing the general crisis of the mid-seventeenth century is intended to invite analysis and critical thinking and use 'deep' approaches to produce an appraisal of the thesis that there was such a crisis in the first place.

These ideas are explored in more detail elsewhere (notably in Pack 2 of these materials, *Students Learning*). I have said enough to claim that good assessment practices are considerate of students in the sense that they help students to appreciate the rules of the game – to understand what is being valued and why, as Extract 1.2 below, *Students' expectations of courses in higher education*, shows.

**Assessment at the heart of learning: assessment as learning**

I have suggested that the ways in which we assess exert a powerful influence over students. Unfortunately, we are not always clear enough about what we want them to do. This is a serious oversight because students are then thrown back on their intuition and have to figure out what they think we want them to give us. (It also means we are making trouble for ourselves as markers, but that is another story.) Students have already developed their preferred approaches to learning and some of the common ones – notably 'strategic' and 'self-protection' approaches – do not sit well with what we say HE is about. If we do not help students to understand what our assessment tasks are for, what we shall reward and how they might go about constructing it for us, then we face the glum prospect of failing students and miserable marking.

This means that we, as teachers, need to be clear about what any assessment task is for. It certainly helps students to know that a task is going to be treated as 'low stakes', will mainly get peer feedback (comments from other students)
and is there to stimulate better learning in the future. It may help us more if formative, low-stakes tasks like this do not need to be graded reliably and by the teacher. They are for discussion, not for certification. The more we are clear about what an assessment task is for, the more we can appropriately emphasise the need for accurate, objective and reliable grading or the value of stimulating worthwhile conversations about authentic, life-like tasks.
EXTRACT 1.2
STUDENTS' EXPECTATIONS OF COURSES IN HIGHER EDUCATION

Paul Ramsden

To begin a course of higher education is for many students to begin a period of uncertainty and confusion. Whether the transition is from school or work, or even from a previous year of study in the same institution, students often have only the slightest idea of what to expect. ... We easily forget how students can experience a sense of disorientation owing to a rapid shift from an ordered and familiar environment to one of considerable freedom. Although it is apparent from several studies of transition from school to university and the early experiences of higher education that the first few weeks of a student's experiences in higher education are critical to success in the remainder of their programmes, the effects sometimes do not show themselves until the second year of a programme of study or even later (see, for example, Entwistle et al., 1989; Entwistle, 1990). It is clear that students often spend a lot of their time simply trying to discover what we want them to learn, and that many of them fail to perceive the links between the academic knowledge they have acquired in school and the work they are required to do at university (Marton, Ramsden and Bowden, 1989) ...

The first question in teaching anything should be 'What do I want my students to learn?' It should be closely followed by a second question: 'How can I express my requirements to my students?' It cannot be emphasised too strongly that satisfactory answers to these questions must precede attempts to address problems such as how to present a lecture, manage a tutorial, or use visual aids. To appreciate the significance attached by students to the clear expression of the pattern, content, and expectations of a course, consider again some of the items from The Course Experience Questionnaire study ... that differentiated the 'good' courses from the 'bad' ones:

- It's often hard to discover what's expected of you in this course.
- The aims and objectives of this course are not made very clear.
- You usually have a clear idea of where you're going and what's expected of you in this course.
- It's always easy here to know the standard of work expected of you.
- The staff here make it clear right from the start what they expect from students.

In the good courses, students tended to disagree with the first two statements listed and to agree with the rest ... It is indisputable that, from the students' perspective, clear standards and goals are a vitally important element of an effective educational experience. Lack of clarity on these points is almost always associated with negative evaluations, learning difficulties, and poor performance ...

Using assessment to improve student learning

[Here we are] concerned with the design of assessment systems, not to measure learning as accurately as possible, but to bring about as much learning as possible. It presents assessment as the most powerful lever teachers have for changing the way students learn. Students’ experience of courses is often dominated by the way they are assessed, and students have always been very strategic in orienting their study behaviour to meet assessment requirements in ways which use the minimum effort.

- [There is more in Chapter 2 on students’ response to assessment systems. A glimpse of that work (Snyder, 1971; Miller and Parlett, 1974: see Extracts 2.2 and 2.1) is given by the student in Snyder’s study who said]: ‘I just don’t bother doing the homework now. I approach the courses so I can get an A in the easiest manner, and it’s amazing how little work you have to do if you really don’t like the course’ (Snyder, 1971, p. 50). And this was a postgraduate student in one of the world’s top research universities!

- Once the ratio of in-class to out-of-class learning hours reaches about 1:3, assessment exerts more influence on the way students go about their learning than what happens in class. Our perception of the centrality of our teaching is not always shared by students, who may see turning up to class as a strategic option to meet assessment requirements economically.

- Diary studies have shown that students may spend as little as 5% of their time on work which is not assessed, and that this narrow orientation to assessed tasks becomes more marked as they progress through their courses (Innis, 1996), quite contrary to the intention to encourage more independent, self-directed learning over time.

- The financial and employment context students find themselves in means that many full-time students allocate their time strategically so that they can earn cash from part-time employment. The proportion of part-time students with competing work or family commitments has greatly increased. Students are under much greater time pressures than used to be the case.

- Social and cultural changes, in the UK at least, have made students more concerned to achieve good qualifications than to rebel (MacFarlane, 1992).

For all these reasons students may pay considerable attention to features of the assessment system. They may allocate more time to some study tasks rather than to others, pay more attention to some aspects of course content than others, and orient themselves to the criteria they think will be used to allocate marks. They may be ‘selectively negligent’ about those parts of your course which they believe will not be assessed or will not contribute significantly to gaining good marks. They may do this whether or not you deliberately use assessment to orient their studying. The side-effects of assessment, though often described as ‘unintended’, are always substantial. It is sensible to try to control these side-effects and turn them to your advantage rather than leave them to chance.

Some teachers may protest at this portrayal of students as extrinsically oriented and only concerned about marks. It is certainly the case that mature students, for example, may be more interested in learning than this account suggests. It may be important for you to explore the extent to which this picture is true for your particular context.
The following piece by Graham Gibbs is drawn from his 1999 paper on using assessment strategically to change the way students learn. The extract contains a series of case studies of changes in assessment that had dramatic effects both on the way students went about their study and on their marks.
EXTRACT 1.3
USING ASSESSMENT STRATEGICALLY TO CHANGE THE WAY STUDENTS LEARN

Graham Gibbs
As is implied by Extract 1.2 *Students' expectations of courses in higher education*, the more we are clear about assessment, the more we are being considerate of students and their learning, ourselves and our practices, and of what teaching and learning in higher education are about. This is a view of LTAC (learning, teaching, assessment and curriculum) as a system operating on students who are in environments that have effects on their learning, as do their individual learning biographies. It suggests that it is going to be hard to change students' learning practices and that chances of success will be slim if the four elements of learning, teaching, assessment and curriculum are not constructively aligned with one another (Biggs, 1999).

**Reflection 1.2 Designing for clarity**

If students are to have a fair chance of succeeding with our assessment tasks and of achieving the complex learning outcomes that we intend, they need to understand:

- what the rules of the game are
- what the thinking behind these assessment practices is (for example, why there are so many formative, peer-assessed tasks)
- what teachers are looking for
- what the task specifications are
- what grading indicators will be used
- what, in practice, will get rewarded.

You are invited to reflect on the 'patter' about assessment that could go in your module handbook to explain your assessment thinking and practices to students. To put it another way: what can you do to make your students 'knowing students', that is, people who know how to succeed in your assessment game?

[Ideally, there would be consistent programme assessment practices, and module-level explanations would only need to identify the occasional distinctive practice, otherwise reminding the student of programme assessment rationale and practices.]

1.4 Making links

Systems thinking has begun to affect the ways in which we think about module design. For example, attempts to define the 'graduateness' of graduates (HEQC, 1997) stimulated the production of benchmark statements for 24 subject areas (see [http://www.qaa.ac.uk/crntwork/benchmark/benchmarking.htm](http://www.qaa.ac.uk/crntwork/benchmark/benchmarking.htm)). Although these benchmarks describe the threshold achievements of a UK graduate, individual modules contribute to these achievements and should be planned with reference to programme specifications (see [http://www.qaa.ac.uk/crntwork/prospec/prospec0600.pdf](http://www.qaa.ac.uk/crntwork/prospec/prospec0600.pdf)).

In the 1990s, as Figure 1.1 suggested, there was nothing exceptional about teachers having good ideas for a module and going ahead and validating the product of their enthusiasm. Now, though, this enthusiasm should be mapped on to the programme specification so that we can see how the enthusiast's module would make a coherent contribution to the programme learning outcomes. That said, individual modules may have one or two learning outcomes that are not in the programme specification. However, because modules have to engage with
some of the programme learning outcomes, there is only room for a couple of module-specific ones.

Figure 1.3 sketches a relationship between learning outcomes at programme and module level and relates them to the design decisions that should be made about the contribution modules make to the programme's learning intentions.

It introduces an important concept, 'scaffolding', which refers to the support that learners need on their route to autonomy or expertise. It may affect task design (tasks for novices spell out what needs to be done; those for more autonomous learners expect them to identify the problem and possible solutions); learning support (making helpful resources readily available for novices while expecting more experienced learners to identify for themselves the resources they need); and social support (by letting inexperienced students work together, often under close tutor supervision, while expecting more experienced students to work independently when necessary). We often find that students who can show evidence of similar achievements have experienced different levels of scaffolding. Reports of achievement should not, therefore, be taken at face value, a point I develop in Chapter 4.

---

**Figure 1.3 Designing assessment arrangements for a module**

Ideally, thinking about the assessment arrangements for a module begins when a set of learning outcomes is selected from the programme specification. If, for example, a learning outcome is chosen that is to do with working effectively in groups, then it follows that we have committed ourselves to assessment methods that can illuminate it. With module learning outcomes selected, the next question would be a direct assessment question: what arrangements will be made for
assessing these outcomes? In this ideal world the assessment question comes before decisions about teaching, learning and sequencing because:

- There is a view that what is assessed gets learned. Conversely, what assessment overlooks, students also overlook. This has been elaborated in *Using assessment to improve student learning* (Section 1.3, above) which showed how assessment plans can capitalise on this in the interests of promoting good, complex learning.
- Assessment tasks are also learning tasks, especially where they are designed as formative tasks. There is a real sense in which assessment is learning.

It follows that assessment decisions, whether at module or programme level, are not trivial. The design of good assessment practices makes for good student learning. This runs counter to an old view that assessment is the enemy of good learning. Phase 3 assessment thinking (see Figure 1.1) was particularly concerned to dispel this view, mainly by describing a wide range of assessment methods that were consistent with good learning intentions. This thinking would have had more impact had it been supported by a good summary of measurement theory in order to disrupt the assumption that all human achievements can be measured (a) reliably, (b) validly, (c) affordably. Because phase 3 writers on assessment did not challenge, head-on, tacit beliefs that assessment is about objectivity, certainty and measurement, attention was diverted into ways of making new assessment techniques, notably peer-assessment, reliable (as you will read in Section 1.5). My claim is more radical. Good assessment makes for good learning because assessment of the complex achievements we want our graduates to have is not primarily a matter of measurement. Assessment is better understood as the creation of plenty of affordances for feedback that can help metacognitive processes.

In this chapter, I am exploring some of the implications of this thinking for the design of module-level assessment plans. In Chapters 3 and 4, I shall make strong links between module design and programme design, arguing that many of the complex learning outcomes that we want students to achieve need programme-level attention – a coherent suite of learning and assessment tasks that stimulates progression in their learning across the undergraduate years. Besides beginning to make that point, here, I want to suggest the corollary – that many assessment problems that get treated as technical ones to be solved by classroom teachers are really design problems for programme teams to tackle. Take, for example, the assessment of practical competence in subjects such as nursing, social work, radiology, teaching and project management. To be reasonably confident about a general judgement of competence we need a series of judgements by several observers, each of whom has been trained to apply a well understood and well designed set of criteria or indicators in a variety of occupational settings. This is not possible in a single module. (But a module can contribute judgements to programme-level judgements of competence.)

1.5 Complex learning, validity and reliability

This section begins to develop an idea to which I shall repeatedly return, namely that we need to have differentiated assessment systems. The argument is that many of the aims of higher education are complex, by which I mean that they cannot be precisely and neatly specified. In the same measure, these complex aims defy measurement. If we want to make valid judgements about achievement in respect of these complex goals, then we cannot do so with certainty. If, on the other hand, we want certainty, then we have to rely on simple measures that are insensitive to the complexities we are supposed to be judging. In technical terms, reliability and validity are in tension with each other. Although we can get
tolerably reliable judgements without gross simplification, the more we wish to have reliability and validity, the more expensive it is.

This is laying the ground for the argument that we should identify those outcomes of learning that are fairly simple and can be assessed reliably, those that are too complex for reliable judgement, and those which we can afford to judge with as much reliability and validity as possible.

The importance of complex outcomes of learning

My view of assessment rests on my claims that higher education is primarily concerned with complex learning and that complex learning resists valid, reliable and affordable assessment. For example, the view is gaining ground that the first cycle of HE – the undergraduate years – should be laying the foundations for lifelong learning, the reasoning being that professionals will be continuously updating themselves as today’s knowledge becomes obsolete and today’s jobs disappear. How is a propensity for lifelong learning to be assessed, let alone preparedness to be a successful lifelong learner?

I take it there is no great problem with the claim that higher education is concerned to promote complex, or advanced, understandings of subject matter. Nor that there would be much objection to the view that it should develop subject-specific and ‘transferable’ skills. Few would argue that critical thinking, analysis, synthesis and evaluation are oven-ready achievements. Lists of the achievements that employers, in particular, look for in new graduates suggest that they also want higher education to encourage graduates’ efficacy beliefs, so that they have self-theories consistent with persistence and strategic thinking. As Dweck’s (1999) book on self-theories insists, this is far from simple. Finally, there is consensus that graduates should be reflective or, to use a word I have already introduced, to have good metacognitive capacities. Again, this is a complex achievement. And all the more so with the whole USEM (understanding, skills, efficacy beliefs, metacognition) complex.

This fits well with the view that higher education institutions are increasingly expected to produce knowledge that is more focused on application in ‘the real world’ than on reflection within disciplinary structures in universities. This, sometimes called ‘mode 2 knowledge’, is ‘output-driven rather than discipline-driven ... its validation lies in its use-value or performativity ... [and] it answers to pragmatic rather than to disciplinary “truth”’ (Usher, 2000, p. 100). Or, to put it another way:

An adequate preparation of people for the market now requires greater subtlety, flexibility, responsiveness, and pragmatism ... skilled workers and trained employees may have to ... adopt new ways of doing things and ‘knowing’ about their work ... It is against this backdrop that capacity-building approaches begin to look desirable ... Unlike earlier approaches to competency-based training, capacity building focuses on the ways that workers need to think and how workers need to be in the world ... workers have to reconceptualise not only their tasks and roles but also themselves.

(Garrick and Clegg, 2000, p.164)

Such thinking confronts universities with complexity because it all calls on them to add to what they have been doing. For instance, helping students to make strong claims to being highly employable people implies some preparation for participating in problem-solving, consultative committees and quality circles; for formal and informal on-the-job training; for flexible team-working; and for
understanding the sorts of identities that are valued in workplaces and to appreciate how to take them on (Chappell et al., 2000). Notice two points. One is that there is ineluctable pressure on higher education to take these expectations seriously. Secondly, the claim that these concerns can be compatible with HE’s traditional concern to promote good understandings of worthwhile content.

Measurement, validity and reliability

The obvious assessment implication is that complex learning outcomes need a varied set of assessment methods, which was the key insight of phase 3 thinking about assessment (see Figure 1.1). It is less obvious that many of these outcomes resist measurement. Measurement depends on the thing to be measured being real – being determinate, in the sense that it can be fully, hence accurately, described. It takes little time to realise that the complex achievements of the USEM (understanding, skills, efficacy beliefs, metacognition) account of the goals of higher education are not determinate like a table, the trajectory of a speeding bullet, or knowledge of conventional historical facts. They are not determinate achievements. There may not even be a lot of consensus about what they are. Consider ‘critical thinking’ for example. There are North American tests of critical thinking which can be objectively administered and reliably scored, but they are more like IQ tests than estimates of the complex processes of informed criticism in Fine Art, History, Sociology or Philosophy. Even within those disciplines, there is argument about what counts as criticism (as opposed to ‘rent-an-opinion’) and there is disagreement between disciplines.

Validity

Validity is a measure of the extent to which assessment measures what it is supposed to measure. The following extract explores what we mean by validity and distinguishes five main types. The claims made in this extract about the predictive ability of degree results for career performance are not borne out within most professions: higher education is not usually at all successful in predicting which students will do well afterwards.

A glance at some of the research on employability shows that higher education is expected to foster: willingness to learn; self-management skills; communication skills; effective learning skills; exploring and creating opportunities; action planning; networking; coping with uncertainty; transfer skills; self-confidence; team-working; managing others; critical analysis; being able to work under pressure; imagination/creativity. These are sometimes called ‘soft skills’. I reckon that the word ‘skills’ is misleading but I still find the phrase ‘soft skills’ a useful one.
EXTRACT 1.4
VALIDITY

G. Brown, J. Bull and M. Pendlebury

Copyright material removed
OpenLearn
Copyright material removed

OpenLearn

We have a double validity problem in this extract. First, is that we cannot agree on the construct validity – we cannot agree what we mean by critical thinking. Second, that it is far from clear that the tests that get used are getting at critical thinking rather than at IQ. The brief paragraph entitled Validity, which precedes Extract 1.4, gives you a basic account of what this key assessment concept can mean. My definition is that valid assessments get at the achievements you want to engage with – not at a sawn-off version, or at a mediating factor, like the students’ reading skills or memory of mathematical formulae. Now, if my programme or module goals are to develop those achievements, then I want assessment to support that development. I do not want to be sponsoring a sawn-off version of, say, problem-working because I can only devise reliable measures of convergent problem-solving, which is a very different sort of achievement. Since I begin from a position of scepticism about the usefulness of our summative assessment practices, I am not going to put reliability, which only matters most in high stakes, typically summative, assessment, ahead of validity.

This is a quick sketch of reliability. There is an extended treatment in Extract 1.5, followed by suggestions for improving reliability in Extract 1.6.

Reliability is crucial whenever the stakes are high – in examinations, for instance. We want to know that the examination procedures are reliable and that our score is not affected by glitches in the paper or in the marking process. We want the whole process to be objective. Reliable estimates are usually based on many judgements or measurements. Single measurements are not trustworthy, as Fleming (1999) shows with essay marking.

Reliability is often bought at the expense of validity. At base, validity is about being sure that we’re getting at the very thing we say we’re trying to judge. However, our assessment tests often compromise validity because reliability thrives on low-inference judgements. But we can only make low-inference judgements of simple achievements; complex achievements call for what Eisner (1985) called ‘connoisseurship’, which is high-inference and unreliable in the sense that two connoisseurs may legitimately disagree. This is why so many tests of critical thinking look more like IQ tests than judgements of the complex, contextual business of appropriately thinking critically.

When we try to improve reliability we simplify; we reduce reading tests to tests of word recognition and ignore the definition of reading as a form of thinking. Critical thinking becomes IQ tests. In the process, validity is shed and we encourage students to settle for simple forms of what are intended to be complex learning outcomes.

We can try to preserve validity but it is expensive and tends to compromise reliability – valid judgements of complex learning need connoisseurs and experts, instead of optical mark reading machines processing thousands of multiple-choice questions an hour.

Reliability

Marks awarded to students are sometimes treated with more respect than they deserve by examinations committees, teachers, and even students. A mark of 59 per cent may be considered to be not of ‘upper second class’ standard because it is not over 60 per cent, but this trust in the accuracy of marks is not well founded. The following extract examines what we mean by reliability and considers evidence about how reliable assessment in higher education might be. It refers to work by Newstead on the reliability of assessment in psychology that appears in Extract 1.6.
Comments by Graham Gibbs and Derek Rowntree

Brown and colleagues distinguished several different kinds of reliability:

- Consistency between markers. Do markers agree with each other and use the same criteria? If any assignments or exams you mark are 'double marked' you will begin to find out about this kind of consistency! In controlled studies of inter-marker reliability, marking has sometimes been found to be effectively random. Student marks often depend on who marks their work.

- Consistency of individual markers. Would you always give the same piece of work the same mark, regardless of when and how you marked it? Do you treat each piece of work in the same way, bringing the same criteria and standards to bear in the same way regardless of the content of the work or who wrote it? If you marked the same assignments six months later would you award exactly the same marks?

- Consistency within tests. For example, do students get much higher marks on some essay or exam questions, or on some design briefs, than on others? If so, students' marks may depend more on choice of question than on what they have learned. Are some fieldwork or workplace assignments actually providing quite different challenges, being much harder than others? If so, students' marks may depend on the luck of the draw in the allocation of assignments.

- Consistency of students. Would students get the same marks if they took the same test a week later or took a different – perhaps bolder and more imaginative – approach to an assignment? Extract 1.5 points out that if students are only tested once the risk of them having an 'off day' is much greater, and that increasing the number of occasions on which students are tested will increase this kind of reliability.

Improving reliability (1)
EXTRACT 1.6
IMPROVING RELIABILITY

S. E. Newstead

Copyright material removed

OpenLearn
Balancing reliability with side-effects

The problems that Newstead identifies are age-old and to some extent intractable. All assessment in all formal educational settings suffers from these kinds of problems. Some solutions themselves cause problems – as Newstead acknowledges with regard to the side-effects of using multiple-choice questions in order to increase reliability. Another example of such a side-effect would be the increased use of exams in order to cut out some of the cheating associated with coursework. This may at the same time reduce the learning activity associated with coursework and induce last-minute cramming which leads to short-term learning. Exams also bear little resemblance to any task students might undertake afterwards and as a result exam marks are a very poor predictor of student success in subsequent employment or even in postgraduate work. Here the by-products of the proposed solution to the cheating problem are probably more damaging than the original problem.

Some of the problems have relatively simple solutions. For example, the level of reliability of marking can be improved by the use of structured marking sheets containing criteria. Such sheets are used increasingly where consistency between markers is an issue and are also used to orient students to what is required and to give targeted feedback (McDonald and Sansom, 1979).

It is not clear that all of the phenomena described as ‘problems’ are in fact undesirable. For example, for students to work collaboratively can be seen as a productive learning activity rather than as a problem of cheating. Newstead says that the rates of cheating are ‘disturbingly high’ but what rate is tolerable is a matter of judgement. Even what cheating means is not as clear as the data presented suggest. Newstead himself admits to confusion about why reproducing lecture notes is accepted while reproducing books is considered cheating. Students are even less clear what plagiarism means or what is ethically acceptable and what is not (Ashworth et al., 1997) but they distinguish between what is only designed to get better marks and what helps their learning. There are also cultural differences: some overseas students have a different understanding of what counts as plagiarism and can inadvertently fall foul of rules.

Overall a balance needs to be struck: damaging problems needs to be addressed, but in ways that are not unacceptably constraining or disruptive. As teachers we should try to make the assessment system of which we are part work as well as it can under the circumstances.

Improving reliability (2)

Reliability is concerned with whether markers would give the same mark to the same piece of work on another occasion, whether different markers would agree with each other, or whether parallel questions or tests on the same topic would produce the same results. Inter-marker reliability can be embarrassingly low – even random – for assignments such as essays or dissertations, and students may perform very differently on alternative exam papers for the same course.

... Reliability can always be improved, for example, 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 be 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 opportunity your students have to show 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.

1 What aspects of the political system of modern Sweden seem to you most worthy of comment?
2 Comment on the political stability of modern Sweden.
3 Explain the political stability of modern Sweden.
4 Identify and discuss three factors that might help explain the political stability of modern Sweden.
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.
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 modernisation?
   (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.

The questions become progressively more specific – leaving the student less and less responsibility for deciding what is significant about the situation referred to, and how to respond to it. For instance, the student answering Question 1 may choose not to discuss political stability ...

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 ...

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 ... When the marker's response is guided in this way it may become more reliable, but at the cost of preventing a personal response to unique features of the student's work which cannot adequately be reflected in such scales. There is a balance to be struck here between reliability and flexibility.

Reliability can also be increased simply by aligning markers' understanding and behaviour more closely with what the course intended and with each other, without changing either the assignment or the criteria. The Open University does this by undertaking special marking exercises, briefing perhaps fifty associate lecturers who will be marking the same assignments, monitoring their marks and intervening where unwanted variations occur.

Of course, local agreements get made, so that in this department, critical thinking is treated in this way and in that department it gets understood like that. How may we trust that these are reliable judgements, in the sense that we can trust that they are not based upon one person's subjectivity, nor that they are artefacts of the tasks set? The measures that are most often used include: using multiple markers; making all markers use the same criteria; training markers to use the criteria; and using the same criteria to get judgements of several tasks. If we also want to know that the judgements are robust, in the sense that they apply to a variety of settings under different levels of constraint, then assessments need to be made in varied contexts. Extract 1.5 Reliability elaborated on the concept and the two pieces on improving reliability amplify these remarks. The point is that it is both complicated and expensive to get tolerably reliable judgements of complex outcomes of learning.

If local agreements on the meaning of complex learning outcomes are the best we can do, then we need to abandon the idea of having reliable measures of achievement. For sure, we can get locally reliable descriptions of achievement, but they are not measures. Measures are transferable. These descriptions are local. So, before I would trust the judgement that a student from department Z tends to think in original ways I would want to know more about what goes on there. Does the department provide affordances that favour original thinking or is there so much scaffolding that it is likely that what it calls original thinking I would call good synthesis?

As you can see, reliability, generalisability and validity are all tangled together. Reliable measures should give us some confidence that we can generalise to future performance. If a university transcript says that a student is excellent at critical thinking, then an employer would infer that this student will bring a critical mind to the workplace. But my experience as a new graduate working in a policy unit was that this inference was wrong. Sure I was smart at critical thinking in the sense of seeing the value of alternative perspectives and frameworks but I was useless at the nit-picking, detail-dredging critical thinking that was needed in this unit. The reliable judgement that I was a critical thinker had been based on a narrow range of tasks. Generalisations to new settings proved to be invalid. A more valid assessment of my critical thinking practices would have been based on performance in a wider range of settings calling upon different forms of critical thinking. This story reminds me that reliable judgements may not be valid, in that reliability may have been purchased by simplifying the thing that was supposed to be assessed. The price can be high because people reasonably assume that we can generalise from reliable
judgements whereas generalisation can be very misleading because the only generalisation upon which we can depend is from the reliable judgement to other circumstances identical to the ones on which the judgement was based.

I do not want to push too far this idea that reliability and validity are at loggerheads because we can, at a price, agree on indicators of achievement that we will use in a subject community (the QAA's subject benchmarks are an example). We can, with care, learn to interpret and apply them in similar ways, using external examiners and insisting on second, unseen marking of anonymous scripts. However, this is expensive, works within subject communities of practice but may not cross subject boundaries so well, and is easiest when done with fairly straightforward evidence of achievement, such as essays. Unfortunately, essay-writing is not much valued outside higher education, where there is a greater interest on how a complex achievement, such as critical thinking, is deployed in complex practical situations. Agreed, we can assess performance in these authentic situations but it is harder to do reliably (because it is all the harder – arguable impossible – to write criteria to identify critical thinking in the turbulent flow of practice) and much more expensive (because it costs more to observe practice).

We can reconcile the reliability–validity tension in three ways.

- Assess simple achievements, where the tension is at its least.
- Invest heavily in improving the reliability of judgements, where this is feasible.
- Take the (wrong) common-sense view that we ought to be able to assess with reliability and validity: the ostrich strategy.

I find it difficult to avoid the conclusion that some learning outcomes, notably information retention and the use of algorithms, are fairly determinate and can be assessed cheaply and reliably. Others can be pushed into fairly determinate shape and assessed fairly reliably, at a price. And, importantly, others resist measurement. Any summative assessment of those complex outcomes will have local meanings, no more. To be blunt: we cannot get – or we cannot afford to try and get – reliable measures of many of the learning outcomes valued in higher education.

This may seem like a hard way to make the point that module-level design decisions involve identifying the learning outcomes that will be assessed reliably, those that will be moulded so that some fairly reliable judgements can be made (albeit at some cost to validity), and those that will be assessed otherwise. I have taken so long to say it because the point is counter-intuitive, especially for those teachers who are at home in traditions that assume that anything of value can and ought to be measured in much the same ways as natural scientists go about measuring things. My claim is that assessment is not all things to all outcomes: different sorts of assessment are needed for different learning outcomes and the more complex the outcome (the more it is typical of higher education), the less likely that it is proper to think of measuring it. If you want to measure, stick to measuring information transmission. Other things may be assessed in their own appropriate ways.

So, what might this mean for the design of modules that are considerate of complex learning outcomes and that also provide data that can be used for grading purposes? Later in this chapter, I will suggest a differentiated approach. First, I need to introduce the important distinction between formative and summative assessment purposes.
1.6 'Formative and summative assessment purposes

The distinctions I have been making call to mind the well-known distinction between formative and summative assessment. I have gone into some detail about this distinction between formative and summative assessment purposes in a booklet for the Generic Centre of the UK Learning and Teaching Support Network (Knight, 2001a. It is available at www.ltsn.ac.uk/genericcentre/projects/assessment/).

![Diagram of summative and formative assessment purposes]

**Figure 1.4 Summative and formative assessment purposes**

Simply put, summative assessment is supposed to yield firm, high-stakes summaries of achievement. Formative judgements are intended to provide learners with feedback. Figure 1.4 extends this distinction. Figures 1.5 and 1.6 give a different view of it by treating it as a distinction between two different assessment systems. They are not, on this view, two faces of assessment but two different systems within a programme’s learning arrangements. Each approach has its own ‘rules of the game’ (shown as the object of assessment) and different tools and rules exist to guide and help the players. More radically, different assessment purposes make different assumptions about who does the assessing, as you can see by comparing Figures 1.5 and 1.6. One implication is that tasks which have formative purposes should be kept separate from those with summative purposes. This is illustrated in the following material.
Using journals and logs

Often when students submit work for assessment they are ‘faking good’: trying to give the best possible impression of what they know or can do, and covering up what they don’t understand for all they are worth. Consider this engineering student talking about submitting problem sheets to be marked:

The average lecturer likes to see the right result squared in red at the bottom of the test sheet, if possible with as few lines of calculation as possible – above all don’t put any
comments. He hates that. He thinks you are trying to fill the page with words to make the work look bigger. Don’t leave your mistakes there either, even corrected. If you’ve done it wrong, bin the lot. He likes to believe that you’ve found the right solution the first time. If you’re still making mistakes, that means you didn’t study enough. There is no way you can re-do an exercise a few months after because you’ve only got the plain result without comments. If you have a go you may well make the same mistakes you’ve done before because you’ve got no record of your previous errors. (Gibbs, 1992, p. 71)

This does not seem to provide the best possible basis for ensuring that assessment supports learning! Instead it would be more effective to encourage students to reveal how they went about their work and how they currently understand the concepts and methods they are learning about, so that the teacher can respond in a constructive way. This is an excellent illustration of the tension between formative and summative assessment. Coursework is meant to be formative but is here treated as summative by the student, who assumes that the teacher treats it similarly; this limits learning.

The following extract from Gibbs and Griffiths’ paper describes the use of journals and logs in the same engineering course that the student quoted above was studying. The journals are used by students to write down their reflections on how they tackle engineering problems, so that they learn which approaches work and which fail. Teachers comment on these for the sole purpose of encouraging more perceptive, accurate and revealing reflection. Comments tend to take the form of further questions such as ‘Does this always happen?’ or ‘Do you experience the same difficulties with free body diagrams?’ They focus on the process of learning and are never marked.

In this case study, logs are used to record work done in a way that explains what students understand and the development of their understanding. Students can use these logs to recognise the progress they have made or to go back and see how they thought their way through past problems. Teachers can use them as a vehicle for discussion in tutorials as they are very revealing about what students have learnt. They focus on the content of learning.

Students’ ability to tackle problems with awareness of what they are doing is vital; it is an ability from which all courses in engineering could benefit. This was the first course in this degree programme that attempted to develop this ability.
The UK Quality Assurance Agency (QAA) has never required that all programme learning outcomes be summatively assessed, although it does expect all to be touched by assessment systems. In other words, formative assessment is acceptable. (In fact there is some powerful evidence that it is more than acceptable: good learning is invigorated by good formative assessment. I summarise the evidence in the Generic Centre assessment materials cited above (www.ltsn.ac.uk/genericcentre/projects/assessment/).

As Figure 1.4 suggests, when the purposes are formative, the stakes are lower and there is less need for certainty. Ideally, there will be some parity of esteem between the learner and the commentator on their work, who may be a tutor, might be another student or, indeed, the learner herself or himself. If we see assessment as conversation rather than as decree, then the notion of certainty becomes redundant because 'a genuine conversation is', in the words of Hans-Georg Gadamer, 'never the one we wanted to conduct' (Gadamer, 1989, p.361). This presupposes some degree of parity between the people in conversation, which is quite counter to the spirit of summative assessment. There is also some psychological evidence (Moshman, 1999) that thinking is better in these more symmetrical power relations: when they are lop-sided, as with summative assessment, then the inferior concentrates on getting the message being sent by the superior rather than on engaging with the assessment task.

This is not a rejection of summative assessment, nor a denial that summative judgements can have formative value when they are fed back to the learner. Notice, though, that they may stimulate conformity rather than fresh thinking and there are practical objections to becoming too enthusiastic about the learning that might come from feedback on high-stakes assessments. First, there is the brute fact that examinations, summative assessment incarnate, rarely provide any feedback except a letter or percentage, let alone feedback that can help future learning. While this is partly the result of modular programmes, in which a term of learning is followed by a couple of weeks of intensive examination, there is no evidence that more traditional courses do any better in getting feedback from terminal assessments to students. Secondly, there are fears (Higgins et al., 2002) that students fix on the grades, not the comments. So, as you can see in the rather sad little vignette in the box below about a teacher who puts his soul into preparing feedback which students ignore, there are real fears that the case for valuing assessment primarily for the feedback it provides to guide further learning is overstated, if not plain naïve.

The tutor invited David to visit him in his office to talk about assessment.

The tutor explained how he put a lot of time into giving students written feedback on their essays. He took an essay from a large pile, and showed David the many detailed and helpful comments he had made on it.

David noticed the date on the essay cover. It was from the previous term. The tutor saw the look of surprise on David’s face as he registered the date.

‘It’s sad’, the tutor said, ‘but they often don’t pick up their essays’.

‘Why not?’

‘I don’t know.’

David Baume asks why this might be happening and recommends you, as teachers, to:
• Talk with your students about feedback.
• Tell them why you think it is important ...
• Ask them what they find useful and less useful in feedback on their work.
• Meet as many of their requests about feedback as you can.
• Continue to talk with them about feedback, even if only for a few minutes, if possible before each assignment and after they have received feedback on each assignment.

However, that is not the conclusion Higgins et al. (2002) reached, although it is true that they heard some concerns about the quality of some feedback. Some teachers give assignment feedback that is too specific (going into detail on material that is unlikely ever to be revisited) and others give vague advice ('you need to improve the structure of your writing'). On the basis of rather limited work, they concluded that:

Our research suggests that, while the grade may be of paramount importance to students, many of those we questioned are eager to read feedback comments. They expect feedback because they believe they deserve it ... Furthermore, there is a perception that higher education is a service and, as such, it is also the tutor's 'duty' to offer feedback ... the argument that feedback will be ignored or only used if it provides 'correct answers' cannot be sustained.

(Higgins et al., 2002 p. 61)

I have come to feel that the formative-summative distinction, like that between 'deep' and 'surface' approaches to learning, has an intuitive appeal that bypasses our critical defences. I found it surprisingly hard to explain the distinction when I tried in 1994 (Brown and Knight, 1994) and I wonder now whether the time is right to downplay this distinction, useful though it has been, and put something more sophisticated in its place.

Figure 1.7 offers an alternative, based on two dimensions. The first dimension is to do with whether assessment judgements are supposed to have local or general meaning. Most formative assessment has local meanings because the judgements are negotiable and tied to the contexts in which they arise. Public examinations and the certificates, licences and warrants that go with them are generalised statements of achievement, saying that a student has a certain proficiency or understanding. The second dimension runs from learning that is treated as determinate to that which is reckoned to be fundamentally complex. My hunch is that the formative–summative distinction is really getting at a distinction between assessment practices that engage with fuzzy learning outcomes and produce judgements that have rich local meanings but which don't generalise well. (as with many attempts to assess authentic achievement in the workplace) and those that home in on more determinate, clear achievements, providing data that can be fairly unproblematically generalised (such as measurements of information retention, skill at solving conventional engineering problems, or essay writing in history).
Figure 1.7  Local and generalised assessments

This may be well and good, but what help is it to the hard-pressed teacher to know that there are tensions between reliability and validity, and between formative and summative purposes that mean that good assessment practices are far more complicated than might have been imagined? What is to be done? I have two answers.

The first is to distinguish between:

- simple outcomes that can easily be assessed reliably
- complex ones that cannot and
- others that will be assessed quite reliably by dint of putting a lot of time and effort into them.

The Improving reliability (1) and (2) material in Section 1.6 contained a lot of ideas for doing a good job when reliability is the object of the exercise. A second answer involves tackling the objection that students won’t take formative assessment seriously by thinking in terms of passports or admission tickets.

Passports or admission tickets

Module teachers might be interested by these distinctions but they are likely to be more interested in workable advice on how to ‘square the circle’ and design assessment arrangements compact enough to fit a semester, reliable enough to contribute some summative judgements to programme-wide assessments of achievement and, at the same time, sufficiently valid to encourage complex learning.
I suggest starting with the need to produce tolerably reliable data. There are no particular problems here with the simpler, more determinate achievements but there is pressure to produce fairly reliable data about less determinate, more complex achievements. At module level, greater reliability is bought by investing in good grade indicators used by more than one, trained marker. In programme terms, there need to be procedures for aggregating judgements of the same learning outcome made at different points in the programme. Many teachers favour examinations because they want to be sure that the work students do is not plagiarised or done for them by others. Rather than resist that thinking, I suggest going with it and putting some time-constrained, individual assessments (TCIA), or examinations, into the assessment plan. Time-constrained, individual assessments can be designed into module assessment plans, fitted quite easily into normal class time and supported with the resources needed to maximise reliability. Because there are limits to how much students can write in two, perhaps three 30 to 40 minute TCIA s, it can be feasible for teachers to spend time trying to get tolerably reliable judgements of fairly complex learning outcomes. This is even more realistic if they are not also expected to spend a lot of time on other assessment tasks.

Some practice-focused programmes use Observation of Structured Clinical Experience (OSCE) to get a lot of fairly reliable data about clinical competence. Observations of Structured Clinical Experience involve students assembling in a hall in which there are a large number of ‘stations’. Each station contains a practical challenge – taking a patient history, showing how to bandage a shoulder, examining an X-ray – and an assessor. Students spend a short time at each station where they have to meet the challenge under the eye of the invigilator, who assigns a score to the performance by reference to standard grading criteria. Short answer questions are written variants. Here students are given problems and told to write responses in, say, fifty words or so. In both cases the point is that short, focused tasks can provide a lot of data about performance and therefore provide a fair basis for making reliable judgements of competence.

Plainly, if resources are invested in TCIA s, they are not available for other assessment tasks, which implies that these tasks cannot be ones that need a lot of teacher attention. But how are students to be encouraged to take seriously assessment tasks which are not examinations and which are not going to get close attention from teachers? Or, to put it rather differently, why should students take low-stakes, often formative assessments seriously when they will naturally be fixated on preparing for their time-constrained individual assessments? Part of an answer assumes that we can persuade students to do the rational thing and take formative assessment seriously by helping them to understand the intrinsic value of formative assessment, and combining this with profiling or personal development planning (Chapter 4). Many teachers will be sceptical of that answer. An alternative is to make ‘satisfactory’ engagement with formative or low-stakes tasks mandatory. Students who do not engage with these tasks do not qualify for the summative assessment – for the examinations. These low-stakes assessments are the equivalents of tickets to a lottery or a passport to travel. Some teachers also allow students to submit a file of their formative assessed work as supplementary evidence, to be considered by assessors if a student’s provisional marks for a module fall just below a class threshold.
This produces a very differentiated system of module assessment, with plenty of low-stakes tasks, many of which are intended to engage students with those learning outcomes most resistant to reliable assessment. These tasks provide feedback for better learning and help students to develop portfolios supporting their claims to achievement. Other learning outcomes are clearly identified with more expensive summative assessment procedures. They could take the form of relatively short high-stakes TCIA tasks, perhaps done in class, which require students to reflect upon, analyse or apply what they have learned through earlier low-stakes tasks. For example:

- Students who have worked on principles of small-scale research design in the social sciences could be told that a week 6 class will be given over to an open-book examination comprising one task: design a study to meet one of the three research briefs printed on the examination paper.
- A portfolio of lab reports could lead to an examination question about good laboratory practices and problems in maintaining them.
- A portfolio of tasks completed during a module could lead to a question asking students to summarise the key themes of the module in order to defend a position on a central substantive issue.
- ‘Research literacy’, which is the ability to read research reports intelligently, could be summatively assessed by a structured, in-class question about a paper distributed 24 hours earlier.

**Reflection 1.3 A differentiated approach to assessment in your module**

I recognise that some teachers do not have the power to change module assessment arrangements, so this is intended as a ‘what if?’ exercise.

I have reviewed several forms of differentiation in assessment, particularly differences in methods and purposes. The culmination has been my suggestion that many tasks can be qualificatory (they have to be done but do not contribute to module grades) or treated as supplementary evidence (they are reviewed and taken into account if module marks fall just short of a grade boundary). High-stakes assessments could then be carefully designed to shed light on target learning outcomes and carefully marked against robust grade indicators. They could be short, piggy-backing on a run of formative tasks to show what students have learned from them.

I invite you to reflect on a module you teach, considering:

- how assessment arrangements could be best linked to module learning outcomes.
- how the range of assessment methods in use might be altered to allow a better match between assessment practices and learning outcomes.
- whether there would be advantages in differentiating more sharply between qualificatory, low-stakes assessment tasks for formative purposes, which might be mainly peer- or self-assessed, and tutor-marked high-stakes assessments with the summative purpose of producing tolerably reliable evidence of achievement in relation to selected learning outcomes.

You may find it particularly helpful to talk to a colleague about your reflections on differentiated assessment.
1.7 Planning module-level assessment arrangements – problems and issues

I have been arguing that module-level assessment practices could be much more systematic than they often are. In the process I have been saying that assessment, learning, teaching and curriculum are all enmeshed and that it makes little sense to talk about one without thinking carefully about the others. In this section, I want to begin to deal with three possible objections to the line I have taken: it is unrealistically systematic; it doesn’t take account of wide-choice programmes; and short, content-heavy modules will not fit the model. In each of these cases, I agree that my preferred approach to assessment could be criticised because of its apparent incompatibility with the ways in which curricula are typically framed in HEIs in the UK.

1. Your approach is too systematic – people don’t really plan modules this way

The approach I have been describing could be put in a family known as ‘rational curriculum planning’ (RCP). There are quite a lot of objections to RCP and one of the most telling is that, whatever theorists say we ought to do, people don’t actually use RCP: even if they’re supposed to they often don’t. I have to make this point by referring to school teaching, which has been much more thoroughly studied than teaching in higher education. Schoolteachers are likely to plan their teaching sequences quite differently from the sequence I have just summarised, even though they are routinely advised to use RCP (Clark and Peterson, 1986). In practice, once they have identified the topic or content area to be covered, they recall successful ‘lessons in memory’ and good tasks. Then they look for ways of adapting this experience of ‘what works’ to the new material (Calderhead, 1996). Sequencing, pacing and editing decisions often follow as they try to construct a roughly viable instructional sequence. Then they are likely to see what learning outcomes might reasonably be expected to arise from this organisation of the material and to check back to their scheme of work (equivalent to a programme specification in higher education) to see whether there are any unfortunate gaps.

I have a great deal of time for this approach, not least because it is so widely used that I assume it must have some good sense behind it. I think it has a lot to offer as a way of designing modules, always given that it does not become a cloak for conservatism. Although it makes sense to stick with ‘what works’, practices that have succeeded in the past may not be the best ones for carrying forward today’s programme learning outcomes. Besides, there is the obvious danger that some outcomes simply get overlooked, although good programme-level audits of assessment practices should reveal gaps like these. A more subtle danger is that teachers claim that what they already do really promotes new learning outcomes and leave it at that. However, unless students are clear about the learning intentions – unless teachers spell out the intended learning outcomes and highlight the ways in which they are engaged by assessment practices – then such claims will be little more than rhetoric. The danger is one of false consciousness, of claiming to assess something that is, in practice, ignored. It may look good on paper to say that all outcomes are assessed, albeit in different ways, but in practice there is a danger that those that are not reliably assessed – those that are covered mainly by local judgements – will not get the same respect as those that are. Tokenism may prevail, with neither students nor teachers taking them seriously.

2. What about wide-choice programmes?

By ‘wide-choice programmes’ I mean those that allow students to construct their own degree from a menu of hundreds of modules. An extreme example would
be the Independent Studies programme at Lancaster, to which I'm quite attached. After the first year, students negotiate what they are going to study and how: they create their own designer degrees. How compatible is my advocacy of aligning module learning outcomes and assessment practices with wide-choice programmes?

In a sense there is no great problem for individual teachers who only need to select from the programme specification and make sure, through validation, that their selection causes no particular problems for programme leaders. In a wide-choice programme there are not likely to be difficulties because the programme will have been designed to accommodate diversity, so teachers can more or less do what they like at the moment — although, I think this laissez-faire approach is under threat.

However, I want to signal another theme to be developed in Chapter 3, which is that this systemic approach with its emphasis on coherence is now causing problems for the designers of wide-choice programmes. If they are to write programme specifications and then to approve module assessment arrangements that should result in programme graduates having similar achievements to show for similar awards, then they are pushed in one of three directions. The first is to write vague aims and objectives, which are of little help to the teacher trying to design or re-validate a single module because they are too open-ended. The second is diametrically opposite, and involves insisting that courses select from a limited range of programme learning outcomes. Given the amount of choice available to students, this is a hard path to take. The third position is a compromise, a programme specification with a core set of learning outcomes, some of which must be addressed by every course, and a much larger set of outcomes from which course designers may select if they wish. Programme designers who take this position are likely to be much more concerned about the design of core modules and those that are, or are likely to be, very popular. Specialised options may not worry them very much.

I agree, then, that wide-choice programme designers have a very difficult task to do, whichever pathway is chosen. In fact, it may be impossible to preserve diversity and provide programme specifications that are of any use to anyone.

3. What about short, content-heavy modules?

There are considerable assessment difficulties when teachers feel that they have to cover a great deal of material in a ten- or twelve-week module. At first sight, it is not clear what the problem should be because if the aim is to make sure that students know the textbook material, then there are no great assessment problems — it's quite easy to assess information retention cheaply and reliably. But a programme that is supposed to be developing more complex learning outcomes really needs modules that do more than cover the material. They need modules to contribute to the development of social practices, to foster reflection and promote 'higher order skills' such as analysis, synthesis and evaluation. Quite reasonably, teachers protest that they cannot cover all of the material and find the time for students to engage with it in ways that encourage critical thinking, autonomy, group work, and suchlike, let alone assess such a broad range of learning outcomes.

In the material headed Assessment and your students in Section 1.1, we heard that lecturers can have unrealistically high expectations of students, set too much reading and overload students. This is a clue that when lecturers protest that they cannot find the space to assess a range of learning outcomes, part of the problem may be that their modules are poorly designed spoil-tips of material. With that in mind, I present below a lengthy reading. Extract 1.8 Planning assessment for a new course, which contains rather a good description of how a
module can be planned so that worthwhile assessment tasks are designed to
cover a good range of learning outcomes. It is supplemented by some additional
material under the heading Allocating marks to assessment components, which
brings in important elements about marking practices that are not covered in
Extract 1.8.
In the following paragraphs my colleague Roger Harrison and I consider how we might assess a new two-semester course on *National Energy Policy*. The course is intended for second-year undergraduate students, most of whom will be studying economics, business management or planning as their main subject. There will be about 80 such students and the course is meant to occupy them for 10 hours a week for a fifteen-week semester: 150 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-emphasized. 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.

**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.

**Outcomes** The student should be able to:

1. List the various forms of energy and describe how one form may be converted into another.
2. Explain why energy is important in modern industrial society.
3. Perform simple calculations concerning energy transformation.
4. Decide appropriate methods to determine the power of an energy transforming device.
5. Explain quantitatively why there are limitations to the proportion of thermal energy which may be transformed into useful energy.
6. List the commercial sources of energy currently and prospectively available and estimate reserves or potential of each.
7. Identify the principal energy flow in a given community (or industry, or factory, etc.).
8. Identify the energy inputs into a given product using data provided.
9. Evaluate a technical proposal with regard to energy consumption.
10. Suggest ways of reducing energy consumption in particular contexts, and assess their feasibility.
11. Identify the economic factors relating to the use of energy.
12. Form a judgement as to the weight to be given to economic considerations in evaluating energy proposals.
13. Apply what has been learnt to real-world situations demanding that the student exercise initiative, selectivity, understanding, judgement, and social skills.

[The outcomes are arranged across three sections.] In brief, the first section of the course lays the scientific foundations; the second section brings in the economic and social issues; and the third section requires the student to apply the theories and concepts acquired so far.

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 attained on the way through. That is, some *integrative* assessment activity seems to be called for. This could be an essay towards the end of the third section 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 Objective 13 clearly suggests a *project*. In fact we can decide
that the last five weeks of the course will be devoted to an empirical enquiry related, for example, to Objectives 7 and 8. Students could negotiate the precise context of their projects and will carry it out in a local factory, hospital or whatever. To include the social skills aspect of Objective 13 and to save on supervision time and practical arrangements with such large student numbers this could be a project which was undertaken in its investigation stage, on site, in a group, but analysed and written up afterwards individually. Such a project will give rise to a report (with graphs, calculations etc.). Clearly this ought to count for the purposes of overall assessment since it relates to a vital learning outcome to which a third of the course time was allocated.

At the same time, if students are concentrating on their projects during the last part of the course they cannot be expected also to produce an integrative essay or prepare for an exam, and a project plus an exam would seem excessive in a single semester. This means that the project has to encompass more than the empirical enquiry to address Objectives 7 and 8 and has to include Objectives 9 to 12. Each project report should therefore be extended to address each of these Objectives in the form of short essay-type answers to a series of set questions, but in relation to the context of each project.

So we have decided on a project, with additional questions, towards the end of the semester. Is anything more needed? Well the scientific/experimental background early in the course 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 Objectives 1 to 6 two or three times during the first six weeks in order to keep students working consistently. Probably multiple choice questions will suffice and these could be computer-based, using some standard testing software, to save marking time and to collate student results quickly and conveniently. While these objective tests are vital we will not pretend that they are very high level. It would not be appropriate to let the tests contribute too much to the students’ overall grade for the course. Let us regard them as qualifying tests. That is, students have got to get satisfactory scores in order to be allowed to continue with the course, but the scores 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. We will have to build in an opportunity to re-take the tests for those who fall below this standard, but with computer-based testing this should not be difficult.

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

1. computer-based multiple choice qualifying tests during the first six weeks (Objectives 1 to 6);
2. project report, during the final five weeks (Objectives 7, 8 and 13);
3. short essay-type questions added to the project report (Objectives 9 to 12).

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 the most and which the least? We can see no reason why either the project (Objectives 7 and 8) or the project questions (Objectives 9 to 12) should be given greater importance, so let them count equally. However, Objective 13 implies five criteria (initiative, selectivity, understanding, judgement, and social skills) by which overall performance could be judged. We could decide to allocate 40% of marks to the project, 40% to the project questions and 20% to these five criteria: 4% to each. We will need to give some thought as to how students can present evidence of initiative and social skills.

So these are 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 around about assessment into the course before the course structure begins to harden. Because, for example, the suggestion that we assess on the basis of a project report in the last five weeks is, indeed, a suggestion also as to the structure of the course. Assessment is a direct expression of the teaching strategy.

The following paragraphs – a discussion of this design process between Graham Gibbs and Derek Rowntree – are followed by some helpful material concerning the allocation of marks to assessment components.
Allocating marks to assessment components

To assess a range of outcomes you will probably need a range of assessment components, each of which assesses a different outcome. Allocating marks to assessment components so that they ‘count’ towards students’ overall marks is a way of making sure that students pay attention to them. It may only be necessary to allocate 4% to each of ten weekly lab reports, for example, to get students to take them seriously. Spreading the marks widely over a range of assignments will spread student effort and also reduce the risks to students of doing badly (or missing deadlines) for any particular component, and so will reduce anxiety. The potential drawbacks include:

- a heavy marking load (though economical assessment methods such as peer assessment are available);
- narrowing the focus of what students study, so that they do nothing except assessed work (though if the assessment tasks are well designed this need not be a problem);
- making it less likely that students will take risks or be creative in their approach to assignments, in case the risks do not pay off in terms of marks (though you could reward creativity in your criteria);
- overloading students with work, which may result in them taking a surface approach to their studies.

Marks from assessment components have to be added together to provide a single mark for the course and decisions have to be made about the weighting of these components. For example, should the coursework count for 25% or 75%? The weighting of assessment components can determine students’ overall marks, their ranking and the average marks for the course independently of their performance on individual components of the assessment. The following are just some of the issues to bear in mind:

- Do the components assess different levels of outcome, and if so would you want to give more weight to higher level outcomes such as those involving application or analysis?
- Is the course organised hierarchically, with the assignments that students tackle towards the end encompassing everything they have learnt at earlier stages, and if so would you want to place more weight on these integrative assignments?
• Are some components much bigger than others, in terms of the time students need to spend on them, and if so would you want either to reward effort or to orient student effort accordingly?

• Are some assessments much better than others at distinguishing between students? You might not want to allocate many marks to an assignment on which most students gain a ‘B’ grade, as often happens with seminar presentations, for example.

• Are there guidelines about the proportion of the total marks that can be allocated to coursework? Institutional rules may specify anything from zero to 100%.

• Are there local conventions about the number of marks that are allocated to particular types of coursework, for example 5% to a lab report, 20% to an essay, 40% to a project?

In allocating marks, and in considering the complex issues associated with designing assessment as support for learning, it is worth remembering that modules are selections from all the material that could be covered, and one way of finding space for good assessment practices could be to re-think the basis for selection. If higher education is about ways in which students get, evaluate and use information, rather than about how much information they get, then modules should concentrate on those intellectual processes, not on getting students to write out an encyclopaedia in weekly instalments. European thinking about the first cycle of higher education (the undergraduate years) is moving in this direction, with the first cycle being characterised as the development of generic achievements that become specialised in the second, postgraduate cycle. In other words, postgraduate students will have to know a lot about their area of specialisation while undergraduates will be less concerned with information acquisition and more with becoming fluent in the processes and practices of a domain. However, to return to a theme introduced at the beginning of this chapter (Section 1.3), this assumes that students get the message.

I have already used some of the material that appears in Extract 1.9 Differences in conceptions of assessment tasks as Extract 1.1, but I think it is worth repeating. It shows that tutors in humanities and social sciences prefer cogent arguments to more or less relevant arrangements of information but unfortunately, some of the students interviewed in this research project hadn’t got the message. So, there is little point in devising good modules that promote understanding of principles and practices over mastery of information if students are going to be assessed in ways that signal the opposite. Good assessment needs ‘knowing’ students.
Two groups of second-year university students took part in the empirical investigation: seventeen whose main subject was history and sixteen whose main subject was psychology ...

The main findings stem from analysis of the interview accounts. Within each of the two subject groups, a fundamental difference was identified between the students' conceptions of what an essay was and what essay-writing involved. The full set of categories of description derived (Hounsell, 1984) cannot be presented here for reasons of space, but Table 1 provides a summary of the main differences in conception. Some history students conceived of essay-writing as a question of argument, coherently presented and well-substantiated; others saw it as concerned with the arrangement of facts and ideas. And amongst the psychology students, essay writing was seen by some as a matter of cogency, where substantive discussion was rooted in a solid and coherent core of empirical findings, and by others as relevance, in the sense of an ordered presentation of material pertaining to a topic or problem.

Table 1 summarizes the differences in students' conceptions of what was required of them in history and psychology essays respectively.

In the first group are the two interpretative conceptions of argument (history) and cogency (psychology). What these two conceptions share is a concern with the making of meaning: an essay is seen as a mode of discourse through which one makes sense of a topic or problem in a way which is individually distinctive ... The student is thus the arbiter of the content and form of the essay, making personal choices about the selection and organization of material, whether in the light of an explicit interpretative stance which had crystallized at an early stage or of a steadily growing sense of the dominant theme. Introductions and conclusions are not simply conventions but devices through which a unity of form and content can be facilitated.

The second group of conceptions comprises the two non-interpretative ones of arrangement (history) and relevance (psychology). In each, an orientation towards the establishment of meaning is absent, and each is correspondingly less explicit than its interpretative counterpart. Although there is an acknowledgement that essays are a medium through which one's own ideas, thoughts and opinions can be conveyed, these tend to have an almost incidental status or are seen as 'value added' rather than as the essay's main justification ... In consequence, students' descriptions of their essay-writing procedures appear flat and mechanical, a series of steps which seem uninformed or undirected by considerations of meaning. And the students' accounts of the content of their essays typically do not include a substantive reference to a conclusion.

### Table 1  Conceptions of essay writing

<table>
<thead>
<tr>
<th></th>
<th>History</th>
<th>Psychology</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ARRANGEMENT</strong></td>
<td>Students define an essay as the ordered presentation embracing facts and ideas</td>
<td>Students define an essay as a well-integrated and firmly grounded discussion of a topic or problem</td>
</tr>
<tr>
<td><strong>ARGUMENT</strong></td>
<td>Students define an essay as an ordered presentation of an argument well supported by evidence</td>
<td></td>
</tr>
<tr>
<td><strong>RELEVANCE</strong></td>
<td>Students define an essay as an ordered discussion of relevant material on a topic or problem</td>
<td></td>
</tr>
</tbody>
</table>
No less importantly, such differences in conception seem to have a more general association with students’ grasp of a discipline. Students ascribed to an interpretative conception tended to see the craft of essay-writing as mirroring the practice of the discipline (in history) or (in psychology) as an activity both fostering and demonstrating one’s mastery of the accumulated knowledge of the discipline. The non-interpretative conceptions, by contrast, do not have these associations. There are instead abundant indications of vagueness and uncertainty about the nature of essays and essay writing and a perceived gap between aspiration and what is achieved in the finished essay.

ACCOUNTING FOR DIFFERENCES IN CONCEPTION

It is not difficult to find parallels between these differences in conceptions of essay-writing and other findings from research into student learning. The concern with abstraction of meaning characteristic of the interpretative conceptions has evident affinities with a deep approach to learning (see for example Marton and Säljö, 1984), contextual relativistic reasoning (Perry, 1970) and a thematic conception of learning (Säljö, 1982). Similarly, the non-interpretative conceptions of essay-writing share the quantitative and reproductive associations of a surface approach, a taken-for-granted conception of learning, and some aspects of the pre-relativistic positions within Perry’s scheme of intellectual development. But how can such differences be accounted for – and in particular, how is it that the non-interpretative conceptions of essay-writing both subsist and apparently persist?

One possible source is the course context within which the activity of essay-writing takes place. At the most obvious level, the interview analysis pinpoints contextual features which could be considered dysfunctional. In both subject groups, the traffic of comment on essays was almost overwhelmingly from tutor to student and in written form. It was only very seldom that one-to-one oral discussion of essays took place, and there was virtually no substantive discussion of essay-writing among peers: whatever discussion did occur was confined to close personal friends and touched upon deadlines, choice of essay topics and useful sources, rather than weightier issues such as options in the treatment of an essay theme or essay-writing strategies. Nonetheless, and despite some variation in the amount of written feedback given, both groups of students were issued with written guidelines on departmental expectations of essays and received page-specific and general comments on individual pieces of work. Broadly speaking then, it does not seem plausible to attribute differences in conception to a manifest lack of guidance on what essays entailed ...

... where students’ conceptions of essay-writing are qualitatively different from those of their tutors, communication cannot readily take place because the premises underlying the two disparate conceptions are not shared or mutually understood. Students misconstrue a tutor’s comments or guidance or fail to grasp the import of these because they do not have a grasp of the assumptions about the nature of academic discourse underlying what is being conveyed to them ... this gulf in communication can nonetheless be found within the interview accounts. Perhaps the most telling of these show up in students’ reactions to specific comments. One psychology student, for example, finds herself thrown by comments which in effect challenge her understanding of expectations, but she also apparently interprets argument in a literal and everyday sense rather than an academic one:

Ellie:

I felt, in actual fact, I’d covered the area very comprehensively, by trying to bring in as many angles as I could. I tried to cover all the different areas. But one of the tutor’s criticisms was why did I just keep going from one to the other. But I thought that’s what I was supposed to do.

Interviewer:

So did you feel you had a clearer sense of what the tutor was looking for when you got the essay back?

Ellie:

Well, from the comments on the essay I gathered the tutor wanted me to argue, about something, but I mean, by presenting the material as the research had demonstrated, it was a mild form of argument. I wasn’t going to get aggressive, in an essay.

75
... guidance about generalization and a concern with details is not firmly apprehended because, for the students concerned, there is not an appropriate conception within which it can be readily construed:

Interviewer:

What did you think the tutor was looking for in this essay?

Pattie:

Ah ... well, this is confusing me. I know the tutor likes concise work, but doesn't like generalizations, and doesn't like too much detail, although on the whole I think he'd like more detail than generalizations. And because it was such a general question, I thought 'oh help!', I don't know what he's looking for.

... students may indicate their awareness (at least, embryonically) that something different is expected, but feel powerless to achieve it.

Wendy (Psychology):

I think really that what the tutor wanted was for us to say what has been said, and then what you think -- Is there a reason for it? But I don't really think I did that. (Laughs) I think it was just what other people said.

Brenda (History):

I don't really have a structure to (my essays). They say you're supposed to, and some of the tutors like you to have, you know, themes, where you deal with various themes as you go along, but mine don't, they just ... as it occurs to me. On the whole I just write it down as it comes to me.

Indeed, nowhere is this important (and seemingly baffling) awareness of expectations more apparent than in what is perhaps the crucible of an academic essay, the conclusion.

Rick (History):

I might draw a conclusion, if I have time, and draw all the threads together. If not, I might just finish, you know, just finish like that.

Brenda (History):

Sometimes I can finish with a quote, sometimes I can sum up with my own feelings, or sometimes it just kind of gets to where there's nothing more to write, but you can't think of anything to sum up.

Rosemary (History):

My conclusions are always terrible. The tutor said if I rewrote the conclusion -- 'cos it was one of those I didn't conclude, he would re-mark it.

Ellie (Psychology):

I'd done this amazing structure (laughs) and then I hadn't put in any ... evaluative conclusion.

But we should also note the dysfunctional side-effects of feedback which does not seem to engage with the student's understanding of the activity. A specific consequence is to see tutor's comments as marginal:

Donna (History):

I never really find them very helpful. (Laughs) Just irrelevant, really.

or as having nothing to offer beyond the confines of a particular essay assignment.

Interviewer:

Generally speaking, do you find tutors' comments helpful?

Holly (Psychology):

Not unless you get that title again, no.

Peter (Psychology):

Sometimes I do read the comments but I find that, I'll never write the same essay again anyway ... I tend to ignore them in some ways, unless there's something very startling.

But there is an even more worrying consequence of repeatedly low marks and of feedback which is critical in tone and opaque in its signification. This is to treat essay-writing extrinsically, as a chore which deserves only minimal effort or an activity whose value lies elsewhere. For example:

Rick (History):

If I knew I could do a good essay, I'd do more research. Sometimes I just give up.

Ellie (Psychology):

Most students really dread [essays]. Perhaps they could be of use, but I am going to just treat them ... with the view in mind that I'm going to try and use them for revision. And that's it.
Frank (History):
It's just work, in a way. Just all these essays, and reading's the worst part, it's just labouring really ... I think it tends to kill the interest, in fact.

Peter (Psychology):
I find that my main objective, when turning out an essay, is just to get two and a half thousand words and finish.

Such comments suggest firstly, that motivation may be less a personal attribute than a function of the relations between an individual and a situation, and secondly, that some students may be locked into a 'cycle of deprivation' as far as constructive feedback is concerned. Since feedback fails to connect, it comes to be viewed as insignificant or invalid, and so is not given considered attention.

In this sub-section on content-heavy modules I have been suggesting that complaints that there is not enough time to use a range of assessment methods may be saying rather a lot about the complainants. There are possibilities, especially:

- If we see ourselves teaching for understanding rather than for information mastery.
- If we capitalise on ideas about passports, tickets and hurdles (Section 1.6).
- If we adopt more efficient practices (Section 2.3).

The following material is offered in support of a tentative claim that time – our most valuable resource -- may also be saved by adopting formative assessment practices using C&IT in working practices.

Formative self-assessment by students
1.8 Summary

This chapter has tried to suggest that, on the one hand, we need a more differentiated approach to assessment and that, on the other, many of the assessment problems that worry module teachers are not so much their problems as course designers' problems.

Although the chapter has contained a lot of information about assessment techniques, more systematic attention needs to be given to practical matters. That is the concern of Chapter 2.