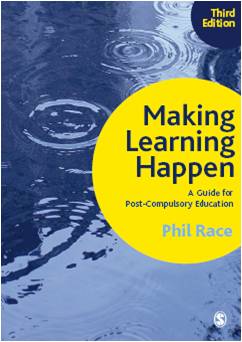
**Assessment Digest: exams, essays, and more?**

**(Phil Race, February, 2015)**

This digest is made up of selected extracts from my two recent books, ‘Making Learning Happen’ (3rd edition, 2014, Sage) and ‘The Lecturer’s Toolkit’ (4th edition, 2015, Routledge), and is designed to flesh out workshops I run on assessment, particularly those where I encourage staff to address the weaknesses of some kinds of assessment which remain in widespread use in higher education assessment. The ‘commentary’ bits of this digest are in ComicMS, and the text in Calibri is taken directly from my disc versions of the manuscripts (the published editions of course have the benefits of excellent copy-editing work from the respective publishers). The first part is taken from ‘Making Learning Happen’ linking exams and essays to the seven factors underpinning successful learning, the theme of the whole book, then leading up to a table critically comparing a range of types of assessment (which was new to the 3rd edition).

Then follow extracts from ‘The Lecturer’s Toolkit’, going into rather more detail about the advantages and disadvantages of various kinds of exam, and essays.

**Extracts from ‘Making Learning Happen; 3rd edition (2014, Sage)**

**(Taken from Chapter 4: ‘Assessment driving learning’)**

## Assessment is broken in higher education!

Several indicators tell us this, including:

* Students’ perceptions (for example, as gathered from final-year students annually in the National Student Survey in the UK since 2005) tell us that assessment and feedback continue to be the least satisfactory elements of the higher education experience, that the assessment criteria are not clear, and that students do not believe that marking and grading are fair.
* Academic staff responding to the starter ‘Teaching would be much better for me if only I...’ mention ‘had more time’ more frequently than any other factor, and when further questioned usually cite the time it takes to mark students’ work as the main oppressing factor. Indeed, staff often grumble that they have to spend more than twice as much time marking students’ work as they actually spend with their students.

# How has assessment become broken in higher education?

What are the factors which make it harder and harder to make assessment work well for students, and keep the processes of marking and giving feedback to students manageable for staff? The factors which have caused our present problems include:

1. Student numbers have grown: we can’t use the same processes and instruments for a system where nearly 50 per cent of the 18–21 year-old population study at university level, compared to 5 per cent a couple of decades ago.

2. The world has opened up, so that our assessment processes and practices need to be more compatible with those in markedly different cultures and traditions.

3. It is widely accepted now that assessment is the major driver for student learning, and if assessment is not working as a *good* driver for learning, learning is not happening well and the effectiveness of our entire higher education provision is jeopardized.

4. We now know much more about the standards we should strive towards in assessing students’ work, in particular the need to improve assessment to make it more valid, more fair, more transparent to students, better linked to the world outside higher education, and more inclusive so that assessment does not disadvantage students with identified special needs, and enables them to demonstrate their learning in ways where they can show their optimum achievement.

5. Rapidly increasing usage of online assessment, including online submission of coursework assignments, and the increasing awareness of the difficulty in establishing veracity in many contexts – in other words, who exactly did the coursework?

6. We need to continue to diversify the assessment processes and instruments we use, so that no students are repeatedly disadvantaged by the predominance of particular assessment formats.

In short, we need a richer mix of high-quality assessment formats, and we also need to *reduce* the overall burden of assessment for ourselves and for our students. We need to measure less, but measure it better. We need to measure a wider spectrum of students’ evidence of achievement, with a broader more versatile set of tools. Presently, we spend far too much of our (and students’) time on things they write, at the expense of other ways they can show evidence of how they have achieved the learning outcomes. But to repeat myself, we still need to *reduce* the burden of assessment: more assessments, but much shorter ones.

Fortunately, help is at hand. There now exists a rich literature about assessment, containing a great deal of authoritative wisdom to bring to bear on assessment design, helping us to address the factors listed above, and work towards making student perceptions of assessment more satisfactory, at the same time as allowing us to make it a more reasonable and realistic proportion of our jobs as academics in post-compulsory education.

# What students think

In the National Student Survey, administered to all final-year students every year in the UK since 2005, in the section on ‘assessment and feedback’, statements 5 and 6 link directly to students’ experience of assessment. The design of this particular survey is far from ideal, but its use right across the higher education sector in the UK makes it an important indicator of at least some of students’ feelings about their higher education experience. Students are asked to make judgements as follows:

* definitely agree
* mostly agree
* neither agree nor disagree
* mostly disagree
* definitely disagree
* not applicable

on the following two statements (numbers 5 and 6 out of the 22 in the survey):

5 The criteria used in marking have been clear in advance.

6 Assessment arrangements and marking have been fair.

Widely across the sector, institutions have been dismayed at students’ negative responses to these statements, compared to most of the rest of the survey, and this trend has continued throughout the use of the Survey as implemented so far. Many factors can account for student dissatisfaction with assessment, including:

* For students, assessment is of course the sharp end of their overall experience in higher education – it’s what determines their qualifications.
* When students are paying for their higher education, they are quick to pinpoint assessment as one of the most important things they are buying, and naturally start looking hard at what they are getting for their money.
* Even when the assessment criteria have been clearly formulated in module handbooks or on course websites, students may not have *heard* the criteria sufficiently in face-to-face contexts, allowing them to know better exactly what they mean in terms of evidence of achievement of the intended learning outcomes.
* Even when assessment and marking *are* fair, students often do not know the lengths institutions go to, to achieve this. Students may know little of the detail of assessment questions being moderated and improved by external examiners, for example.
* Overall, students do not have sufficient opportunity to get their heads inside the assessment culture of higher education – it is sometimes a ‘black box’ to them, where they do their best and hope it will all work out well in the end.
* Students have not sufficient opportunity to really get their heads around the assessment criteria applied to their work, for example by *making informed judgements* themselves on their own and each other’s work.

Sometimes, of course, students are right, and assessment is far from being perfect! Assessment criteria are not always made clear to them in advance, and sometimes assessment and marking are far from fair! However, going back to the design of the survey, both of the statements involved in this section are in fact ‘bipolar’. In other words, is it ‘clear’ or ‘in advance’ (or both) which triggers the student’s choice of options for statement ‘5’, and is it ‘assessment arrangements’ or ‘marking’ (or both) being ‘fair’ which triggers students’ choices for statement ‘6’? I regard it as unfortunate that such a widely used instrument incorporates (and repeats annually!) such clumsiness! Nevertheless, the discussion in this chapter aims to address all of the reasons why students are dissatisfied with the assessment culture they meet in higher education.

Similar problems with assessment are shared in many parts of the world. Flint and Johnson (2011) in Australia have built their excellent book ‘Towards fairer university assessment’ around the concerns of students themselves, which are very close to the reservations expressed above. They report that:

Student evaluations frequently reveal poor assessment practices that:

1. Lack authenticity and relevance to real world tasks;
2. Make unreasonable demands on students;
3. Are narrow in scope;
4. Have little long-term benefit;
5. Fail to reward genuine effort;
6. Have unclear expectations and assessment criteria;
7. Fail to provide adequate feedback to students;
8. Rely heavily on factual recall rather than on higher-order thinking and problem-solving skills.

(Flint and Johnson, 2011, p2)

I have now showed this list to several thousand staff at workshops and keynotes, each time inviting them to ‘raise a hand if not guilty on any of these eight counts’ and to date have only had one hand raised (by someone who had just been appointed the day before).

# Assessment and deep, surface and strategic learners

*Deep learning* gets a good press in the scholarly literature. Deep learning is, we might argue, closer to developing real *understanding.* But we’ve already seen that this is difficult or even impossible to measure. So deep learning may be the wrong approach to wean our learners towards when our assessment may only be measuring something rather less than deep learning. Deep learning may, of course, be much more appropriate for those learners going on to higher levels, and is doubtless the kind of learning which leads to the most productive and inspired research. Perhaps that is why deep learning is regarded so favourably by educational researchers on the whole. However, ‘Save your deep learning for your postgraduate years. For now, your priority is to make sure that you getto *having* some postgraduate years’ could be wise advice to give undergraduates!

*Surface learning* gets a bad press in the literature. However, probably most of the learning done by most people in post-compulsory education is actually only surface learning. Learners learn things ‘sufficient to the day’ – the exam day or the assessment week or whatever. When it has been learned successfully enough to serve its purpose – pass the module, gain the certificate, whatever – it’s ditched. It’s not entirely wasted, however. Something that’s been surface learned is a better starting point for re-learning, or for learning more deeply, than something which has not been learned at all. But learners can all tell us tales of the countless things they have learned only well enough to give back when required to demonstrate their achievements, which have been quite deliberately ‘eased out’ of their minds as they moved on to the next stage of their learning journey. ‘You are what you learn’ may be a noble sentiment, but it can be argued that our assessment processes and instruments cause learners to learn far too many things which aren’t important, diluting the quality of learning that is afforded to those things that *are* important.

Despite the criticisms of surface learning approaches, sometimes it is a fit-for-purpose choice. Where a limited amount of factual information needs to be available at will in a particular scenario, but will not be needed after that scenario is completed, surface learning can be a wise enough choice. There are things that just are not important enough to warrant a lot of time and energy being invested in learning them deeply. An example could be the statistics relating to stopping distances in wet and dry conditions, which need to be learned to pass parts of the driving test in the UK. Few experienced drivers can quote these facts and figures correctly a few years after passing their driving tests, but probably are perfectly capable of judging stopping distances well enough simply based on experience. This aspect of the learning for the test seems to be almost entirely a surface learning business.

(There are several additional sections in Chapter 4, which I have left out of this digest, so it could focus more on exams and essays, rather than the whole picture of assessment).

# What’s wrong with *strategic* learning?

Strategic learning has perhaps had the worst press of all. It’s not just *accidental* surface learning. It is perhaps sometimes *deliberate* surface learning, consciously engaged in at the expense of deeper learning. Strategic learning is regarded as ‘learning for the exam’. It’s associated with ‘seeking out the marks or credit’ quite consciously in essays, reports, dissertations and theses, and extends readily to preparing strategically for job interviews, promotion boards, and so on. Moreover, it can be argued that strategic learners may make informed judgements about *what* to learn deeply and what to learn *just* at a surface level.

Strategic learners tend to be successful, or at least moderately successful. Deep learners may well *deserve* success, but quite often shoot themselves in one foot or the other by mastering *some* parts of the curriculum very very well while leaving other parts of the curriculum underdeveloped, and not getting the overall credit that they might have achieved had they spread their efforts more evenly across the curriculum. Surface learners can also fare well enough if and when all that is really being measured in our assessment systems is surface learning. Strategic learning is often thought of in terms of doing the *minimum* to get by. But there are various ‘minima’. In the present degree classification system in the UK, perhaps there’s the minimum to get by and get a degree at all, the (different) minimum to get by and get a 2.1, the (different again) minimum to get by and get a first-class degree, and perhaps the minimum to get by and get a first-class degree with a margin for safety?

So what *is* strategic learning? We could regard it as making informed choices about when to be a deep learner and when to be a surface learner. It could be viewed as investing more in what is important to learn and less in what is less important to learn. It could be regarded as setting out towards a chosen level of achievement and working systematically to become able to demonstrate that level of achievement in each contributing assessment element.

There is growing recognition that the curriculum in post-compulsory education is content-bound. There is just so much subject matter around in every discipline. Any award-bearing programme of study necessarily involves making informed decisions about what to include in the curriculum and what to leave out. But is not this the very same thing that strategic learners do? Isn’t being an *effective* strategic learner to do with making wise and informed choices about where to invest time and energy, and where not? It can be argued that strategic learning, when done well, is a demonstration of a useful kind of *intelligence –* that of handling quite vast amounts of information, narrowing the information down to a smaller proportion and then processing only that smaller proportion into knowledge. It can also be argued that those learners who go far are the strategic ones, rather than the deep ones. It can be argued that they know *when* to adopt a deep approach and when it is sufficient to adopt a surface approach. In the UK, for example, every year there is an annual clamour about the A level results. The clamour echoes the usual protests that standards have not fallen, that there has been no ‘dumbing down’. Could it not be that A level candidates are becoming better prepared to achieve at A level? Could it not be that they know more about what is being looked for in good exam answers? Could it not be that they are more aware about what is required for good grades in associated coursework? Could it not, indeed, be that they are now better versed in the virtues of strategic learning? And is this really a ‘bad thing’?

# Validity, fairness, ‘whodunit’?, transparency and real-world dimensions

We’ve already seen that it is widely accepted that for most learners assessment drives learning to a quite profound extent. This is particularly the case for cue-seeking learners and strategic learners, and unsurprisingly they fare best in most common assessment processes and procedures. But is this state of things satisfactory? Institutional policies on teaching, learning and assessment make much of the design of assessment processes and instruments being adjusted to address the following qualities:

* validity – how well are we actually measuring what we’re supposed to be measuring – evidence of achievement of intended learning outcomes?
* fairness – how reliable is the assessment is – e.g. will different markers award identical marks for the same piece of work or exam script? Fairness can also be thought of as ‘justice’.
* ‘whodunit’? – was the assessed work done by the candidate, or by other people? Another word for this could be ‘veracity’ (but I quite like ‘whodunit?’ as it reminds us of the central question involved).
* transparency – how well do learners know how the system works?
* real world dimensions – how closely does the assessment link to things learners may need to do in employment? This issue is also often referred to as authenticity.

So assessment should be valid, fair, transparent and linked to the real world, and we need to be certain regarding who did the assessed work. Anyone who cares about the quality of the assessment they design for learners will say how they strive to make it so. We are also *required* in the UK, for example, to make assessment valid, fair, and transparent in higher education by the Quality Assurance Agency. As an example of what we need to do to ‘mend’ assessment, it is worth listing in full in this Chapter the eighteen indicators of sound practice listed in the QAA’s ‘UK Quality Code for Higher Education’ (2013) in the section on ‘Assuring and enhancing academic quality’ for the ‘Assessment of students and the recognition of prior learning’ as follows:

The basis for effective assessment

1. Higher education providers operate effective policies and processes which ensure that the standard for each award is rigorously set and maintained at the appropriate level, and which ensure that student performance is equitably judged against this standard.
2. Assessment policies, regulations and processes, including those for the recognition of prior learning, are explicit, transparent and accessible to all intended audiences.
3. Those who might be eligible for recognition of prior learning are made aware of the opportunities available, and are supported throughout the process of application for recognition and its assessment.
4. Higher education providers assure themselves that everyone involved in the assessment of prior learning, student work, and associated assessment processes are competent to undertake their roles and responsibilities.
5. Assessment and feedback practices are informed by reflection, consideration of professional practice, and subject-specific and educational scholarship.
6. Students are provided with opportunities to develop an understanding of, and the necessary skills to demonstrate, good academic practice.
7. Staff engage students to promote a shared understanding of the basis on which academic judgements are made.
8. Feedback on assessment is timely, constructive and appropriately developmental.
9. The volume, timing and nature of assessment enable students to demonstrate the extent to which they have achieved the intended learning outcomes.
10. Assessment tasks provide, through inclusive design wherever possible, and through individual reasonable adjustments wherever required, every student with an equal opportunity to demonstrate their achievement.
11. Assessment is carried out securely.
12. Degree awarding bodies assure themselves that the standards of their awards are not compromised as a result of the language in which assessment is conducted.
13. Processes for marking assessments and for moderating marks are clearly articulated and consistently operated by those involved in the assessment process.
14. Higher education providers operate processes for preventing, identifying, investigating and responding to unacceptable academic practice.
15. Degree awarding bodies specify clearly the membership, procedures, powers and accountability of examination boards and assessment panels, including those dealing with the recognition of prior learning; this information is available to all members of such boards.
16. Boards of examiners/assessment panels apply fairly and consistently regulations for progression within, and transfer between, programmes and for the attainment of awards.
17. The decisions of examination boards and assessment panels are recorded accurately, and communicated to students promptly and in accordance with stated timescales.
18. Degree awarding bodies systematically evaluate and enhance their regulations and assessment processes and practices.

Institutional teaching and learning strategies in the UK are now required to embrace these qualities in the design of assessment. But hang on – why have we all got ‘teaching and learning’ strategies in our institutions? Why have most institutions got ‘teaching and learning’ committees? (Or, indeed, ‘learning and teaching’ committees – small difference?) Why haven’t we got ‘teaching, learning and assessment’ strategies – or, indeed, ‘assessment, learning and teaching’ committees, which would be the way round I would name them? Because assessment is the weakest link, I suggest. It’s much easier (and safer) to fiddle around with the quality of teaching or learning than to tackle the big one: assessment. It’s actually quite hard to *prove* that some teaching has been unsatisfactory, but only too easy to demonstrate when something has gone wrong with assessment. But, as shown below, there are significant shortfalls in the extent to which many of the most common assessment practices measure up to bringing these qualities to bear on assessment, particularly as we shall see, exams and essays.

## Validity?

Valid assessment is about measuring that which we should be trying to measure. But still too often, we don’t succeed in this intention. We measure what we can. We measure echoes of what we’re trying to measure. We measure ghosts of the manifestation of the achievement of learning outcomes by learners. Whenever we’re just ending up measuring what they *write* about what they *remember* about what they once *thought* (or what we once *said* to them in our classes), we’re measuring ghosts. Now, if we were measuring what they could now *do* with what they’d *processed* from what they thought, it would be better.

‘But we *do* measure this!’ assessors justify themselves. But ask learners, they know better than anyone else in the picture exactly what we end up measuring. For a start, let’s remind ourselves that we’re still very hung up on measuring what learners *write.* We don’t say in our learning outcomes ‘When you’ve studied this module you’ll be able to write neatly, quickly and eloquently about it so as to demonstrate to us your understanding of it’. And what do we actually measure? We measure, to at least some extent, the neatness, speed and eloquence of learners’ writing. What about those who aren’t good at writing? Or to be more critical, what about those learners who have at least some measure of *disability* when it comes to writing? And in exams, the ‘writing’ tends to be *handwriting*. Where else, one may ask, in this age of keyboards, computers, tablets and smartphones are significant tracts of wording composed with a pen in hand? How divorced has our assessment become from everyday life?

In the UK, the writing is on the wall for us regarding any tendency for our assessment instruments and processes to discriminate against learners with disabilities. Since 2002, the Special Educational Needs and Disabilities Act (SENDA) and subsequent amendments to disability legislation have caused us to make far-reaching changes to our assessment just to keep it within the law. We are required to make ‘reasonable adjustments’ so that no learner should be unfairly discriminated against by our education provision, not least the assessment-related aspects of this provision. Legislation also requires these reasonable adjustments to be made in an *anticipatory* manner. In other words, they should not just deal with instances of discrimination when they are found to have happened. This is a tricky situation, as in one sense the purpose of assessment *is* to *discriminate* between learners, and to find which learners have mastered the syllabus best, and least, and so on. If we’re honestly discriminating in terms of ability, that might be lawful. But if we’re discriminating in terms of disability, it won’t be lawful. But aren’t they the same thing? Where does ability stop and disability begin?

For a long time already, there have been those of us strongly arguing the case for diversifying assessment, so that the same learners aren’t discriminated against *repeatedly* because they don’t happen to be skilled at those forms of assessment that we over-use (such as, in some disciplines, tutor-marked, time-constrained, unseen, written examinations, tutor-marked coursework essays and tutor-marked practical reports).

We’re entering an era where *inclusive* assessment will be much more firmly on the agenda than it has ever been to date. We now know much more about the manifestations of dyslexia in assessment, and are beginning to work out the effects of dyscalcula, dysgraphia, dyspraxia, and so on. Many of us are beginning to realize for the first time that, even in that packed lecture theatre, we do have learners with disabilities, not just the occasional learner visibly in a wheelchair, but perhaps a quarter or a third of our learners who are affected at some times in their learning by factors which we don’t know about and which many of them don’t even know about themselves. So is it ever going to be possible for us, in our assessment practices, to be satisfied with the levels of validity or fairness to which we aspire?

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

## Fairness?

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

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

Sadler (2009a) suggests four propositions which have a bearing on the difficulties we face in ensuring that assessment is fair and reliable. He proposes:

(1) Students deserve to have their work graded strictly according to its quality, without their responses to the same or similar tasks being compared with those of other students in their group, and without regard to the students’ individual histories of previous achievement.

(2) Students deserve to know the basis on which judgements are made about the quality of their work. There should be few if any surprises.

(3) Students deserve their grades to have comparable value across courses in the academic program in which they enrol, and across the institution. Courses should not exhibit characteristically tough or lenient grading.

(4) Students deserve grades that are broadly comparable across institutions and maintain value over time, so that the standing of their educational qualifications is protected not only by the college or university in which they study, but also in higher education as a social institution. (Sadler, 2009a: 809)

Sadler’s propositions here show that there is much more to be achieved, in the context of fairness, than simply reliability within the marking of a set of assignments.

So why else is fairness so important? Not least, because assessing learners’ work is the single most important thing we ever do for them. Many staff in education regard themselves as teachers, with assessment as an additional chore (not to mention those who regard themselves as *researchers*,with teaching and assessing as additional chores). Perhaps if we were all to be called *assessors* rather than teachers it would help? And perhaps better still, if we all regarded ourselves as researchers into assessment, alongside anything else we were researching into, it would help more? ‘Students can escape bad teaching, but they can’t escape bad assessment’ says David Boud (1995).

In countries with a degree classification system, our assessments can end up with learners getting first-class degrees or thirds. This affects the rest of their lives. Now if our assessment were really fair, we could sleep easily about who got firsts or thirds. The learners who worked hardest would get better degrees and the learners who lazed around wouldn’t. This indeed is often the case, but most of us can think of exceptions, where learners got good degrees but didn’t really deserve them, or where learners who seemed worthy of good degrees didn’t come up with the goods in the assessed components of their courses, so we couldn’t award these to them. So perhaps it’s not just that our assessment isn’t too fair, it’s our discrimination that’s sometimes faulty too.

## Whodunit?

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

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

However, probably too much of the energy which is being put into tackling plagiarism is devoted to *detecting* the symptoms and punishing those found guilty of unfairly passing off other people’s work as their own. After all, where are the moral and ethical borderlines? In many parts of the world, to quote back a teacher’s words in an exam answer or coursework assignment is culturally accepted as ‘honouring the teacher’. When learners from these cultures, who happen to be continuing their studies in the UK, find themselves accused of plagiarism, they are surprised at our attitude. Prevention is better than the cure. We need to be much more careful to explain exactly what is acceptable and what is not. While some learners may deliberately engage in plagiarism, many others find themselves in trouble because they were not fully aware of how they are expected to treat other people’s work. Sometimes they simply do not fully understand how they are expected to cite others’ work in their own discussions, or how to follow the appropriate referencing conventions.

It is also worth facing up to the difficulty of the question ‘Where are the borderlines between originality and plagiarism?’ In a sense, true originality is extremely rare. In most disciplines, it is seldom possible to write anything without having already been influenced by what has been done before, what has been read, what has been heard, and so on.

## Transparency?

One way of describing ‘transparency’ is the extent to which learners know where the goalposts are. The goalposts, we may argue, are laid down by the intended learning outcomes, matched nicely to the assessment criteria which specify the standards to which these intended outcomes are to be demonstrated through evidence produced by learners, and also specify the forms in which learners will present that evidence of their achievement. There’s a nice sense of closure matching up assessment criteria to intended learning outcomes. But how well do learners themselves appreciate these links? How well, indeed, do assessors themselves consciously exercise their assessment-decision judgements to consolidate these links? Learners often admit that one of their main problems is that they still don’t really know where the goalposts lie, even despite our best efforts to spell out syllabus content in terms of intended learning outcomes in course handbooks and to illustrate to learners during our teaching the exact nature of the associated assessment criteria – and sometimes even our attempts to clarify the evidence indicators associated with achievement of the learning outcomes are not clear enough to learners. In other words, learners often find it hard to get their heads inside our assessment culture – the very culture which will determine the level of their awards.

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

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

**The real world**

Under the heading ‘whodunit?’ we’ve so far taken up the matter of *ownership* of assessed work. There is, however, another thing to consider – the extent to which the work being assessed relates to the real world beyond post-compulsory education. We need to be thinking more about making assessed tasks as close as possible to the performances that learners will need to develop in their lives and careers in the real world. Doctors, lawyers, accountants and managers don’t, in their day-to-day work sit *writing* about medicine, law, accountancy and management – they *do* these things. There is often a considerable gap between what we get learners to do in our assessment, and what they will need to be good at throughout their careers. The more we can bridge that gap by making assessment feel relevant to learners, the more we can expect them to take ownership of the need to become able to evidence their achievement in assessed contexts.

## A critical look at exams and essays!

In the analysis which follows, I am selecting two of the most common assessment processes, traditional exams and essays, and suggesting how they may impact on the factors underpinning successful learning. Although I am only interrogating two of the available assessment processes and instruments, they presently represent a large proportion of the assessment in post-compulsory education in the UK, for example, and I hope that this may help you to look in a similar way at other assessment processes you employ, and think through the implications in parallel to my analysis below. The analysis which follows is based not just on my work helping teaching staff in post-compulsory education to develop assessment processes and instruments, but even more on my parallel work over three decades in helping learners to develop the skills they need to demonstrate their optimum performance in a range of different assessment conditions and environments.

After this analysis, I would like to widen the field, and get you thinking about the pros and cons of a wider range of assessment possibilities, continuing to hold in mind the detail of the two particular kinds of assessment interrogated below against the factors underpinning learning.

# Traditional exams

In particular, let’s take the example of time-constrained, unseen written examinations. In other words, candidates don’t know the questions until they see them in the exam room. They work against the clock, on their own, with pen and paper. Assessment systems in the UK are quite dominated by this kind of assessment, usually at the end point of increments of learning. The assessment can therefore be described as summative.

As an assessment process, exams can be *reliable or fair –* if there is a well-constructed marking scheme, each candidate can be reasonably confident that the marking will be fair and consistent.

The main problem with many traditional exams is that they don’t rate highly on *validity.* In other words, too often they measure what the candidate can *write* about what they have learned, in the relatively artificial conditions of solemn silence, against the clock. Where, however, exams are based on problem-solving, case study analysis, and so on, validity can be much higher.

Exams can be improved in terms of *transparency* where candidates have been involved in applying assessment criteria to their own or other people’s exam answers, and have found out all they need to know about how the examiner’s mind works.

One of the major advantages of exams is that we are reasonably certain (with due precautions) that the work of the learner is being marked – in other words, that side of *whodunit?* is assured. The the extent to which the assessed performance relates to the normal conditions in which the learning is intended to be applied – is less assured, and in some traditional exams the conditions under which achievement are measured are quite alien.

## 1 Traditional exams and wanting to learn

For many exam candidates, the ‘want’ to learn is damaged by the mere thought of looming exams. Many learners, if given the choice, go for learning modules that are continuously assessed rather than assessed by examination because of their fear – and even dread – of exams. Few assessment processes induce such high emotions. This is not the case for everyone, however. Some candidates love exams – and are very good at preparing for them and doing them. Not surprisingly, the cue-seekers mentioned earlier in this chapter are among those who are good at traditional exams. Their cue-seeking approach is thus rewarded by this pervasive assessment format.

## 2 Traditional exams and needing to learn

This is where the intended learning outcomes should come into their own. Ideally, if learners have systematically prepared to demonstrate their achievement of these outcomes, and practised doing so sufficiently, they should automatically remain able to demonstrate the same achievements under time-constrained, written exam conditions. However, there is often a gulf between the intended learning outcomes as published and what is *actually* measured by traditional exams. Due attention to achieving constructive alignment can overcome this problem. But there is another side to needing to learn. Candidates who prepare successfully for exams by mastering the intended learning outcomes so that they can demonstrate their achievement in answering likely exam questions often concentrate very firmly on what they perceive they need to learn, and don’t invest time or energy in things they decide can’t (or won’t) come up in the exams. We are therefore favouring strategic learners by the use of exams (and, of course, cue-seeking strategic learners do best).

## 3 Traditional exams and learning by doing

There is plenty of learning by doing *before* traditional exams. But not much further learning by doing happens *during* traditional exams. It can, however, be claimed that a looming exam is as good a way as any of causing learners to get their heads down and do some learning. We could argue, however, that preparing for an oral exam (viva) would have just as much effect on learning by doing.

## 4 Traditional exams and learning through feedback

This is where traditional exams do really badly. As far as feedback is concerned, they are mostly lost opportunities. By the time the scripts are marked, learners have often moved on to the next part of their learning and are no longer really interested in which questions they answered well and why, or (more importantly) in where they lost marks. Many learners were *very* interested in these matters immediately *after* the exam, and spent some hours in post-mortem mode trying to work out how their answers measured up to what was being looked for by their examiners. All the feedback that most learners receive – after some time – is their score, or their grade, or simply whether they passed or failed. It is feedback of a sort, but hardly formative feedback. We can, of course, argue that exams are intended to be summative measures, but they still seem to represent lost feedback opportunities. Where feedback *is* provided very quickly after an exam (for example, in computer-marked multiple-choice exams, where a feedback printout can be handed to each candidate on leaving the exam room), feedback can, indeed, play a much more powerful role even in summative testing.

## 5 Traditional exams and making sense of what is being learned

This, too, links badly to traditional exams. As with learning by doing, a great deal of making sense of the subject matter occurs *before* an exam and, indeed, could be argued to be happening *because of* the exam. But few exam candidates report later that the moment when the light dawned was *during* the exam. More often, they report that they only found out that the light *had not* dawned during the exam. And then we need to ask whether traditional exams are measuring the extent to which learners have made sense of what they learned. Too often, exams seem to measure what learners can *reproduce* rather than what they can *do.* Many learners can tell us about the frequent occasions where surface learning was all that they needed to engage in to address the task of answering a particular exam question.

## 6 Traditional exams and learning through verbalising

In Chapter 2, we explored the significant benefits to learners resulting from them verbalising orally, explaining things to each other, coaching each other, and so on. Sadly, traditional exams only really get at verbalising in writing – and with a pen at that usually. One of my worries about traditional exams is that learners tend not to discuss things with each other, but to go into competitive mode, and hide their learning achievements from each other rather than celebrating such achievements. We can, of course, try to counter this tendency, and encourage learners to work together in their preparation for exams, quizzing each other, explaining things to each other, and so on, providing good rehearsal for doing similar things in writing on their own in the exam room.

## 7 Traditional exams and learning through assessing – making informed judgements

We could argue that in the context of traditional exams, most of this kind of learning occurs in the minds of examiners, not learners! Too often, *exactly* *how* the informed judgements are made by examiners is hidden from learners; examiners seem to fear the consequences of sharing with learners details of how marking schemes work in practice, possibly dreading future appeals by learners against ‘academic judgement’. We can indeed encourage learners to self-assess practice exam performance, and to peer-assess each other’s practice as they head towards exams, but the competitive ethos of exams militates against them doing either of these wholeheartedly. Moreover, there is fear involved – fear of finding out that the performance is not going to be up to the standards desired – and this can lead to self-fulfilling prophecy and lower attainment in exams.

Perhaps the main problem regarding learning through making informed judgements in the context of traditional exams is that such assessments are often the ‘mystery black box’ in nature, where learners do the best they can, hoping it will be found to be satisfactory or better. Other assessment formats tend to be more open to learners regarding exactly how they work (though this is not always the case).

## Traditional exams: summary

The picture painted above of the links between traditional exams and the factors underpinning successful learning is very bleak. It does not *have to be* so bleak, however. With care, for example, exams can be designed which are much better at measuring ‘making sense’ than suggested above. Problem-solving exams and case study exams are much better at *not* rewarding reproductive learning. But the concerns remain about the damage that can be inflicted on many candidates’ *want* to learn, the artificial way that exams can skew the *need* to learn and the fact that so much work may be done by examiners making sure that the exams have been fair and reliable, yet very little feedback usually reaches learners. In some ways, it seems that traditional exams are diametrically opposed to all of the central factors underpinning successful learning! Couple this to the problems of achieving validity, fairness and transparency, and it is surprising that in some assessment cultures (including much post-compulsory education provision in the UK) traditional exams continue to hold sway to the extent that they do.

# Other kinds of exams

The discussion above focused on the most common kind of exams – against-the-clock, written exams, and with candidates not seeing the questions until they sit at their exam desks. There are, however, many other kinds of exam, which overcome some of the problems about fairness, validity, transparency and links to the real world in suitable contexts and discipline areas. These alternatives can also be thought of in terms of the factors underpinning successful learning, and some ‘food for thought’ implications are summarized below for just two of the alternatives.

## Computer-marked multiple-choice exams

If candidates are aware that it is their decision-making that will be measured rather than their ability to put their knowledge into words in writing, then *wanting* to learn will not be threatened as much as it is by traditional exams. Ownership of the relevant *need* to learn can also be improved, so long as learners become practised and rehearsed regarding *which aspects* of their achievement of the intended learning outcomes can indeed be measured by this sort of exam. *Learning by doing* in such exams is primarily of the decision-making variety, but with skilful attention to the design of questions and option choices, decision-making can cumulatively be used to yield a good measure of the extent to which learners have *made sense* of what they have learned. At least we can be assured that the *learning by doing* that is measured by computer-assisted assessment is not skewed by such mundane factors as the speed of handwriting or its legibility. Perhaps the most significant link between computer-assisted assessment and making learning happen is *feedback.* There are many possibilities. Learners can be provided with on-screen feedback as they go through a computer-based exam, allowing them to avoid the possibility of carrying forward errors of thinking into their answers to the next questions they meet. Or they can be given feedback on-screen or in print-outs at the end of each exam, when at least the feedback is quick enough for them to still remember what their thinking – and their decisions – were as they answered the questions. The availability of speedy and specific feedback can help learners to *make sense* of the subject matter they have been working with, admittedly too late for the computer-based exam they have just undertaken, but better late than not at all. Learning by *making informed judgements* is of course involved in multiple-choice exams, especially when there is plenty of this kind of practice in advance of the actual test. Learning by *explaining* can be involved if much of this practice is done in small groups of students, discussing the reasoning for choosing particular options.

I include more discussion about multiple-choice exams in Section 4.3 of this chapter, where you may find it easier to compare its pros and cons with those of several other assessment approaches.

## OSCEs

Objective structured clinical examinations (OSCEs) are widely used in medical education and health care studies, and lend themselves to many other disciplines where practical *doing* is important in the intended learning outcomes. Essentially, OSCEs are exams where each candidate *does* something at each of a number of assessment stations located around the exam room. In medicine, for example, candidates may visit successive stations and perform a series of assessed tasks, such as:

* interpreting some X-rays
* looking through a set of notes on a patient and approaching a diagnosis
* prescribing medication for a given condition in a given context
* briefing a ward sister about the pre-operative preparation of a patient
* talking to a patient to diagnose a condition (though in practice the ‘patient’ is an actor, as it is hard to get real patients to tell the same story to successive doctors).

The key claim made for OSCEs is that the assessment is valid, in that candidates are assessed on exactly the sorts of things they have been intended to become able to *do* in practice and not just on what they may have written in traditional exams about hypothetical cases.

Clearly, OSCEs link closely to learning by doing – practice, repetition and trial and error. Furthermore, the more feedback candidates get on their practice before such an exam, the more they can improve their performance. OSCEs also link strongly to well-defined *needing to learn* agendas, and as practitioners can see the relevance of developing their skills and knowledge to cope with such situations, the *want* to learn is enhanced. The variety of tasks which can be built into an OSCE add to the depth of *making sense* of what is being learned and assessed, as triangulation is possible, approaching key tasks from different angles. While it can take a considerable amount of time to design a good OSCE, when candidate numbers are large, this is time well spent, and the time spent *marking* an OSCE can be much less than a corresponding written exam, not least because most of the assessment decisions can be made at the assessment stations while the exam is in progress. In practice, it is wise to get groups of learners to design OSCE scenarios – they will often design better ones than we can! This also maximizes the learning pay-off they gain from discussing and arguing with each other, and making informed judgements about the material involved and the assessment criteria being addressed.

# Essays

In some subject areas (notable exceptions include maths, science and technology-based disciplines), essays are key elements of both coursework and exams. We can again pose questions about how successfully essays relate to validity, fairness, transparency and the world outside education. Essays do not do very well as an assessment method on such interrogation. But perhaps worst of all, essays (except handwritten ones in exams) fall short regarding ‘whodunit’?, as it becomes everyday easier to download information, and paste it into word-processed essays. At least in exam-based essays we can be reasonably certain whose work is being marked, but in coursework essays we can’t. However, in time-constrained essay-type exams we are perhaps penalizing the slower learners – perhaps by measuring speed of writing rather than quality of thought.

There are also particular problems with *fairness* where subjectivity in marking is all too easily present and inter-marker reliability is a problem (different markers giving the same essay different marks), as also is intra-marker reliability (the same marker giving the same essay different marks on different occasions – for example, among the first half-dozen marked or the last half-dozen marked).

*Validity* is perhaps the weakest link for essays as an assessment device. If we look hard at ‘what are we *really* measuring?’, it is often essay-writing *skills* rather than mastery of the subject matter concerned. Academics often defend the importance of essay-writing skills, but in practice for most learners, these tend to be skills that they are unlikely to need when they leave post-compulsory education, unless they too are heading towards becoming academics! Moreover, writing scholarly contributions to the literature involves much more than essay-writing skills – not least addressing fully the intended target audience of the writing and critically reviewing the existing literature in the field. Writing essays for an examiner is just a limited special case of this wider picture.

*Transparency* can be improved a lot by involving learners in self-assessing and peer-assessing essays so that they become much more aware of how marks are earned and lost, and how the assessment criteria work in practice – and, indeed, how the assessment links to the associated intended learning outcomes.

Connections between essays and the real world is problematic. The link between essays and the context in which learning may be intended to be applied – is often tenuous. There are many learners in post-compulsory education who will never again put pen to paper (or fingers to keyboard) to compose an essay after leaving education.

Meanwhile, let’s continue with our analysis of how essays may relate to the factors underpinning successful learning. I should point out at once that there are *very* significant differences here between coursework essays (with feedback in due course) and exam-based essays. As many factors relating to the latter overlap with what I’ve already said about traditional exams, the discussion which follows is mostly about the coursework essays.

## 1 Essays and wanting to learn

The effects here are widely variable. Some learners really enjoy ‘sorting out their minds’ by putting pen to paper to construct essays, particularly when they then get detailed and helpful feedback on their learning. Such feedback is unlikely to be forthcoming for exam-based essays. For other learners, actually getting round to putting pen to paper (or fingers to keyboard) is a major challenge. Ask a group of learners ‘What was your best work-avoidance tactic that you used to delay starting to put together that essay?’ and you will soon see how, for many learners, the task of getting started was the daunting part.

## 2 Essays and needing to learn

On one level, essays help learners to take ownership of the need to learn, by giving them something to do to cause them to get their heads into the books and resources relating to the task. However, the agenda of taking ownership of the intended learning outcomes is less successfully addressed, as all too often the links between these outcomes and a particular essay-writing task are not spelled out clearly enough in the briefings learners receive.

## 3 Essays and learning by doing

Essays certainly involve learning by doing. There are several kinds of *doing* in play, including information retrieval and sorting, planning, communicating in writing, comparing and contrasting ideas, making decisions and judgements, and summarizing. So this aspect of learning can be regarded as being satisfactorily addressed by the use of essays. Similarly, during the processes of drafting and redrafting an essay, a great deal of reflection and deepening of ideas can take place, and the act of writing the essay becomes much more than simply learning by doing.

However, it is worth asking how many of the same aspects of learning by doing are involved in constructing *essay plans* rather than fully-fledged essays. Such plans may miss out on some of the finer points of communicating in writing and on the reflective dimension, but making essay plans can involve many of the other important aspects of learning by doing. And if, let us suppose, 10 essay plans can be produced in the same time as it takes to write one fully-fledged essay, the learning pay-off associated with writing essay plans becomes all the more attractive.

Where, however, essays are primarily being used to train learners in the arts of marshalling their ideas, presenting them coherently and logically, and coming to a well-thought out conclusion or summary, and these are the primary intended learning outcomes, writing full essays will meet these aims to a much greater extent than simply preparing essay plans.

## 4 Essays and learning through feedback

Coursework essays can be very valuable in the context of making feedback available to learners. Feedback in general is discussed in more detail in the next chapter of this book. Meanwhile, it is worth bearing in mind that the timing and nature of the feedback on formative essays need to be managed well for optimum learning through feedback. It can be well worth considering ensuring that at least some of the feedback can be intentionally developmental.For example, if an essay is ‘marked’ three times, once where feedback is given on an essay plan, again when a rough draft is submitted and, finally, when the last version of the essay is completed, feedback on the first two stages can lead to much higher quality in the final products. This clearly takes extra assessor time, but the two earlier feedback stages do not need to be quantitatively ‘marked’, and can be required simply as conditions to be satisfied before the final essay version is submitted. However, most higher education institutions don’t offer feedback on draft work, as learners who’ve been used to formative feedback at school find to their dismay – the handed-in version is the one that determines their mark.

## 5 Essays and making sense of what is being learned

Coursework essays coupled with formative feedback can be very valuable in helping learners to get their heads around ideas and concepts, and also in helping them make sense of other people’s ideas from the literature. It is often the act of trying to communicate an idea which causes the human brain to clarify it and put it into perspective. This is equally true of oral responses, but writing out ideas and progressively making them more coherent is probably one of the best ways of causing reflection and deepening learning. ‘I don’t know what I think until I’ve written about it’ is said by many authors, who recognize the value of putting ideas down on paper as a way of helping the brain to make sense of them. Coursework essays can also cause learners to find and retrieve information from the literature and from other sources, and then to sift it and analyse it, and distil from the source materials their own conclusions or thinking about a topic, issue or question.

## 6 Essays and learning by verbalising – explaining orally

Though coursework essays do indeed involve the development of skills of *written* explaining, essay-writing tends to be a somewhat solitary activity, and learners are rather unlikely to spontaneously involve themselves in orally explaining things to each other or coaching each other, especially if the competitive nature of assessment-by-essays is in the forefront of their consciousness. We can, of course, encourage learners to think that their final essays are likely to be much better if they have spent a fair amount of time and energy working together at least before nearing their final drafts, but both learners and assessors have justified worries that at least some plagiarism may then happen.

## 7 Essays and learning through assessing – making informed judgements

More often than not, it’s the assessors who gain all of this learning – not the learners. This can be counteracted by well-planned use of peer assessment, allowing learners to benefit from seeing work that is better than their own and worse than their own. Furthermore, learners can be encouraged to undertake self-assessment of their coursework essays, using the same criteria as will be used by their assessors. This practice can also help them to become more self-assessing when they come to write essays in exams.

## Essays: summary

As can be seen from the above analysis, essays used formatively in a coursework context (rather than summatively in exam contexts) can involve many of the seven factors underpinning successful learning. Perhaps partly because they are time-consuming to plan, draft and polish, they are perhaps better than many assessment-related artefacts in enabling reflection and consolidation (important aspects of ‘making sense’). They are, however, often solitary learning journeys, at least until the points where feedback is received. Peer-review, peer-assessment and peer-editing processes can be used profitably to enable learners to benefit from feedback along the way.

## A critical comparison of eleven assessment types In the discussion in the previous section, two of the most commonly used kinds of assessment (traditional exams and essays) fared rather poorly in one way or another, when we looked at how they linked to making learning happen. In the next section, we will compare several assessment types (including these traditional ones) in terms of advantages and disadvantages, but this time also discussing validity, fairness, whodunit?, links to the real world, and the extent to which feedback to learners can be provided and made useful.

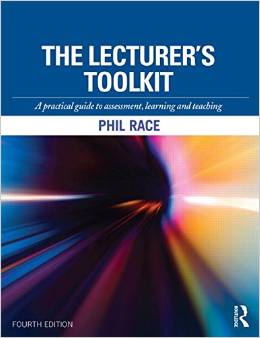
The table below illustrates the pros and cons of several kinds of assessment. In each case, in the ‘status’ column I have included judgements regarding how well (or how poorly) each assessment type listed measures up to validity, fairness, whodunit? (whether there could be serious doubts or not regarding whose work is being assessed), links to the real world, and the extent to which feedback to learners may be available, or useful.

| **Type of assessment** | **Status** | **Advantages** | **Disadvantages** |
| --- | --- | --- | --- |
| **1 Traditional Exams**  Exams are often referred to as the ‘gold standard’ because of their widespread use in secondary and higher education. Still the most common kinds of exam, are handwritten, invigilated, against-the-clock, with questions not being known by candidates beforehand.  Exams remain prevalent on many post-compulsory education courses, sometimes where questions are set by external examiners, and sometimes by the staff who teach the learners. Timescales vary, but two hours and three hours are relatively common in universities, though much shorter exams are perfectly possible (and perhaps desirable).  Many traditional exams offer candidates a choice of questions (e.g. attempt any 5 out of 8 questions, each carrying equal marks), but increasingly there may be a compulsory section, then a section providing choices. | **Validity:** poor, limited to what comes out of pens.  **Fairness:** can be good, but poor when answers are essay-type, and different markers would award very different marks for the same essay.  **Whodunit?:** relatively safe (though stories of ingenious cheating are legion!).  **Real world:** written exams are not at all close to the workings of the real world; most people never do a written exam again after leaving university.  **Feedback to learners:** very limited indeed, usually just a score/grade or pass/fail, which can leave candidates having very little idea about what they did well or badly. | Can avoid plagiarism and cheating.  Give data which can be ranked and handled quantitatively.  Exams are relatively familiar to learners entering higher education, as they’ve already experienced them at school.  Exams are already ‘hard-wired’ into many university systems, so a case doesn’t have to be made for continuing to use them.  Written exams are much better for some subjects than others: for example they can work well for mathematical and quantitative subject matter, and tend to work really badly for ‘wordy’ or descriptive matter. | As has been argued already in this book, exams tend only to measure what comes out of pens, a poor proxy for what might be in heads.  Many otherwise capable learners never show their best efforts under exam conditions.  It can take a long time to mark a set of exam scripts (properly). (There are economies of scale for large numbers of candidates as examiners become familiar with the marking scheme being used).  Problems with speed of writing and legibility, and difficulties candidates face when using a se.  ‘Sudden death’: a bad day can mar a lifetime.  A snapshot of achievement, rather than a real measure of it.  One of the main skills measured tends to be time-management – dividing the available time sensibly between the questions being attempted.  Promotes surface-learning: filling heads with information to use ‘on the day’ and forget as quickly as possible thereafter.  Question-spotting by candidates can pay off substantially, meaning that at least some candidates pass without having learned the whole syllabus reasonably well.  Where candidates have a choice of questions, it is really hard to get all of the questions to be of equal difficulty – leading to at least some candidates ending up with an easier exam than others overall.  Long-answer written exams contribute to the continuing trend for post-compulsory education to remain ‘elitist’ – i.e. to favour those who are good at such exams. |
| **2 Short-answer exams**  Short written responses to a large number of questions.  Usually the whole exam is compulsory, reducing the tendency for candidates to us question-spotting as a means of deliberately only learning some of the syllabus. | **Validity:** often a lot better than long-answer exams, as evidence of achievement of learning outcomes can be covered much more fully, rather than the ‘write down everything you happen to know’ tendency which long-answer questions can engender.  **Fairness:** can be good.  **Whodunit?:** safe under exam conditions.  **Real world:** closer than traditional exams, as success may involve knowing the overall subject really well.  **Feedback to learners:** poor, usually just a score or grade. | Can cover a wide range of topics in a limited time.  Not so much affected by speed of writing or legibility.  Can be somewhat faster to mark than long-answer exams (but not always).  Can measure breadth of knowledge.  Is fairer than exams where there are choices of question, as all candidates are effectively taking exactly the same exam. | Can miss out on depth of knowledge.  Can deprive high-fliers of the opportunity to excel.  It can take much longer to design a short-answer exam paper than a traditional long-answer one.  It can be quite difficult to apportion marks across the various elements in a short-answer exam paper. |
| **3 Multiple-choice exams**  Paper-based or computer-based: e.g. select the ‘best’ option from 4 or 5 alternatives for each of a fairly long list of questions.  Such exams are usually time-constrained, but often most candidates complete the exam with plenty of time to spare.  Note that these can be extended to be **multiple-response exams,** where each question is along the lines of ‘Which (one or more) of the following options is true?’. This can cause candidates to think harder about *all* of the options, rather than just picking the ‘best’ then moving on. | **Validity:** can be good if questions well-designed. A wide range of syllabus knowledge can be addressed in quite a short time.  **Fairness:** can be really good, if questions well designed and trialled.  **Whodunit?:** safe under exam conditions; not safe if asynchronous, for example in distance learning contexts.  **Real world:** can be better than other kinds of exams, as multiple-choice exams tend to measure what goes on in heads, and aren’t limited by what comes out of pens – they can measure decision making well.  **Feedback to learners**: possible to be excellent in speed and quality (but not often achieved). Excellent feedback can address, for each option: ‘was I right?’ and (particularly) ‘if not, *why* not?’ | Gets away from ‘what comes out of pens’ limitations of other kinds of exams.  When questions well-designed, can quickly test quite a wide range of subject knowledge.  Can be useful in areas where rapid decision-making is a useful skill for learners.  Possible to provide feedback on-screen after each decision in computer-based uses, or a complete feedback printout (feedback on the distractors as well as the correct choices) on leaving the exam. (Such printouts can make really useful revision tools for future candidates)  Can be useful for candidates who have difficulty stringing together fluent prose in written answers, but who can still think clearly through options. | It is much harder than people think to design really good multiple-choice questions.  The ‘key’ is the best option, and is the one intended to be chosen by candidates who know the topic properly. However, It is sometimes hard to design a ‘key’ which is *always* right – high-fliers can often spot when even this choice is not correct.  It can be difficult to design good ‘distractors’.  Options such as ‘all of the above’ or ‘none of the above’ are still too often included (usually though laziness regarding thinking of more distractors), and which are very rarely the ‘best’ option in any case.  Questions need to be well piloted and tested before being used in exams.  Still an element of ‘luck’ picking the best option (in a test using four-option questions, the average monkey should score 25%).  Quite a lot of emphasis now goes onto *reading* the questions and options well; learners with limited skills or speed in reading can be disadvantaged. |
| **4 Essays**  Essays are highly regarded as an assessment type, despite the many disadvantages listed in this table!  At university level, a coursework essay is often an extended account of length such as 3000 words, usually submitted in word-processed form, and often passed through plagiarism-detection software.  However, handwritten essays (shorter) in exams are still widely used.  In the case of coursework essays, learners may be given an essay title or theme, or may be allowed to choose from a list of topics, or may have leeway to choose an appropriate topic of their own, or negotiate a topic with a tutor. | **Validity:** rarely good (prose which comes out of a pen in an exam, or through a keyboard in coursework, is rarely the best way of measuring evidence of achievement of intended outcomes).  **Fairness:** poor (a great deal of research evidence is available showing that different markers award very different marks).  **Whodunit?**: very unsafe, except for essays under exam conditions. Essays can be commissioned and purchased online from well-practised skilled writers! However, concerns about Whodunit? are minimal with small groups of learners, where tutors regularly talk to learners and would usually quickly know if submitted work was not their own.  **Real world:** not close to the sorts of writing relevant to most careers.  **Feedback to learners:** can be useful, but usually too late, not least because of the length of time it takes to mark a set of essays. There is usually no feedback at all on essays handwritten in exams. | Allow learners to demonstrate ability to construct written arguments, and to write fluently.  Can give candidates who show ‘depth’ fair reward.  Can give an indication of the quality, depth and breadth of reading that has been done by candidates. | Essays take forever to mark, and marking is unreliable (unfair) anyway, as proved by a great deal of research!  Tends to advantage learners who are good at written ‘waffling’!  Unless there are tight word-limits, a longer essay will usually score higher than a shorter one.  Where there is an element of choice (e.g. coursework essays) some choices may prove harder to bring off in practice than others, disadvantaging some learners.  In coursework essays, there can be a tendency to copy in (suitably rephrased) sentences from literature sources, without really thinking about the meaning of the elements copied in.  Spelling, punctuation and grammar may disproportionately affect marking.  ‘Coherence’, flow, ease of reading essays disproportionately influences most markers. A ‘smooth’ essay is usually awarded higher marks than a ‘jerky’ one, even if the content of the latter is much better.  Where essays are handwritten in exams, it is not at all easy to edit and adjust along the way, e.g. to go back and rephrase the start of the essay appropriately after the main thrust has been addressed.  Handwritten essays in exams are subject to concerns about measuring ‘what comes out of a pen’ rather than ‘what’s in a head’, and are subject to the effects of speed of writing, legibility and so on.  Handwriting an essay in an exam is quite a different game than composing a word-processed coursework essay, so coursework is poor preparation for the exam experience, and feedback on coursework essays may not help exam candidates. |
| **5 Annotated bibliographies**  This kind of assessment works best when various elements of the task are specified, including one or more of:   * Overall word count, preferably with quite tight limits; * Number of sources to include – preferably an exact number in practice; * Any expected balance between kinds of sources, e.g. journal articles, reviews, book chapters, web sources. * Whether to include (say) four ‘given’ sources, and four more that individual learners have found; * Whether to prioritise the sources in order of any particular aspect: e.g. usefulness, authority, relevance to topic, and so on; * Whether the list should be a comparative one, e.g. with pros and cons of each source in the context of the bibliography; * Whether there is free choice regarding the ‘age’ of the sources. For example, in some contexts it can be useful to ask limit contents to sources published within the last 3 years. * The extent to which the comments on each item should include the learner’s own view. | **Validity:** can link well to learning outcomes about breadth of reading, and prioritising quality of sources.  **Fairness:** can be much better than essays or reports, as high-flying candidates quickly distinguish themselves by the quality of their comments on sources, and their rationale for the sources they choose.  **Whodunit?:** good, as any unwanted collaboration or copying would be fairly obvious (e.g. same sources selected, in same order, and with identical mistakes in the referencing!).  **Real world:** high relevance to many careers, where learners will need to be able to review a range of sources and select and justify those most relevant to a given context.  **Feedback to learners:** this can be harder to achieve than with some other assessment types, but can be really valuable when done well (e.g. it would have been good for you to have included source ‘x’ *because*...., source ‘y’ was not a good one to include, *because*...., and so on). | A really useful way to cause learners to read around a topic, rather than just dip into random sources during their studies.  ‘Wikipedia’ (for example) could be allowed to be linked to only one of several required sources.  Learners can be given the opportunity to demonstrate the breadth of their learning (range of sources) as well as depth (their judgemental comments about respective sources).  It can be useful if one of the assessment criteria relates directly to the correct ‘citing’ of each source (e.g. ‘Harvard system’), so that learners gain practice in getting references exactly right, which may be important for future research-related writing.  A set of annotated bibliographies can be retained as an online resource, to show future cohorts of learners how to do this task, and for these learners to practise making judgements to establish criteria for their own work.  A ‘300-word annotated bibliography prioritising five sources’ can be far faster to mark than would be an essay where similar literature reviewing was intended – and can be marked much more reliably (fairly) than such an essay. | The easiest sources to find tend to be via Google (especially the first page in a search) and Wikipedia, and it may be necessary to prevent learners from making more than limited use of these sources.  Annotated bibliographies essentially require learners to demonstrate their academic literacies, but any lack of relevant information literacies may get in the way of this.  There is the danger that the extent of sources chosen by learners may mask the depth of thinking about individual sources, therefore it is important not just to get learners listing a lot of sources.  Learners may need some rehearsal before undertaking this sort of task for assessment. For example, a whole-class session could be taken up with learners assessing some past examples of annotated bibliographies, so that they found out how the assessment worked in practice before making their own contributions to the genre. |
| **6 Reports**  For example, write-ups of practical work, field work, investigations, and so on.  Usually word-processed these days. | **Validity:** can be reasonably high.  **Fairness:** not great, but much better than essays. Can allow room for good candidates to shine.  **Whodunit?:** can be unsafe, unless other ways of checking, e.g. face-to-face quizzing.  **Real world:** can be good: report writing relevant to many careers.  **Feedback to learners:** can be useful, but usually comes too late. | Avoids ‘sudden death’ aspects of assessment, as reports are usually built up over a period of time.  Assessment can be broken down usefully, such as agreed proportions of marks for ‘Abstract’, ‘Method’, ‘Interpretation of Data’, ‘Conclusions’, ‘Plans for further work’ and so on.  Can be on work done collaboratively, but with individual write-up.  Learners who write-up quickly (before they have forgotten what they actually did) can be advantaged, therefore encouraging good study habits of ‘keeping up’ and ‘avoiding backlogs’. | Word-limit may need to be controlled strictly, as long-reports would otherwise almost always score more marks than short ones, whereas in the real world a really good short report may be much more useful in practice.  There is the danger that learners can spend more time on writing reports, than is reflected by the marks they carry overall in the bigger picture of assessment.  Those learners who end up with a backlog of reports may spend far too much time catching up on this backlog at the expense of preparing for summative exams, which may carry much more weight in the overall assessment. |
| **7 Portfolios of evidence**  For example, built up over a period of time on a course or module, often with intermediate feedback opportunities for learners.  In the present climate of ‘evidence-based practice’, assessment by portfolios is naturally gaining momentum. As can be seen in this table, the ‘status’ aspects (validty, etc.) are favourable, and there are many advantages, but the crunch comes in the disadvantages column – portfolios take a great deal of time to mark, and there are distinct problems regarding fairness of assessment. | **Validity:** can be good, as different elements of a portfolio can relate to each different aspect of evidence of achievement of a range of learning outcomes.  **Fairness:** can be good, but different assessors may be looking for different things in a portfolio, in which case fairness can be poorer.  **Whodunit?:** can be questionable, but improved where face-to-face probing is also used.  **Real world:** can be better than many other kinds of assessment, depending on what’s included in the portfolio specification.  **Feedback to learners:** can be good, especially if their progress is reviewed at various times during construction of a portfolio. | Allows for a wide range of kinds of evidence of achievement, for example drawings, photos, videos, recordings, reviews, reflective commentaries.  Can extend well beyond the ‘read-write’ domain.  Can build in opportunities for learners to reflect on their learning, and provide evidence of such reflection.  Can be interdisciplinary, helping learners to link together aspects of different subjects and topics.  Allows candidates to demonstrate originality and creativity.  Portfolios can be useful evidence to show prospective employers.  Portfolios can be maintained and updated beyond the assessment period. | It can take forever to mark a set of portfolios, and they can be very bulky to carry around from one marking place to another (work to home, and so on).  Can be difficult to balance marks for portfolios evidencing different strengths.  A big portfolio will normally attract higher marks than a small one, so there is a tendency to reward ‘cramming in as much as possible’ rather than quality of evidence.  One of the most significant dangers with such a ‘big’ assessment element is non-completion.  When portfolio assessment is used, the portfolio may often make up a substantial part of the overall assessment, which can mean for those candidates who are not at their best in this element, assessment is unduly prejudiced against them. |
| **8 Oral exams**  These are essentially about interrogating individual learners face-to-face, often with two or three assessors present. Sometimes the questions may be about the course or module in general, and/or about a particular submitted element of coursework (for example a portfolio or dissertation).  An oral exam does not necessarily have to be lengthy. Even a 5-minute oral exam can be useful to gain assurance about the ‘whodunit?’ aspect of a portfolio or dissertation.  (Note that in many parts of the world there is much more oral assessment than written assessment, as was the case in the UK before written exams started in 1791 towards becoming endemic! | **Validity** can be high, but still can favour candidates who can ‘talk well’ over those who know it just as well but are not so good at ‘talking it’.  **Fairness:** can be good when successive candidates are asked exactly the same questions (but ‘security’ of questions needs to be safeguarded, so they don’t ‘leak’ to later candidates).  **Whodunit?:** one of the safest kinds of assessment.  **Real world:** strong links to the sorts of questioning learners will need to be able to handle in most careers.  **Feedback to learners:** can be quick and useful, but also can be stressful and rather transient (e.g. quickly forgotten). | Allows for probing questions, to test real mastery rather than surface learning.  Gets over reservations about ‘Whodunit?’!  Learning payoff during preparation for an oral exam can be wider, deeper and better than (for example) just writing an essay or report, as anything may be asked.  When learners are encouraged to engage in practice and rehearsal, they can learn a great deal from each other, as well as improving their skills relating to oral performance. | Some candidates can be let down by nerves.  With a large class of learners, a round of ‘orals’ can take a lot of time, and ‘the word inevitably gets around’ about the sort of experience it is going to be.  Evidence of achievement as demonstrated orally is somewhat ephemeral, and it can be difficult to think back over several candidates performance and remember accurately who did better or worse.  It can be difficult to guarantee fairness, when different degrees of probing may have been used with different candidates. |
| **9 Individual presentations**  Typically, assessed presentations are given before a relatively small group rather than a whole large class, in the presence of a tutor (who assesses) but often some peer-assessment is built in as well.  Sometimes presentations are made by groups, but then it is much harder to allocate credit appropriately to individuals, so for assessed presentations, individual performance is preferable.  It is useful if the briefing is really clear to learners in advance, including:   * The duration * A range of topics to choose from – or the chance to choose something original * Any guidance about supporting materials (e.g. slides, handouts, exhibits). * The assessment criteria * The extent of any ‘question and answer’ episode after each presentation. | **Validity:** can be good when learning outcomes include oral communication skills.  **Fairness:** it can be difficult to maintain fairness during a set of presentations, as there is a tendency for later candidates to learn from earlier ones, and do better.  **Whodunit?**: one of the safest forms of assessment.  **Real world:** high relevance; in many careers candidates will need the skills involved in giving presentations.  **Feedback to learners:** possible to give quick and useful feedback, but his may be somewhat ephemeral and quickly forgotten. Peer-feedback during rehearsal can however be really valuable. | Allows oral communication skills to be demonstrated alongside mastery of subject matter.  Can allow candidates as much time as they need for preparation and rehearsal.  Depth of learning tends to be high; learners tend to remember very well things they researched and practised as preparation for a presentation.  Can be used in a peer-assessment context, there learners can gain a lot from making informed judgements on each others’ presentations.  Can allow individual learners to demonstrate particular strengths.  Can include the opportunity for ‘probing’ to test depth of knowledge, where questions are posed for a few minutes after each presentation. | It can take forever to assess a large number of presentations.  Choice of topic can affect marks significantly. What seemed like an interesting and stimulating topic can end up being harder than imagined, and so on.  There can be some drift in standards, where ‘later’ candidates are judged more rigorously than ‘earlier’ ones, or conversely benefit themselves from seeing earlier presentations.  Some candidates can be unduly disadvantaged by nerves.  Impression marks associated with the quality of slides or handout materials used during the presentations may overshadow the quality of the actual mastery of the topic concerned.  It is sometimes difficult to collect evidence to put forward for moderation (for example external examiner scrutiny), though recordings can be made for this purpose. |
| **10 Posters**  For example, preparation of a visual display in a specified format, on (e.g.) an A1 sheet, using photos, drawings and text to address a particular brief. | **Validity:** can often be high, allowing good links between evidence of achievement and intended learning outcomes.  **Fairness:** likely to be at least some subjectivity when it comes to judgements, but this can be offset by having multiple judgements (possibly peer-assessment, and assessment by externals).  **Whodunit?:** questionable, as learners may use varying amounts of external help in producing posters, and may collaborate with each other. The ‘Whodunit?’ aspect can be made much better when the assessment also included learners ‘talking about their poster’ or ‘being questioned about their poster’.  **Real world:** visual displays are used in many professions, for example to back-up a proposal, or present findings to colleagues.  **Feedback to learners:** can be very effective, especially when tutor and peer-feedback is offered on drafts in class (i.e. nothing ‘secret’) before the preparation of the final submission. | Can allow learners flexibility in choice, where they have significant control of the topic and the way they present their findings.  Gives room for learners to organise their thinking visually rather than in words alone.  Exhibitions of posters can be kept online, and used for future learners as indications of the kinds of evidence they may aim to emulate, and (better) for getting future learners to learn by assessing, before they set out to make their own posters. | It can be really hard to make relative assessment judgements about different topics handled in different ways to different depths.  Wealth may come into the picture; learners who can afford good colour printing and photos may be advantaged over those whose resources are more limited.  The visual aspects of the poster can dominate too much when being assessed.  Judgements on visual evidence such as in posters is always to some extent subjective, with different assessors looking for different things in a ‘good’ poster. (This problem can be offset by having several assessors, or including one or more externals in the assessment). |
| **11 Artefacts**  These can include paintings, designs, models, sculptures, items of metalwork, engineering outputs, teaching materials, plans, accounts, prototypes, furniture, display items and so on. | **Validity**: this can be high, where the intended outcome of a particular curriculum element includes the production of specific items.  **Fairness:** this can be harder to achieve in assessment of artefacts, as originality and creativity are likely to be among the assessment criteria, and therefore there is bound to be at least some subjectivity in the assessment.  **Whodunit?** On this dimension, artefacts are relatively safe, although there remain possibilities for others to have helped in their production.  **Real world:** artefacts often link strongly to the sorts of skills which learners need for particular vocations or careers.  **Feedback to learners:** this can be done well, for example by ‘things I like best about this example’ and ‘one suggestion which could have improved this example would have been.....’. | When a course or module involves learners in practical work in workshops or studios, it makes a big difference if the quality of their work there counts towards their overall assessment.  The competitiveness which is encouraged by some sort of measure of the quality of the things learners make encourages them to put more effort into their practical work.  Artefacts can be retained by learners after assessment, and can be useful evidence of their achievement to show prospective employers.  A photographic record of assessed artefacts (or the artefacts themselves if not needed by their creators) can provide the next cohort of learners with valuable targets to aim towards – and exceed. | Where some learners may have benefited from external help in the production of artefacts, the fairness of assessment can be compromised.  When individuals are to be assessed on the basis of artefacts they produce, collaboration between learners is discouraged, and may deprive them of things they could have learned from and with each other.  It can sometimes be difficult to work out how much the assessment of artefacts should contribute to the overall assessment of a particular curriculum element.  Where some learners have special needs which limit how well they can produce particular kinds of artefact, it can be difficult to make ‘reasonable adjustments’ to allow them alternative assessment possibilities. |

**Extracts from ‘The Lecturer’s Toolkit: 4th edition (2015, Routledge)**

**(From Chapter 2 ‘Designing assessment and feedback to enhance learning’)**

In the ‘Toolkit’, Chapter 2 follows straight on from an exposition of the ‘Ripples’ model of learning, based on the answers I’ve had from over 200,000 people to the questions about their learning experience. There’s quite a lot more ‘theory’ and quotes, but I’ve included here the main ideas which link to the design of exams and essays. The chapter in the published edition goes on to give fairly detailed suggestions about how we can approach several other assessment formats and processes. The second part of the chapter is a detailed discussion of a variety of feedback processes.

****

**Pre-quotes**

**On assessment:**

Assessment of student learning is a fundamental function of higher education. It is the means by which we assure and express academic standards and has a vital impact on student behaviour, staff time, university reputations, league tables and, most of all, students’ future lives. The National Student Survey (in the UK), despite its limitations, has made more visible what researchers in the field have known for many years: assessment in our universities is far from perfect. (HEA, 2012 p.7)

Assessment is a central feature of teaching and the curriculum. It powerfully frames how students learn and what students achieve. It is one of the most significant influences on students’ experience of higher education and all that they gain from it. The reason for an explicit focus on improving assessment practice is the huge impact it has on the quality of learning. (Boud and Associates, 2010, p.1)

**And on feedback:**

Higher education teachers are often frustrated by the modest impact feedback has in improving learning. The status of feedback deserves to be challenged on the grounds that it is essentially about telling. For students to become self-sustaining producers of high quality intellectual and professional goods, they must be equipped to take control of their own learning and performance. How can students become better at monitoring the emerging quality of their work during actual production? Opening up the assessment agenda and liberating the making of judgments from the strictures of preset criteria provide better prospects for developing mature independence in learning. (Sadler, 2013b)

This is the longest chapter in the *Toolkit.* This is partly because most lecturers actually spend far more time marking students work and designing feedback for them, than they spend on preparing and giving lectures. But it’s also because there is nothing more important than assessment – if we get that wrong we damage students’ futures. And getting assessment and feedback right are the hardest things we do. The chapter is therefore divided into four main parts as follows:

* **Putting assessment and feedback into perspective:** exploring the implications of some of what we can learn from the huge literature of evidence-based practice around assessment;
* **Pros and cons of sixteen assessment processes:** a detailed comparison of a range of different ways to go about assessing students’ evidence of achievement of intended learning outcomes, including a detailed and critical probe into essays and exams;
* **Making formative feedback work:** a discussion of various ways of going about helping students to benefit from feedback on their assessed work
* **Involving students in their own assessment:** rounding off much of what has been discussed in this chapter and the previous one, with some suggestions for getting students to deepen their learning by being involved in their own and each others’ assessment, using self- and peer-assessment.

**Putting assessment and feedback into perspective**

This chapter addresses the most important and intimate interactions of lecturers and students: assessment and feedback. Often, in fact, lecturers spend more time and energy on assessing students’ work and providing feedback for students, than on preparing and delivering teaching. I am therefore starting the chapter with some detailed references to some of the most important recent publications about assessment and feedback, before we begin a point-by-point exploration of what we can do to live up to some of the recommendations which have been made about radically overhauling assessment in higher education.

Whether we think of ourselves as lecturers, or teachers, or facilitators of learning, the most important thing we do for our students is to assess their work. This is why, in this book, I have gone straight into assessment after thinking about learning. It is in the final analysis the assessment we do that determines their diplomas, degrees, and future careers. One of the most significant problems with assessment is that just about all the people who do it have already survived having it done to them. This can make us somewhat resistant to confronting whether it was, when we experienced it at the receiving end, valid, fair, authentic and transparent, and explains why so many outdated forms of assessment still permeate higher education practice today.

Over the last decade, many of us have seen our assessment workload grow dramatically, as we work with increasing numbers of students, who are ever more diverse. Consequently, the time we have available to devote to assessing the evidence of achievement arising from each student has fallen. Even those methods and approaches which used to work satisfactorily with relatively small numbers of students are now labouring as we try to extend them to a mass higher education context. It is therefore more important than ever to review the way we design and implement our assessment.

Brown and Glasner began the conclusion of their edited collection *Assessment Matters in* *Higher Education* with the words:

Assessment does matter. It matters to students whose awards are defined by the outcomes of the assessment process; it matters to those who employ the graduates of degree and diploma programmes; and it matters to those who do assessing. Ensuring that assessment is fair, accurate and comprehensive – and yet manageable for those doing it – is a major challenge. It is a challenge which has been grappled with by many, ... Despite the fact that there is a considerable body of international research about assessment and related issues, we experiment largely in ignorance of the way others have effected positive change, and we have limited opportunity to learn from the lessons of others. (Brown and Glasner 1999)

Their book makes good place from which to work backwards through the literature on innovative assessment during the last decade of the twentieth century, and more recently Knight and Yorke (2003) explore in depth some of the things that are still going wrong in assessment at the opening of the present century, and the collection edited by Peelo and Wareham (2002) confronts both the experiences of students who fail, and the ways in which assessment in higher education can be regarded as failing students.

But even now, we have not learned the lessons which are presently available from a very substantial literature on how best to design assessment and feedback. In ‘A marked improvement’ published in 2012 by the Higher Education Academy in the UK, collecting together the wisdom and experience of a distinguished group of experts on assessment, the following warning is sounded:

Assessment practices in most universities have not kept pace with the vast changes in the context, aims and structure of higher education. They can no longer do justice to the outcomes we expect from a university education in relation to wide-ranging knowledge, skills and employability. In a massified higher education sector where tutor-student ratios have gradually been eroded, students can remain confused about what is expected of them in assessment. Efforts to make this transparent through learning outcomes, assessment criteria and written feedback have proved no substitute for tutor-student interaction and newer groups of students are particularly likely to need this contact. (HEA, 2012, p.7)

My overall aim in this chapter is to challenge your thinking on how best to assess students’ learning, and how to optimise the impact of our feedback on students’ learning – whether that learning has proved successful or not. I hope too to provide food for thought to enable you to confront the difficulties in order to move towards making assessment demonstrably fair, valid, authentic and reliable. As a prelude to this chapter, I would like to share some overarching thoughts and questions about teaching, learning and assessment, and the relationships between these processes. Then I will outline some ‘concerns’ about unseen written examinations, and about continuous assessment. The remainder of this chapter is intended to offer some thoughts about sixteen particular forms of assessment, each with its pros and cons, and with some suggestions for making each work better, to improve student learning. (extracts from the first five of these are presented here).

**What are the main purposes of assessment?**

The UK Quality Assurance Agency for Higher Education preface the chapter on ‘Assessment of students and the recognition of prior learning’ of the ‘UK Quality Code for Higher Education’ with this explanation:

Assessment is a complex topic since it involves two distinct aspects. First, it forms an essential element of the learning process. Students learn both from assessment activities and from their interaction with staff about their performance in those activities. This interaction has two elements: a focus on their learning and the extent to which that has been demonstrated in the assessment, and a focus on furthering their learning, which may itself subsequently be assessed. The latter element is often referred to as 'feedforward'.

Second, it is the means by which academic staff form judgements as to what extent students have achieved the intended learning outcomes of a programme, or of an element of a programme. These judgements form the basis for the grading of student performance through the allocation of marks, grades and (where applicable) classification, and (provided the learning outcomes have been met) for the award of the credit or qualification to which the programme leads. (QAA, 2013, p.3).

This neatly shows that assessment and feedback need to be driving learning in appropriate directions, but assessment also needs to be capable of being the basis for the achievement of standards and the award of qualifications. Ideally, these two purposes need to be served at the same time, but some of the problems experienced with assessment are when the two purposes are not addressed in harmony. Boud *et al* (2010) go further in their influential publication ‘Assessment 2020: Seven propositions for assessment reform in higher education’ by proposing that students themselves should be involved in assessment much more intimately and productively than has been done in the past, and that assessment is not only the business of academics. They propose that getting students involved in assessing is a prerequisite for enabling them to produce work of optimum quality as follows:

Assessment is the making of judgements about how students’ work meets appropriate standards. Teachers, markers and examiners have traditionally been charged with that responsibility. However, students themselves need to develop the capacity to make judgements about both their own work and that of others in order to become effective continuing learners and practitioners.

Assessment plays a key role in both fostering learning and the certification of students. However, unless it first satisfies the *educational* purpose of ensuring students can identify high quality work and can relate this knowledge to their own work, the likelihood that they will reach high standards themselves is much reduced. (Boud *et al,* p.1)

Throughout this Chapter, I suggest ways of helping students get their heads around how assessment works, not only so that they can optimize their performance in various assessment contexts, but also so that they can deepen their learning while in education, and far beyond.

**The language of assessment**

In any discussions of assessment, most of the following words appear: validity, reliability, authenticity, transparency and inclusiveness. It’s worth explaining briefly a few of the key terms relating to assessment at this point.

* **Validity** is about whether the particular assessment format under consideration is the most appropriate for the intended purpose. In other words, is it the best way of measuring evidence of achievement of the related intended learning outcomes?
* **Reliability** is about how well different assessors would agree on the mark or grade awarded for a particular piece of students’ work. This is also, of course, about **fairness**and indeed **justice** as perceived by students and others.
* **Authenticity** is about how well the assessment correlates to the sorts of things students need to be able to do in their career after leaving the educational institution. It’s about the **real-world**relevance of the assessment activity.
* **Transparency** is particularly about how well students can see how the assessment works in practice, and how marking occurs. Professional bodies also are keen to see this.
* **Inclusiveness** is about how well the assessment can be taken by a range of candidates with additional learning needs, including dyslexia, dyspraxia and so on. It is about minimising unfair discrimination towards candidates with particular needs.

Alongside these factors, it is important to consider how **manageable** each particular type of assessment proves in practice – in other words how much time it takes for marking and moderation to be done, and indeed how much time and energy it needs from students themselves.

And finally, there’s the issue sometimes referred to as **veracity** – in other words, who did it? (In workshops, I can’t resist calling this the ‘whodunit?’ factor!). This is about the extent to which it can be guaranteed that no plagiarism, copying or cheating has occurred. In exams, for example, this issue can (normally) be regarded as ‘safe’, but in word-processed assessed coursework, reports, dissertations and thesis it can be ‘unsafe’!

It can be argued that the ready availability of source materials on all subjects imaginable has changed how we regard reference materials. I do not agree that the learning process itself has been ‘radically reshaped’ however, but agree that how people view other people’s work has altered probably irreversibly.

Returning to the wider picture of assessment, in this chapter, I offer various practical suggestions regarding how assessment can be improved, particularly so that assessment can be:

* more *valid,* measuring that which we really intend to measure, rather than ‘ghosts’ of students’ real learning;
* more *reliable* and *consistent,* moving away from the subjectivity that can cause assessment to be unfair;
* more *transparent,* so that students know where the goalposts are, and so that external reviewers can see clear links between intended learning outcomes as spelled out in course documentation, and assessment criteria applied to students’ work;
* more *authentic,* bridging the gap between what students do in assessed contexts at university, and what they need to do in the world outside when in employment;
* more *diverse,* so that individual students are not disadvantaged unduly by particular forms of assessment;
* more *manageable,* both for our students and for ourselves;
* more useful in terms of *feedback,* so that students’ learning is enhanced;
* more successful in promoting *deep* learning, so that students get a firmer grasp of the important theories and concepts underpinning their learning.

Radically overhauling and re-designing assessment and feedback in higher education is a daunting task, but HEA (2012) express desirable benefits which may be achieved thereby, not least by increasing the proportion of formative assessment and feedback:

...Where programmes plan for more formative assessment and feedback, there is a better chance that a greater proportion of students pass modules at their first attempt, thereby saving staff time in relation to demand for extra support, resits, appeals and complaints. Improved pass rates and reduced attrition bring obvious financial benefits for institutions and positive outcomes for students. Overall, a radical review of assessment can bring cost savings and better use of teaching resources. (HEA p.11).

Having ‘scoped the task’ of overhauling the design of assessment in higher education, let us now proceed to a step by step analysis of some of the things we need to address, starting with establishing a sound rationale for assessment.

**Concerns about assessment: we can’t go on like this!**

Boud *et al* (2010) are in no doubt: the status quo in assessment is not an option. In their words:

Universities face substantial change in a rapidly evolving global context. The challenges of meeting new expectations about academic standards in the next decade and beyond mean that assessment will need to be rethought and renewed. (Boud *et al,* p.1).

Before it is possible to persuade people to review what they are presently doing, and to consider implementing changes, it is useful to take a critical look at whether current practices actually work as well as we think they do. Therefore I continue this chapter with a critical review of the two principal areas of assessment which most students encounter: traditional time-constrained, unseen written exams, and assessed coursework. In each case I will list some general concerns, starting with concerns about the links between these kinds of assessment and the factors underpinning successful learning drawn from Chapter 1 of this book: wanting to learn, needing to learn, learning by doing, learning through feedback and making sense of what has been learned, and also the need for students themselves to gain experience in *verbalizing* and *assessing* their own learning before it is formally assessed by us. For most of the concerns, I will add hints at how the repercussions they cause be ameliorated – or at least confronted.

Finally in this section on ‘concerns’ I move on briefly to the increasingly prevalent climate of standards and benchmarks, and quote Sadler’s 2014 paper positing whether or not attempts to prescribe or define standards are doomed to be futile! However, despite all these concerns and doubts, assessment needs to go on, so later in the chapter I offer a range of practical pointers suggesting how even the most traditional methods of assessment can be put to good use, as well as exploring the pros and cons of a variety of alternative ways of measuring students’ evidence of achievement.

1. ***Concerns about traditional exams***

Much has been written about the weaknesses of traditional examinations – in particular time-constrained unseen written exams. In many subject disciplines, this assessment format seems to be at odds with the most important factors underpinning successful learning. Moreover, there is abundant evidence that even in discipline areas where the subject matter is well defined, and answers to exam questions are either correct or incorrect, assessors still struggle sometimes to make exams valid, reliable, authentic or transparent to students. In disciplines where the subject matter is more discursive, and flexibility exists in how particular questions can be answered well, it can be even harder to achieve demonstrable reliability (and that means fairness) in assessment, even when validity is well achieved.

Overall in higher education at present, with staff time under more pressure than ever before, there is evidence of a drift back to reliance on exams, which have been argued to be one of the more time-efficient and cost-effective methods of assessment, where it is fairly easy to achieve fairness and reliability, and with the added bonus that plagiarism or cheating cause less headaches to markers than in many other forms of assessment. However, when one takes into account the time it takes to design, set, mark, moderate, and process the results of exams, I remain sceptical of the time-efficiency of traditional exams compared to many alternative forms of assessment.

Some of the principal concerns that can be expressed about unseen written exams are summarised below. Half of the fourteen concerns below are conflicts between this kind of exam and the seven factors underpinning successful learning which we considered in Chapter 1. It could be said that unseen written exams are diametrically opposed to these factors working successfully. Several of the remaining concerns relate to our own procedures for marking students’ exam scripts.

1. **Exams don’t do much to increase students’ ‘want’ to learn.** Students often make choices in modular schemes strategically, so that they avoid this kind of assessment if they can. This can lead them to choose subjects in which they are less interested than those which they fear to select because they will be subjected to exams. Ask students for words that come to mind at the mention of exams. Words include ‘dread’, ‘looming’, ‘fear’, ‘scared’ and ‘threat’ far more often than ‘challenge’ or ‘pleasure’.
2. **Exams are not often a good way of alerting students to what they really need to learn.** Admittedly, students will often only get down to serious learning when an impending exam causes them to revise actively, but the fact that in unseen exams the actual assessment agenda has to be guessed at rather than worked towards systematically means that the resultant learning can be unfocused, and the assessment judgement becomes too dependent upon the success of the agenda-guessing.
3. **Exams are not ideal occasions for learning by doing.** Though students may do a lot of learning *before* formal unseen written examinations, their actual experiences of learning *in* such situations is extremely limited. In other words, a note could be placed on the door of the exam room stating ‘exam cancelled; you’ve already done all the learning that this exam could have caused’! The learning payoff during an assessment element should be considered more – assessment *as* learning rather than mere assessment *of* learning. It is therefore worth our while revisiting our testing processes to search for forms of assessment which are in themselves better learning experiences.
4. **The amount of feedback that students receive about exams is far from optimal.** Most systems require marked exam scripts to be regarded as secret documents, not to be shown to students on any account! It is worth asking what reasons underlie this philosophy? You might have noticed that among the student demands in their charter in 2010 was a request for feedback on exams. It is useful to reconsider the value that students can derive from seeing their marked examinations papers, where it should be possible to be able to demonstrate to students that the examination marking has indeed been reliable, fair, and valid. Moreover, the natural process of learning from mistakes should always be accommodated, even when the assessment judgements have already been taken down to be used in evidence against the candidates.
5. **Exams tend not to do much to help students make sense of what they have learned.** While there may be a significant amount of making sense of concepts and theories during the time leading up to exams, the assessment experience itself does little to help students to gain any further deepening of their grasp of these. One of the consequences of modularising the curriculum can be that some subject matter is introduced too close to an impending exam for the content to be satisfactorily digested.
6. **Written exams don’t do much for learning through *verbalizing.*** However, practising putting learning into spoken words pays enormous dividends even when the eventual testing is written. We need therefore to encourage students not to study for exams in solitary silence, but to work for at least some of the time explaining answers to likely answers to each other (or anyone else who will listen).
7. **Written exams don’t encourage students to learn by *assessing.*** The tendency is for students to do what they can and hope for the best. However, the more we can encourage students to practise making judgements on their answers to typical exam questions, the better they can make similar judgements while they are answering questions in exams. The more they know about how marks are gained and lost, the better they can structure their efforts towards doing written exams successfully.
8. **We mark exam scripts in a rush.** Most staff who mark exams agree that the task usually has to be completed in haste, in preparation for timetabled exam boards. The situation has been worsened by modularization and semesterisation developments in most institutions, which give tighter turn-round intervals between examinations and progression to the next element of study. While our marking may still be relatively fair and reliable (at best), it can be shocking to students who have spent a great deal of time preparing for unseen written exams to find out that their scripts are marked so hastily.
9. **Unseen written exams can lead to us placing too much emphasis on unimportant factors in candidates’ answers.** For example, factors such as quality of handwriting, or neatness of overall presentation of scripts can influence examiners, consciously or subconsciously. Many students nowadays are much more comfortable composing essays or reports using a keyboard, and adjusting their writing on-screen, cutting and pasting to bring their writing to a logical or coherent whole; this is well nigh impossible to do well with pen and paper, against the clock, in a threateningly silent environment. Moreover, there’s a factor we don’t see but that students know all too well – speed of handwriting.
10. **We’re often tired and bored when we mark exam scripts.** Because of the speed of marking, and the pressure to do the task well, we may not be functioning at our best while undertaking the task. Our fairness can suffer. We can suffer from ‘halo’ effects after marking an excellent answer, and the reverse after marking a poor one.
11. **We’re not good at marking objectively.** There is abundant data on the problems both of inter-assessor reliability and intra-assessor reliability, particularly with the more qualitative or discursive kinds of exam answer.
12. **Unseen written exams tend to favour candidates who happen to be skilled at doing exams!** We’ve created an examinocracy! If we look at exactly what skills are measured by unseen written exams, the most important of these from students’ point of view turns out unsurprisingly to be the techniques needed to do unseen written exams, and the same students can get rewarded time after time! This skill may have little to do with the competences we need to help students to develop to become professionals in the subject disciplines they are learning.
13. **Unseen written exams force students into surface learning, and into rapidly clearing their minds of previous knowledge when preparing for the next exam.** Students are encouraged to clear their brains of the knowledge they have stored for each exam in turn. This of course is quite contrary to our real intentions to help students to achieve deep learning.
14. **There are many important qualities which are not tested well by traditional exams.** For example, unseen written exams are limited or useless for measuring teamwork, leadership, and even creativity and lateral thinking, all of which have their parts to play in heading towards graduateness.

Two final concerns are expressed by Gibbs (2010):

Exams can have the effect of concentrating study into a short intense period at the end of the course with, for example, little study of lecture notes until many weeks after the lecture. (Gibbs, p.10)

Teachers rarely set tests or exam questions with the deliberate intention of inducing a surface approach, but they do often allow students to accumulate enough marks to pass without ever doing anything more sophisticated. For students, that may be all the encouragement they need. Gibbs, p.23).

Despite all these concerns, there is a lot we can do to make exams work better or in different ways, for example open book exams, open notes exams, time-unconstrained exams, in-tray exams, OSCEs and so on. Enough for the moment about traditional exams; not all is well, however, with continuous assessment either: read on!

**Pros and cons of sixteen assessment processes (extracts on exams and essays only in this digest)**

Assessment can take many forms, and it can be argued that the greater the diversity in the methods of assessment, the fairer assessment is to students. Each and every one of the forms of assessment I consider in this chapter can be claimed to disadvantage those students who do not give of their best in the particular circumstances in which it is used. Therefore, diversifying assessment so that students experience a range of assessment methods evens out the situation, and increases the chance that all students will be able to demonstrate their best performance in at least some of the formats. The art of assessing therefore needs to embrace several different kinds of activity. I would like to encourage colleagues to broaden the range of assessment processes, (and to reduce sometimes quite dramatically the *size* of at least some individual assessments – for example a 300 word argument instead of a 3000 word essay, and so on), and I have tried to provide practical suggestions about how to maximise the benefits of each of a number of methods I have addressed below.

It must be added, however, that *rehearsal* is critically important to students encountering any new or different kinds of assessment from those to which they have become accustomed. Students develop their own ways of coping with, and preparing for the kinds of assessment they already know. To show themselves at their best in any unfamiliar kinds of assessment, they need time to adjust their preparation approaches, as well as the ways in which they provide in these different circumstances evidence of their learning.

In the next part of this chapter, I will look systematically at each of sixteen forms of assessment, listing a few advantages, some disadvantages, and I will offer some suggestions (sometimes a few, sometimes a lot) for making the particular assessment device work better. None of these lists should be considered as anything more than a starting point. Nor should the sixteen kinds of assessment I happen to have chosen be taken as representative of a sufficiently diverse range of assessment processes. Some of this discussion is further developed in Race *et al* (2005), Brown and Race (2012) and particularly now in Race (2014).

**1 Traditional unseen, time-constrained written exams**

Traditional unseen written exams still make up the lion’s share of assessment in higher education, though in some disciplines, for example mathematics, engineering and sciences courses, this situation is considerably balanced by the inclusion of practical work, projects and other contributions to the evidence on the basis of which we grade and classify students. Despite growing concern about the validity and fairness of traditional exams, for all sorts of reasons they will continue to play a large part in the overall assessment picture. Despite many concerns about exams, I have tried in the following discussion to suggest a number of ways that the use of exams can be improved. I have given more suggestions about setting exam questions than for setting any of the other types of assessment explored in this chapter as, in general, good practice in writing exam questions overlaps with, or extends across, many of the other types.

Nowadays, this kind of exam does not have to be constrained to handwritten answers, but keyboards and computers can be used, and with suitable invigilation such exams can be conducted online. With computers or online environments there are of course significant complications regarding access to information or data from outside the exam, and typing speeds and accuracy may interfere with performance.

ADVANTAGES

* **Exams are relatively high on transparency.** Most students have experienced them before, and have a reasonable idea of how they work, and indeed have fared relatively successfully with exams to get into higher education in the first place.
* **Exams can be reasonably high on reliability.** Where rigorous moderation of marking is in place, and where high-quality marking schemes are designed and adhered to, the fairness of exam marks can be reasonably well assured.
* **Relatively economical.** Exams can be more cost-effective than many of the alternatives (though this depends the duration of the exams, and on economies of scale when large numbers of students are examined, and also on how much time and money needs to be spent to ensure appropriate moderation of assessors’ performance). However, any form of assessment can only be truly said to be cost-effective if it is actually *effective* in its contribution to students’ learning.
* **They don’t have to be long!** In higher education it is still common to have exams of two or three hours duration, but with different kinds of structured question it is possible to have a much shorter exam and still test a great deal.
* **Equality of opportunity.** Exams are demonstrably fair in that students have all the same tasks to do in the same way and within the same timescale. (However, not all things are equal in exams – ask any hay-fever sufferer, or candidate with additional needs such as dyslexia, or with menstrual problems).
* **We know whose work it is.** It is easier to be sure that the work being assessed was done by the candidate, and not by other people. For this reason, exams can be considered to be an ‘anti-plagiarism assessment’ device, and although there are instances of attempting to cheat in exam rooms, good invigilation practice and well-planned design of the room (and the questions themselves) can eliminate most cheating.
* **Teaching staff are familiar with (long!) written exams.** Familiarity does not always equate with validity, but the base of experience that teaching staff already have with traditional unseen exams means that at least some of the problems arising from them are well known, and sometimes well addressed.
* **Exams cause students to get down to learning.** Even if the assessment method has problems, it certainly causes students to engage deliberately with the subject matter being covered by exams, and this can be worthwhile particularly for those more-difficult topics where students may not otherwise spend the time and energy that is needed to make sense of the subject matter.

DISADVANTAGES

* **Written time-constrained exams are often ‘low’ regarding validity.** In many subjects, exams are not the best way of getting from students evidence of achievement of the intended learning outcomes – oral assessment can be much more valid. However, in some subjects (maths, science, engineering) exams can be a valid way of allowing students to demonstrate their skills at problem-solving and numerical manipulation.
* **Speed of handwriting, and legibility may be what is really measured.** In an age where far less than formerly is handwritten, and where most writing is done through a keyboard, handwritten exams may discriminate significantly against those who are not used to writing with a pen!
* **Where answering questions is computer-based, speed and accuracy of typing/keyboarding may be too important.** This may be as serious as the disadvantages regarding speeds/legibility of writing in handwritten exams.
* **Students get little or no feedback on their exam performance.** Their work preparing for and sitting exams is therefore wasted as far as feedback is concerned. Though it can be argued that the purpose of exams is measurement rather than feedback, the counter-argument is that most exams, to some extent, represent lost learning opportunities because of this lack of feedback. Add to this picture the fact that most exam markers write comments on the scripts anyway, for those who may be moderating the standard or the marking, and these comments are wasted as far as students are concerned (while in practice, students are often fascinated when they have the chance to see this kind of comment on their answers). Where students are given the opportunity to see their marked scripts (even with no more feedback than seeing the subtotals and total marks awarded along the way), they can learn a great deal about exactly what went wrong with some of their answers, as well as having the chance to receive confirmation regarding the questions they answered well.
* **Badly set exams encourage surface learning.** Students consciously clear their minds of one subject as they prepare for exams in the next subject. In many discipline areas, it is inappropriate to encourage students to put out of their minds important subject matter, where they will need to retain their mastery for later stages in their studies.
* **Exam technique is too important.** Exams tend to measure how good (and fast) students are at answering exam questions, rather than how well they have learned. The consequence is that those students who become skilled at exam technique are rewarded time after time, while other students who may have mastered the subject material to a greater degree may not get due credit for their learning if their exam technique repeatedly lets them down, or if they ‘shoot themselves in the foot’ by running off at great detail about something they’re really interested in, failing to leave time to answer the rest of the questions properly.
* **Exams only represent a snapshot of student performance, rather than a reliable indicator of it.** How students perform in traditional exams depends on so many other factors than their grasp of the subject being tested. Students’ state of mind on the day, their luck or otherwise in tackling a good question first, their state of health, and many other irrelevant factors creep in.

***Helping students themselves to tune in to what written exams actually measure?***

The table below presents an exercise which can be given to students to alert them to some of the many factors which can affect their performance in exams. Even better, it can be useful as a class discussion exercise, allowing students to question tutors about some of the items on the agenda. The exercise can also be valuable in its own right to those setting and marking exams, as an indicator of the extent to which exams are achieving their desired goals. You might well wish to add to (and subtract from) the forty questions presently in this table.

| **Which factors are measured by written exams?** | ***Measured very well*** | ***Measured to some extent*** | ***Not really measured*** |
| --- | --- | --- | --- |
| 1. How much you know about your subject on the day. |  |  |  |
| 1. How much you *don’t* know about your subject on the day. |  |  |  |
| 1. How well you’ve been concentrating in related lectures. |  |  |  |
| 1. How conscientiously you’ve done suggested tasks between lectures. |  |  |  |
| 1. How widely you’ve read around the subject. |  |  |  |
| 1. How long you’ve spent online working at the subject. |  |  |  |
| 1. How well you’ve kept the published intended learning outcomes in mind as you studied. |  |  |  |
| 1. How well you’ve studied information about the assessment criteria. |  |  |  |
| 1. How often you’ve practised assessing your own answers to typical exam questions. |  |  |  |
| 1. How much you’ve learned from tutor-feedback on coursework assignments. |  |  |  |
| 1. How well you’ve kept your head down working on your own at the subject. |  |  |  |
| 1. How much you’ve talked to others about the subject and learned from them. |  |  |  |
| 1. The *quantity* of revision that you have done. |  |  |  |
| 1. The *quality* of revision that you have done. |  |  |  |
| 1. How intelligent you are. |  |  |  |
| 1. How determined you are to get a really good mark. |  |  |  |
| 1. How well you maintain your concentration during the exam. |  |  |  |
| 1. How much ‘polishing’ work you’ve done the night before. |  |  |  |
| 1. How well you keep your cool in the run up to the exam. |  |  |  |
| 1. How well you keep your cool on the day of the exam. |  |  |  |
| 1. How unruffled you remain if things go wrong while attempting a question. |  |  |  |
| 1. How well you resist getting carried away when you know a lot about a question. |  |  |  |
| 1. How good your memory is. |  |  |  |
| 1. How good you have been at question spotting. |  |  |  |
| 1. How carefully you read each question before choosing to attempt it. |  |  |  |
| 1. How fast you think. |  |  |  |
| 1. How fast you write. |  |  |  |
| 1. How legible your handwriting is. |  |  |  |
| 1. How well you manage your timing as you answer questions. |  |  |  |
| 1. How much practice you’ve had at thinking through how to answer exam questions. |  |  |  |
| 1. How much practice you’ve had at actually writing answers to exam questions. |  |  |  |
| 1. How carefully you re-read the questions while you’re answering them. |  |  |  |
| 1. How wisely you choose the questions that you attempt. |  |  |  |
| 1. How well you leave time to re-read and edit your answers. |  |  |  |
| 1. How well you keep exactly to the questions in your answers. |  |  |  |
| 1. How well you set out your answers to the questions. |  |  |  |
| 1. How skilled you have become at solving problems. |  |  |  |
| 1. How easy you make it for the marker to see how exactly you’ve worked out things in your answers. |  |  |  |
| 1. How carefully you read your own answers after writing them. |  |  |  |
| 1. How well you edit/improve your answers after reading them. |  |  |  |

***Setting unseen written exam questions: some practical suggestions***

Many experienced lecturers remember with some horror the first time they put pen to paper to write exam questions. Sometimes they felt well equipped to do so, as they had been involved in exams as candidates for most of their lives, and thought that it was quite straightforward to write good questions. But then the realisation dawned that the words and tasks used in exam questions could determine students’ future careers, prospects, incomes and lifestyles. Often, only when marking the exam scripts do lecturers first become aware of just how sensitively the questions need to be designed, and how clearly the assessment criteria and marking schemes need to be laid out to anticipate as many as possible of the different ways that even the most unambiguous looking question can turn out to be answered in practice. The suggestions below can help to spare you from some of the headaches which can result from hastily written exam questions.

1. **Aim towards writing shorter exams!** Remember that long exam questions tend to measure students’ legibility and speed of handwriting (or keyboarding), rather than mastery of the subject. Always ask yourself ‘is an answer to this exam question the best way for students to show that they’ve achieved the relevant learning outcome(s)’.
2. **Remember that in an exam question, you’ve normally only got print on paper.** There is no ‘tone of voice’ or body-language or any other means of clarifying what the question really means, or how you intend students to structure their answers.
3. **Whatever else your question is, students are very keen that it’s *fair.*** External examiners are keen on this too. There should be no tricks or hidden depths. Check carefully that each question is clearly linked to the course documentation which students have.
4. **Don’t write questions on your own!** Make sure you get feedback on each of your questions from colleagues. They can often spot whether your question is at the right level more easily than you can. Having someone else look at one’s draft exam questions is extremely useful. It can be better still when all questions are discussed and moderated by teams of staff. Where possible, draft questions *with* your colleagues. This allows the team to pick the best questions from a range of possibilities, rather than use every idea each member has.
5. **Ask colleagues: ‘what would you say this question really means?’** If they tell you anything you hadn’t thought of, you may need to adjust your wording a little. Try this out on students too.
6. **Get one or two colleagues to *do* your questions!** Sometimes even sketch answers can be helpful. This may be asking a lot of busy colleagues, but the rewards can be significant. You will often find that they answered a particular question in a rather different way than you had in mind when you designed the question. Being alerted in advance to the ways that different students might approach a question gives you the opportunity to accommodate alternative approaches in your marking scheme, or to adjust the wording of your question so that your intended or preferred approach is made clearer to students.
7. **Have your intended learning outcomes in front of you as your draft your questions.** It is all too easy to dream up interesting questions which turn out to be tangential to the learning outcomes. Furthermore, it is possible to write too many questions addressing particular learning outcomes, leaving other outcomes unrepresented in the exam.
8. **Keep your sentences short.** You’re less likely to write something that can be interpreted in more than one way if you write plain English in short sentences. This also helps reduce any discrimination against those candidates whose second or third language is English.
9. **Work out what you’re really testing.** Is each question measuring decision-making, strategic planning, problem solving, data processing (and so on), or is it just too much dependent on memory? Most exam questions measure a number of things at the same time. Be upfront about all the things each question is likely to measure. In any case, external scrutiny of assessment may interrogate whether your questions (and your assessment criteria) link appropriately with the published learning outcomes for your course or module.
10. **Don’t measure the same things again and again.** For example, it is all too easy in essay-type exam questions to repeatedly measure students’ skills at writing good introductions, firm conclusions, and well-structured arguments. Valuable as such skills are, we need to be measuring other important things too.
11. **Include data or information in questions to reduce the emphasis on memory.** In many subjects, case-study information is a really good way of doing this. Science exams often tend to be much better than other subjects in this respect, and it is appropriate to be testing what candidates can *do* with data rather than how well they remember facts and figures.
12. **Make the question layout easy to follow.** A question with bullet points for separate parts can be much easier for (tense) candidates to interpret correctly than one which is just several lines of continuous prose.
13. **Don’t overdo the standards.** When you’re close to a subject, it’s easily possible that your questions get gradually harder year by year. For example, in exams including quantitative questions, there is the danger that numerical problems become more difficult in each successive exam, partly because of the wish to stretch students a little further than did the worked examples they may have seen in lectures, or the problems students tackled online or in tutorials or coursework assignments.
14. **Write out an answer to your own question.** Don’t leave this till after the exam has been taken – that’s just going to make marking much harder work. A prepared answer is handy when you come to mark answers, but also you’ll sometimes find that it takes *you* an hour to answer a question for which candidates have only half an hour. Lecturers setting problem-type questions for students often forget that familiarity with the type of problem profoundly influences the time it takes to solve it. Students who get stuck on such a question may end up failing the exam more through time mismanagement than through lack of subject-related competence.
15. **Decide what the assessment criteria will be.** Check that these criteria relate clearly to the syllabus intended learning outcomes. Make it your business to ensure that students themselves are clear about these intended outcomes, and emphasise the links between these and assessment. When students are aware that the expressed learning outcomes are a template for the design of assessment tasks, it is possible for them to make their learning much more focused.
16. **Work out a tight marking scheme.** Imagine that you are going to delegate the marking to a new colleague. Write it all down. You will find such schemes an invaluable aid to share with future classes of students, as well as colleagues actually co-marking with you, helping them to see how assessment works.
17. **Use the question itself to show how marks are to be allocated.** For example, put numbers in brackets to show how many marks are attached to various parts of the question (or alternatively, give suggested timings such as ‘spend about ten minutes on Part 2’).
18. **Try your questions out.** Use coursework and student assignments to do pilot runs of potential components of your future exam questions, and use or adapt the ones that work best for exams.
19. **Proofread your exam questions carefully.** Be aware of the danger of seeing what you *meant,* rather than what you actually *wrote*! Even if you’re very busy when asked to check your questions, a little extra time spent editing your questions at this time may save you many hours sorting out how to handle matters arising from any ambiguities or errors which could have otherwise slipped through the proofreading process.

***Designing marking schemes***

Making a good marking scheme can save you hours when it comes to marking a large pile of scripts. It can also help you to know (and show) that you are doing everything possible to be uniformly fair to all students. As your marking schemes will normally be shown to people including external examiners and quality reviewers, it’s important to design schemes in the first place so that they will stand up to such scrutiny. Well-prepared marking schemes can also be enormously valuable to students themselves, to aid them practise making informed judgements on their own, and each others’ answers to typical questions.

The following suggestions should help.

1. **Write a model answer for each question.** (I suggested that you should have done this when you set the question of course). This can be a useful first step towards identifying the mark-bearing ingredients of a good answer. It also helps you see when what you thought was going to be a 30-minute question turns out to take an hour! If you have difficulties answering the questions, the chances are that your students will too! Making model answers and marking schemes for coursework assignments can give you good practice for writing exam schemes.
2. **Make each assessment decision as straightforward as possible.** Try to allocate each mark so that it is associated with something that is either present or absent, or right or wrong, in students’ answers.
3. **Aim to make your marking scheme usable by a non-expert in the subject.** This can help your marking schemes be useful resources for students themselves, perhaps in next year’s course.
4. **Aim to make it so that anyone can mark given answers, and agree on the scores within a mark or two.** This is to maximise *reliability* (fairness) of assessment in due course. It is best to involve colleagues in your piloting of first-draft marking schemes. They will soon help you to identify areas where the marking criteria may need clarifying or tightening up.
5. **Allow for ‘consequential’ marks.** For example, when a candidate makes an early mistake, but then proceeds correctly thereafter (especially in problems and calculations), allow for some marks to be given for the ensuing correct steps even when the final answer is quite wrong.
6. **Pilot your marking scheme by showing it to others.** It’s worth even showing marking schemes to people who are not closely associated with your subject area. If they can’t see exactly what you’re looking for, it may be that the scheme is not yet sufficiently self-explanatory. Extra detail you add at this stage may help you to clarify your own thinking, and will certainly assist fellow markers.
7. **Think ahead to ‘honourable exceptions’.** Ask yourself whether your marking scheme is sufficiently flexible to accommodate a brilliant student who hasn’t strictly conformed to your original idea of what should be achieved. There are sometimes candidates who write exceptionally good answers which are off-beam and idiosyncratic, and they deserve credit for these.
8. **Consider having more than 20 available marks for a ‘20-mark question’.** Especially in essay-type answers, you can’t expect students to include all the things you may think of yourself. It may be worth having up to 30 or more ‘available’ marks, so that students approaching the question in different ways still have the opportunity to score well.
9. **Look at what others have done in the past.** If it’s your first time writing a marking scheme, looking at several other people’s ways of doing them will help you to focus your efforts. Choose to look at marking schemes from other subjects that your students may be studying, to help you tune in to the assessment culture of the overall course.
10. **Learn from your own mistakes.** No marking scheme is perfect. When you start applying it to a pile of scripts, you will soon start adjusting it. Keep a note of any difficulties you experience in adhering to your scheme, and take account of these next time you have to make one.

***Marking examination scripts to optimise reliability***

Reliability is about fairness. The research literature on assessment shows that some kinds of exam question are not at all reliable – especially essay-type questions. The following suggestions may help you approach the task of marking exam scripts efficiently, while still being fair and helpful to students.

1. **Be realistic about what you can do.** Marking scripts can be boring, exhausting and stressful. As far as constraints allow, don’t attempt to mark large numbers of scripts in short periods of time. Put scripts for marking into manageable bundles. It is less awesome to have ten scripts on your desk and the rest out of sight than to have the whole pile threatening you as you work.
2. **Scan through a number of scripts before you really start marking.** It’s useful to get a feel for the overall standard of the answers before you get deep into the scripts themselves. This can help you to pitch your marking at the right level from the outset and save you the painful task of having to go back and make significant adjustments to your marking of the first few scripts.
3. **Avoid halo effects.** If you’ve just marked a brilliant answer on a script, it can be easy to go into the *same* student’s next answer seeing only the good points and passing over the weaknesses. Try to ensure that you mark each answer dispassionately. Conversely, when you look at the *next* student’s answer, you may be over-critical if you’ve just marked a brilliant one.
4. **Watch out for prejudices.** There will be all sorts of things which you like and dislike about the style and layout of scripts, not to mention handwriting quality. Make sure that each time there is a ‘benefit of the doubt’ decision to be made, it is not influenced by such factors.
5. **Recognise that your mood will change.** Every now and then, check back to scripts you marked earlier, and see whether your generosity has increased or decreased. Be aware of the middle-mark bunching syndrome. As you get tired, it feels safe and easy to give a middle-range mark. Try as far as possible to look at each script afresh.
6. **Remind yourself of the importance of what you’re doing.** You may be marking a whole pile of scripts, but each individual script may be a crucial landmark in the life of the student concerned. Your verdict may affect students for the rest of their careers.
7. **Take account of the needs of second markers.** Many universities use a blind double-marking system, in which case you should not make any written comments or numbers on the scripts themselves, to avoid prejudicing the judgement of a second marker (unless of course photocopies have already been made of each script for double marking). You may find it useful to use post-it notes or assessment proformas for each script, so you are able to justify the marks you give at any later stage. Such aides-memoirs can save you having to read the whole scripts again, rethinking how you arrived at your numbers or grades.
8. **Make notes to explain your assessment decisions to second-markers.** These may be even more necessary to justify your decisions when necessary to external examiners, or to help you re-visit a particular script when a student ends up on a borderline or in case of appeals.
9. **Compose feedback for students too.** In most exams, the system may not allow you to write on the scripts the sort of feedback you would have given if the questions had been set as assessed coursework. However, students may still need feedback, and making notes for yourself of the things you would have explained about common mistakes can help you prepare some discussion notes to issue to students after the exam, or can remind you of things to mention next time you teach the same subjects.
10. **Devise your own system of tackling the marking load.** You may prefer to mark a whole script at a time, or just Question 1 of every script first, and so on. Do what you feel comfortable with, and see what works best for you.
11. **Provide feedback for yourself and for the course team.** As you work through the scripts, note how many students answered each question, and how well they performed. You may begin to realise that some questions turned out to have been very well written, while others could have been framed better. You will find out which questions proved to be the hardest for students to answer well, even when all questions were intended to be of an equal standard. Such feedback and reflection should prove very useful when designing questions next time round.
12. **Set aside time for a review.** Having marked all the scripts, you may wish to capture your thoughts, such as suggestions about changes for part of the course or module, or the processes used to teach it. It is really useful, however tired you feel, to write a short draft report on the marking as soon as you have completed it. Otherwise, important things which are still fresh in your tired mind will all too quickly evaporate away.

***Using exam questions as class exercises***

Answering exam questions well is still one of the principal skills which students need to develop to succeed in their studies in most subjects. In our attempts to increase the learning payoff of taught sessions, we can help students to develop their exam skills by making use of past exam questions. The following suggestions may help you to build related activities into your lectures and tutorials – but don’t try to implement more than two or three of these suggestions with any one cohort – you haven’t got time!

1. **Let a class have a try at an exam question under exam conditions.** Then ask students to exchange their answers, and lead them through marking their work using a typical marking scheme. This helps students to learn quickly how examiners’ minds work. It is well worth using the whole of at least one lecture slot for such an exercise; the learning payoff for students is likely to be considerably more than if you’d just spent an extra hour with one small element of their curriculum.
2. **Issue two or three old exam questions for students to try in preparation for a tutorial.** Then lead them through assessing their work using a marking scheme during the tutorial. Ask them to prepare questions on matters arising from the exercise, both on subject content and requirements for exams, and use their questions to focus tutorial discussion.
3. **Display an exam question on-screen in a large-group lecture.** Ask students in groups to brainstorm the principal steps they would take in the way they would approach answering the question. Then give out a model answer to the question as a handout, and talk the class through the points in the model answer where marks would be earned. It can also be useful to give out a flawed answer, and get the students to see the weaknesses in this. All this can be achieved in less than half of the overall time of a typical lecture, and you may be surprised at the levels of interest and attention which students pay to such elements in a lecture slot.
4. **In a lecture or a tutorial, get students in groups to think up exam questions themselves.** You can base this on work they have already covered, or on work currently in progress. Where possible, display each of these in turn, giving feedback on how appropriate or otherwise each question is in terms of standard, wording, length and structure. (You will get many questions this way which you can later use or adapt for next year’s exams or future coursework assignments!).
5. **Use exam questions to help students to create an agenda.** In a lecture or tutorial, give out two or three related exam questions as a handout. Ask students in groups to make lists of short questions that they don’t yet know the answers to. Then allow the groups to use you as a resource, quizzing you with these questions. You don’t have to answer them all at once – for some your reply will be along the lines ‘We’ll come to this in a week or two’, and for others ‘You won’t actually be required to know this’.
6. **Get students to make marking schemes.** Give them a typical exam question, and ask groups of students to prepare a breakdown of how they think the marks should be allocated. Discuss each of these in turn with the whole group, and give guidance to how closely the marking schemes resemble those used in practice.
7. **Get students to surf the net.** Ask them to find appropriate exam questions on the subjects they are studying. Suggest that they work in twos or threes, and bring the questions they find to the next class session. You can encourage them to download the questions they find, and then assemble a question bank on a course web page.
8. **Ask students in groups to think up a** **‘dream’ question.** Ask the groups to make bullet-point lists of the ten most important things that they would include in answers to these questions. These questions will give you useful information about their favourite topics.
9. **Ask students in groups to think up** **‘nightmare’ questions.** With these, you can open up a discussion of the causes of their anxieties and traumas, and can probably do a lot to allay their fears, and point them in the right direction regarding how they might tackle such questions if needed.
10. **Ask students to think of way-out, alternative questions.** Suggest that they think of questions which are not just testing of their knowledge and skills, but which get them to think laterally and creatively. This encourages deeper reflection about the material they are learning, and will probably give you some interesting ideas to use in future exams.

**2 Open-book exams**

In many ways these are similar to traditional exams, but with the major difference that students are allowed to take in with them sources of reference material. Alternatively, candidates may be issued with a standard set of resource materials that they can consult during the exam, and are informed in advance about what will be available to them, so that they can prepare themselves by practising to apply the resource materials. Sometimes, in addition, the ‘timed’ element is relaxed or abandoned, allowing students to answer questions with the aid of their chosen materials, and at their own pace.

There is the possibility of an online-equivalent of an open-book exam, where in a computer-based exam all students are provided with the same range of online resource materials. You can choose whether or not to inform them in advance of the range of materials they will be able to consult as they answer the questions.

ADVANTAGES

Open-book exams have many of the advantages of traditional exams, with the addition of:

* **Less stress on memories!** The emphasis is taken away from students being required to remember facts, figures, formulae, and other such information.
* **Questions can be used which would not be possible in traditional exams.** For example, where it would have been quite inappropriate to expect students to remember all the detail of a wide range of information, open-book exams can test how well they can navigate provided information, and pick out trends, review alternative points of view and so on.
* **Higher on authenticity.** Open book exams are closer to the real-world situation where relevant information can be assembled and consulted when tackling problems.
* **There is less emphasis on reproducing things.** There is no point in an open-book exam just writing out extracts from the available materials, and questions can get candidates to go deeper into making judgements, comparing sources, prioritising options and so on.
* **Retrieval skills can be measured.** It is possible to set questions which measure how well students can use and apply information, and how well they can find their way round the contents of books and even databases.
* **Slower writers helped?** If coupled with a relaxation in the timed dimension (e.g. a nominal ‘2-hour’ paper where students are allowed to spend up to three hours if they wish) some of the pressure is taken away from those students who happen to be slower at writing down their answers (and also students who happen to think more slowly).

DISADVANTAGES

* **Not enough books or resources!** It is hard to ensure that all students are equally equipped regarding the books they bring into the exam with them. Limited stocks of library books (and the impossibility of students purchasing their own copies of expensive books) means that some students may be disadvantaged.
* **Open-book exams require different sorts of exam questions.** It would not be enough simply to ask students to *find* things in the available resource-materials, and questions need to probe what students can actually *do* with the resource materials.
* **Estimating timing is not as straightforward as in ‘normal’ exams.** It can be hard to anticipate how long is a reasonable time to allow for students to undertake a task with resource materials available, as part of the time needs to be spent navigating the materials, detracting from the time available to actually answer the questions.
* **Reliability can be compromised.** Where different students have access to different resources, their work can reflect the range or quality of resources they have available, rather than the quality of their learning.
* **More desk-space is needed.** The tables in traditional exams are normally quite small, allowing a venue to accommodate quite large numbers of students. Students necessarily require considerably more desk space for open-book exams if they are to be able to use several sources of reference as they compose their answers to exam questions. This means fewer students can be accommodated in a given exam room, and therefore open-book exams are rather less cost-effective in terms of accommodation and invigilation.
* **Speeds of reading and of writing may limit candidates’ performance.** Marks can only be awarded on the basis of what candidates manage to do with the question, and when they have to spend significant time on browsing the resource materials, they have less time to do themselves justice in their written answers.

*Tips on setting open-book exam questions*

Many of the suggestions already offered regarding traditional exam questions still apply. In addition:

1. **Decide whether to prescribe the books or articles students may employ.** This is one way round the problem of availability of books. It may even be possible to arrange supplies of the required books or articles to be available in the exam room.
2. **Consider compiling a source-collection for the particular exam.** Check on copyright issues, and see if it is cost-effective to put together a set of papers, extracts, data, and other information from which students can find what they need to address the questions in the particular exam.
3. **Set questions which require students to do things with the information available to them,** rather than merely summarising it and giving it back.
4. **Provide rehearsal opportunities.** Where students have not met open-book exams before, they need at least some practice to familiarise themselves with the approach they need to adopt. It can be useful to provide a dry-run, asking students to time themselves accordingly when doing a set task with particular resource materials on their own, and bring their results to a class session where they are guided through marking their own (or each others’) work, helping them to see what may constitute a good response to the question.
5. **Make the actual questions particularly clear and straightforward to understand.** The fact that students will be reading a lot during the exam means that care has to be taken that they don’t read the actual instructions too rapidly.
6. **Focus the assessment criteria on what students will have done with the information,** and not just on them having located the most relevant or appropriate information.
7. **Plan for shorter answers.** Students doing open-book exams will be spending quite a lot of their time searching for, and making sense of, information and data. They will therefore write less per hour than students who are answering traditional exam questions ‘out of their heads’.

**3 Open-notes exams**

These are similar to open-book exams described above, but this time students are allowed to bring into the examination room any notes that *they* have prepared for the purpose. In other words, we are talking about a situation of ‘legitimised crib-notes’! Your first thought may be that this is all very strange, but in fact such exams can work surprisingly well. Many of the advantages and suggestions for open-book exams continue to apply – the following additional matters arise.

ADVANTAGES

* **Students can achieve a very significant learning payoff simply making the notes in the first place.** The act of making revision summaries can have high learning payoff. It is best not to place stringent limits on the amount of materials which students can bring in. Those who bring in everything they have ever written about your topic will be disadvantaging themselves in that it will take them much longer to search for the relevant parts of their notes, compared to students who have been really selective in summarising the important parts of your topic.
* **Higher on authenticity.** In real-world contexts, people usually have the opportunity to assemble the information they need before undertaking a task.
* **The emphasis on memory is reduced, allowing competence to be tested more effectively.** Open-notes exams can also spread candidates’ abilities out more fairly, as the better candidates will have made better notes in the first place.
* **Drafting out answers to likely questions is legitimised.** This may enable students to do ‘better-quality’ revision as they prepare for this kind of exam.
* **You can write shorter questions.** When it is up to the students to ensure that they have with them important information or data, you don’t have to put so much into the questions themselves.

DISADVANTAGES

* **Students need rehearsal at preparing for open-notes exams.** They may take two or three practice runs to develop the art of making comprehensive but manageable summaries of the important data or information you intend them to make available to themselves.
* **Candidates whose open notes were not very suitable are penalised quite severely.** Some of these candidates may have been better at answering traditional exam questions with no notes.
* **Extra desk space is needed, just as for open-book exams.**
* **We’re** **thinking of paper-based notes, including print-outs.** In practice nowadays, students are more likely to prefer to make computer-based notes, and may find it much easier to search their notes electronically rather than by browsing pages physically.

*Tips on designing open-notes exams*

1. **Think of giving a topic menu in advance.** This can save candidates from trying to prepare open notes on everything they have learned about your topic. It does, of course, also mean that you are letting them off the hook regarding trying to learn some of the things that you *don’t* include in your menu.
2. **Consider having an inspection process.** For example, let it be known that yourself or your colleagues will be keeping an eye on the range and content of the open notes, or even that they may be temporarily retained after the exam.
3. **Give students some practice at writing open notes.** For example, set a coursework task where students prepare open notes to a given specification (length, number of references quoted, and so on) on a given topic, then facilitate peer-assessment of the open notes in class (ideally having each participant assessing at least two peers’ notes) on quality, depth, and coverage. Students learn a great deal from seeing each others’ ways of going about the task, helping them to make much better open notes where they are to be used for ‘real’ exams.

**4 Structured exams: for example multiple-choice questions**

These include multiple-choice exams, and several other types of formats where students are not required to write ‘full’ answers, but are involved in making true/false decisions, or identifying reasons to support assertions, or fill in blanks or complete statements, and so on. It is of course possible to design mixed exams, combining free-response traditional questions with structured ones. Some kinds of structured exams can be computer-based, either online synchronously in a normal ‘exam’ environment, or synchronously but at different locations (with appropriate invigilation). Software can be used both to process students’ scores and to provide feedback to them. In the following discussion, I will concentrate on the benefits and drawbacks of multiple-choice and multiple-response questions. Many of the same points also apply at least in part to other types of structured exam questions, such as true–false, short-answer, and sequencing questions.

ADVANTAGES

* **Greater syllabus coverage can be achieved.** It is possible, in a limited time, to test students’ knowledge of a much greater cross-section of a syllabus than could be done in the same time by getting students to write in detail about a few parts of the syllabus.
* **Multiple-choice exams can be high on validity.** In other words, they can be a better way of measuring whether candidates have made sense of a topic, than just asking candidates to write out things that they know.
* **Multiple-choice exams can be high regarding reliability.** The possibilities of unfair subjectivity in marking are eliminated.
* **Multiple-choice exams can be high regarding authenticity.** In many disciplines, the ability to make decisions between options can be closer to real-world tasks.
* **Multiple-choice exams can test how fast students think,** rather than how fast they write. The depth of their thinking depends on how skilled the question-setters have been.
* **Students don’t have to waste time writing out complex things like equations.** For example, questions can already show, for example, formulae, definitions, equations, statements (correct and wrong) and students can be asked to select the correct one, without having to write it out for themselves.
* **Staff time and energy can be reduced dramatically.** With computer-based assessment platforms, or optical mark readers for paper-based answer scripts, it is possible to mark multiple-choice exams very quickly and cost-effectively, and avoid the tedium and subjectivity which affect the marking of traditional exams.
* **Computer software can be used to analyse the effectiveness of the questions.** As well as processing all of the scores, computer software can work out how each question performs, calculating the discrimination index and facility value of each question. This allows the questions which work well as testing devices to be identified, and selected for future exams.
* **Higher-level thinking can be tested.** Multiple-choice exams can move the emphasis away from memory, and towards the ability to interpret information and make good decisions. However, the accusation is often made that such exams seem only to test lower cognitive skills, but this is usually because the questions themselves have not been made more challenging. There are numerous, examples where high level skills are being tested effectively.
* **Multiple-response can be used, as well as multiple-choice.** Multiple-response is where more than one option can be correct, with the stem asking (for example) “Which (one or more) of the following options is correct?”. These can go much deeper than simple multiple-choice questions, as some of the ‘correct’ answers can be less-obvious than others, and considerably more thought and knowledge can be required of candidates.
* **Multiple-choice questions can have more than one ‘layer’.** For example in computer-based exams, the first layer can ask which is the correct option out of a series, then the next layer can ask *why* this option is the best one, and *why* each distractor is wrong, asking candidates to select the best *reason* in each case. This substantially reduces the ‘guess-factor’ of the overall question.

DISADVANTAGES

* **The guess factor.** In simple multiple-choice questions, students can often gain marks by lucky guesses rather than correct decisions.
* **It is surprisingly hard to write good multiple-choice questions.** What seems like a good question can turn out to be too easy, with most candidates selecting the correct option, not always for a good reason. Conversely, sometimes a distractor can be too distracting, and can cause casualties with even the best candidates.
* **Designing structured questions takes time and skill.** It is harder to design good multiple-choice questions than it is to write traditional open-ended questions. In particular, it can be difficult to think of the last distractor or to make it look sufficiently plausible. It is sometimes difficult to prevent the correct answer or best option standing out as being the one to choose.
* **Black and white or shades of grey?** While it is straightforward enough to reward students with marks for correct choices (with zero marks for choosing distractors), it is more difficult (but not impossible) to handle subjects where there is a ‘best’ option, and a ‘next-best’ one, and so on.
* **Where multiple-choice exams are being set on computers, it is necessary to check that the tests are secure.** Students can be ingenious at getting into computer files that are intended to be secret!
* **The danger of impersonators?** The fact that exams composed entirely of multiple-choice questions do not require students to give any evidence of their handwriting increases the risk of substitution of candidates.

*Designing multiple-choice exams*

1. **Continuously try out questions with colleagues and with large groups of students.** Make sure that you select for exam usage questions where people are selecting correct options for the right reasons – and not because in one way or another the question gives away which is the correct option.
2. **Make sure that distractors are plausible.** If no one is selecting a given distractor, it is serving no useful purpose. Distractors need to represent anticipated errors in students’ knowledge or understanding.
3. **Try to avoid overlap between questions.** If one question helps students successfully to answer further questions, the possibility increases of students picking the right options for the wrong reasons.
4. **Avoid options such as ‘none of the above’ or ‘all of the above’.** These options are a let-out for students who find it hard to decide between the other alternatives, and are often chosen by weaker students in surface-thinking mode. Also, it is surprisingly rare for such options to be in fact the correct one, and test-wise candidates will already have guessed this. To complicate matters, the best students will sometimes spot weaknesses with the option which is intended to be correct, and select ‘none of these’ because of this.
5. **Avoid where possible the restrictions of being limited to a set number of options.** For example, four-option or five-option questions are often used, even when thinking of the ‘last’ distractor is difficult, resulting in the option concerned having little real value in the test. It is better when as many distractors as are realistic errors-of-thinking are allowed, so that each item serves a useful purpose.
6. **Pilot questions in formative tests before using them in summative exams.** Ideally, multiple-choice questions that appear in formal exams should be tried-and-tested ones. It is worth consulting the literature on multiple-choice question design and finding out how to assess the discrimination index and facility value of each question from statistical analysis of the performance of substantial groups of students.
7. **Remember that students can still guess.** The marking scheme needs to take into account the fact that all students can score some marks by pure luck! If most of the questions are, for example, four-option ones, the average mark which would be scored by a monkey would be 25 per cent, so the real range lies between this and 100 per cent. It is important that people are indeed allowed to get 100 per cent in such structured exams, and that this does not cause any problems when the marks are blended with more traditional exam formats where written answers in some subjects still attract marks only in the 70s even when they’re reckoned to be first-class answers.
8. **Design feedback responses to each option.** Where possible, it is useful to be able to explain to students selecting the correct (or best) option exactly *why* their selection is right. It is even more useful to be able to explain to students selecting the wrong (or less-good) options exactly what may be wrong with their thinking. When multiple-choice questions are computer-marked, it is a simple further step to get the computer to print out (or display on-screen if the exam is computer-based) feedback responses to each student. This practice can equally be applied to formative multiple-choice tests, and to formal multiple-choice exams. Furthermore, the availability of feedback responses to each decision students make lends itself to extending the use of such questions in computer-based learning packages.
9. **Ensure that students are well-practised at handling multiple-choice questions.** Answering such questions well is a skill in its own right, just as is writing open answers well. We need to ensure that students are sufficiently practised, so that multiple-choice exams measure their thinking and not just their technique.
10. **Look at a range of published multiple-choice questions.** For example, in the UK several Open University courses have multiple-choice assignment questions, as well as multiple-choice exams. You may be surprised how sophisticated such questions can be, and may gain many ideas that you can build into your own question design.
11. **Gradually build up a large bank of questions.** This is best done by collaborating with colleagues, and pooling questions that are found to be working well. It then becomes possible to compose a multiple-choice exam by selecting from the bank of questions. If the bank becomes large enough, it can even be good practice to publish the whole collection, and allow students to practise with it. Any student who has learned to handle successfully a large bank of questions can normally be said to have learned the subject well.
12. **Involve students in groups in designing multiple-choice questions.** When students have learned a topic relatively recently, they can still remember the things that confused them at first, and this helps them to design good distractors. The act of designing such questions has high learning payoff, and serves as good rehearsal for the relevant technique students need for this kind of exam.
13. **When you’ve got a large bank of questions, there is the possibility of on-demand exams.** Students can then take a multiple-choice test with a random selection of questions from the bank, at any time during their studies, and ‘pass’ the component involved as soon as they are able to demonstrate their competence with the questions.

**5 Essays: in exams and in coursework**

In some subjects, assessment is dominated by essay-writing. This has often been the case for a long time, and essays have become a firmly-established form of assessment in traditional (and open-book) written exams. Assessed coursework often takes the form of essays, formerly handwritten but nowadays almost always word-processed. It is well known that essay-answers tend to be harder to mark, and much more time-consuming to assess, than quantitative or numerical questions. There are still some useful functions to be served by including some essay questions in exams or coursework assessments, but perhaps we need to face up to the fact that reliability in marking essays is often unsatisfactory, and refrain from using essays to the extent that they are used at present.

ADVANTAGES

* **Essays allow for student individuality and expression.** They are a medium in which the ‘best’ students can distinguish themselves. This means, however, that the marking criteria for essays must be flexible enough to be able to reward student individuality fairly.
* **Essays can provide students with opportunities to demonstrate their own particular ‘take’ on a topic.** While this may be an advantage, it is also a disadvantage in that it can turn out to be particularly troublesome to *assess* students’ own ‘take’ fairly and without prejudice.
* **Essays can reflect the depth of student learning.** Writing freely about a topic is a process which can demonstrate understanding and grasp of the material involved.
* **Essay-writing is a measure of students’ written style.** It is useful to include good written communication somewhere in the overall assessment strategy. The danger of students in science disciplines, where essays are used less, missing out on the development of such skills is becoming increasingly recognised.
* **Students are relatively familiar with essays.** They’ve normally done this sort of writing beforehand, but assessors at university level justifiably grumble that students don’t seem to have had any real training in structuring essays, particularly when it comes to logical argument and coming to a resounding conclusion.

DISADVANTAGES

* **The assessment of essays is well proven to be unreliable.** Different markers often award the same essay quite different marks, and students are quick to notice such unfairness. Essays are demonstrably the form of assessment where the dangers of subjective marking are greatest. Essay-marking exercises at workshops on assessment show marked differences between the mark or grade that different assessors award the same essay – even when equipped with clear sets of assessment criteria.
* **The validity of essays as an assessment device remains questionable.** Students’ knowledge of the subject is only tested to a limited extent, and technique for essay-writing is tested rather better sometimes.
* **Essay-writing is very much an art in itself.** Students from some backgrounds are disadvantaged regarding essay-writing skills as they have simply never been coached in how to write essays well. For example, a strong beginning, a coherent and logical middle, and a firm and decisive conclusion combine to make up the hallmarks of a good essay. The danger becomes that when essays are over-used in assessment strategies, the presence of these hallmarks is measured time and time again, and students who happen to have perfected the art of delivering these hallmarks are repeatedly rewarded irrespective of any other strengths and weaknesses they may have.
* **Essays take a great deal of time to mark.** Even with well thought out assessment criteria, it can be difficult to ‘get into one’s stride’ applying a marking scheme, and it is not unusual for markers to need to work back through the first dozen or so of the essays they have already marked, as they become aware of the things that the best students are doing with the questions, and the difficulties experienced by other students.
* **‘Halo effects’ are significant.** If the last essay answer you marked was an excellent one, you may tend to approach the next one with greater expectations, and be more severe in your assessment decisions based upon it.
* **Essays take time to write (whether as coursework or in exams).** This means that assessment based on essay-writing necessarily is restricted regarding the amount of the syllabus that is covered directly. There may remain large untested tracts of syllabus.
* **With coursework essays, it is increasingly difficult to guarantee that the essay is the work of the candidate.** Ready-made essays can be purchased online on just about any subject, or even directly commissioned. Plagiarism is always a possibility, and there is the need to check authenticity by some other means, such as face-to-face questioning to guarantee that the work is the students’ own. This adds even more time to what is already a high-burden kind of assessment.
* **Traditional ways of giving students feedback on essays are known to be problematic, and rarely worth the time taken.** Writing comments on students’ work is particularly troublesome, in that it is well known that students don’t often make good use of the feedback, not least because it often arrives too late, when they’ve already moved on in their studies. Using the comments function in word-processing ‘track-changes’ can be much better, however (this is discussed in some detail later in this chapter).

*Tips on setting and using essay-type questions*

Many of the suggestions given earlier in this chapter about writing traditional exam questions continue to apply – whether essays are to be used as assessed coursework or as exam questions. Some further suggestions are given below.

1. **Help students to see exactly how essays are marked.** Alert students to the credit they gain from good structure and style. One of the best ways of doing this is to involve classes of students in looking at examples of past (good, bad and indifferent) essays, and applying assessment criteria. This can be followed by involving students in peer-assessment of each other’s essays, and indeed in self-assessment of their own essays. This helps them to put their own efforts into perspective, and to learn things to emulate (and things to avoid!) by seeing how other students go about devising essays.
2. **Don’t leave students to guess the real agenda.** Some essay questions are so open ended that it is hard for students to work out exactly what is being sought. The authors of such questions will defend their questions by saying ‘well, it’s important to find the students who know what to do in such circumstances’, but the fact remains that it is an aspect of study technique which is being rewarded, rather than mastery of the learning involved in answering the question.
3. **Subdivide essay questions into several parts, each with marks publicly allocated.** This helps to prevent students from straying so far off the point that they lose too many of the marks that they could have scored.
4. **Give word limits.** Even in exams, it can be useful to suggest to students that an essay answer should lie between (for example) 800 and 1200 words say for a 30-minute question, and so on. This helps to avoid the quantity-versus-quality issue, which leads some students into simply trying to write a lot, rather than thinking deeply about what they are writing – and it also helps reduce the time it takes to mark the essays.
5. **Be even firmer about word limits in coursework essays sometimes.** For example suggest ‘exactly 400 words’ with penalties for each word over or under. Using word-processing software it is easy for students to track the length of their essay on-screen as they compose it, and students often enjoy the challenge of ‘putting their best foot forward’ in this constrained amount of words. Writing to length is an important skill. Besides, it is demonstrably fair if all students are constrained in the same way, as it can be argued that anyone could score more marks if they had more words at their disposal.
6. **Help students to develop the skills required to plan the content for essays.** This is particularly important in those disciplines where students will be more accustomed to handling structured questions and problems. The danger then is that students tackling essay questions in exams spend far too long on them, and penalise themselves regarding time for the rest of the examination. One of the best – and most time-effective – ways of helping students to become better at handling essay questions is to set class or coursework tasks which require students to prepare essay-plans rather than fully finished masterpieces. A concept-map or diagram can show a great deal about the eventual ‘worth’ of students essays, and can avoid distraction from the elements of style and structure. Students can put together at least half-a-dozen essay plans in the time it would take them to complete one essay, and making the plans involves far more payoff per unit time in thinking and learning.
7. **Don’t assess essays too often.** Any assessment form advantages those students who happen to be skilled at delivering what is being measured. This applies to essays too, and there is a significant danger that those students who happen to become good at planning and writing essays continue to be advantaged time and time again.
8. **Have a clear, well-structured marking scheme for each essay question.** This can save a lot of time when marking, and can help guarantee that students’ answers are assessed fairly and consistently. That said, even the most carefully-designed marking schemes for essays always seem to need adjustment as soon as you start marking a batch, as students’ answers always range more widely than one anticipated.
9. **Don’t assume that longer equals better.** It is often harder for students to write succinctly than to just ramble on. However, students need to be briefed on how best we want them to develop their art in writing concisely.
10. **Help students to improve their technique through feedback.** Consider the range of approaches you can use to give students useful feedback on their essays, including statement banks, assignment return sheets and email messages, and try to minimise the time you spend writing similar feedback comments onto different students’ essays.
11. **Let technology help you to give useful feedback.** The track-changes function on word processing software allows you to speed up giving comments when marking on-screen. In particular, the *comments* can be linked visually to the word, phrase or paragraph you’re giving feedback on, and allows you to say more than you could squeeze in between the lines of an essay.
12. **Use some class time to get students to brainstorm titles for essays.** This helps them to think about the standard they could anticipate for essay questions in forthcoming exams, and gives them topic areas to base their practice on.

***Essay 1***

**Technology made large populations possible; large populations make technology indispensable**

The population of every country in the world has increased in the last 150 years, and in some countries the rate of increase has been much greater than in Britain. When populations increase, more food has to be produced to keep these people nourished. More houses have to be built, with increased water supply and facilities for sewage disposal. More schools and hospitals are needed. Forests are cleared and burnt to give fertile land to grow crops but if the increase in food production and provision of housing and other facilities does not take place at the same rate as the increase in population, then the standard of living falls.

The life expectation up to about 150 years ago rose slowly from an average of 20 years for early man to the 50s in the 19th century. Today’s life expectancy is 70–80 years and is therefore a population increase of 50%. Infant and child mortality has, until now, been high – in 1866 only 40–45% of those aged between 0 and 15 years were expected to survive. Diseases such as cholera, typhoid, tuberculosis and diphtheria took their toll, mainly on the poor living in slums. Improvements in the standard of hygiene and sanitation reduced this rate so that the survival rate of 0–15 year olds increased to between 62 and 75% in 1961 in Britain, but this is still lower in many parts of the world today.

The world population in 3000 BC was about 100 million. This slowly increased to about 350 million in 100 BC. In the Middle Ages the population was steady at around 350–500 million until the Black Death caused a drop to 250 million people. The population then increased rapidly to around 1000 million in the 1800s and then very rapidly to nearly 4000 million in the early 1980s. The projected rise is around 6500 million people by the year 2000.

Darwin stated in his book *The Origin of Species* that ‘there is no exception to the rule that every organic being naturally increases at so high a rate that if not destroyed, the earth would soon be covered by the progeny of a single pair’. He also recognized that there are many ‘checks to increase’ which limit the size of populations. For humans, the checks have included diseases and epidemics as well as famine over which until recent times, man has had very little control, and wars over which he does. Even now, his control is far from complete where diseases are concerned. The degree to which the incidence of fatal diseases such as smallpox, cholera and tuberculosis has been reduced since 1955, has resulted in an enormous increase in population throughout the tropics. With more mouths to feed, the threat of starvation has increased. Starving people are not as well able to work and increase food production as healthy, well-fed people.

The rise in population is exponential and one wonders what the next ‘check to increase’ can be. Already, shortage of food and water, and wars, are checking the population growth in many parts of the world. A check on the birth rate would enable the increase in food production to catch up and improve the standard of living of those surviving in the poorer parts of today’s world.

High technology is an essential component of medicine today and it can have substantial benefits when appropriately used. A positive and accurate diagnosis made quickly with little discomfort and risk, can lead to quick and efficient appropriate management with resulting reduction in mortality and morbidity. Restoring physiological function (by pacemakers, renal dialysis, open heart surgery or joint replacement) can restore patients to a full and near-normal length of life. Powered prostheses for the severely disabled and simpler aids for other disabilities, can enhance the quality of life for years by allowing greater degrees of independence and easier communication.

The world population cannot continue to grow at its present rate, its resources are running out through use as well as destruction. Conservation has to be now; in ten years time it might well be too late for us all!

***Essay 2***

**Technology made large populations possible; large populations make technology indispensable**

At first sight this seems an excellent example of cause and effect. In this case, technology can be regarded as the cause, and large populations the effect. Looking back over human history, populations increased relatively slowly for millions of years, and the growth in population only became alarming in the last hundred years or so – the age of technology.

There are many examples supporting how technology makes large populations possible. Medical science and technology have allowed many former killer diseases to be eradicated, and consequently life expectancy in the developed world has increased more in the last hundred years than in our evolutionary history. Technology has made it possible to sustain large populations. For example, agricultural technology and fertilisers have allowed greater productivity from limited amounts of farmland. It has been possible to convert to agricultural use land that was previously unsuitable.

Large populations have indeed become dependent on technology. In the developed world, transportation (for example) is an essential part of the way we live, and without technology life as we know it would be difficult – if not impossible. In the home, less and less time is spent on basic tasks such as cooking, washing, cleaning and so on. Technology has been used to automate many of these. More and more time is devoted to leisure activities – particularly that of sitting at home and watching one particular form of technology for entertainment and enlightenment: the television set.

So it would appear that large populations have made technology indispensable. However, our planet is of finite size. The resources upon which many forms of technology draw are limited in availability. The amounts of fossil fuels are limited. The amounts of metal ores are limited. The ability of our ecosystem to absorb the toxic products of some forms of technology are limited. In other words, technology itself has limits. There comes a point where greater uses of technology start to produce irreversible changes in our ecosystem. The holes in the ozone layer may be an example of this.

While it has always been recognised that a finite planet can only support a finite population, it has only recently become recognised that our uses of some forms of technology can only take place on a limited scale. Despite the logic of limiting the population of our planet, there has not been any evidence of any real readiness to accept the implications of the steps necessary to limit population globally. Similarly, it seems difficult for people to agree to curb some forms of technology which are becoming life-threatening.

All is not bad news, however. There are some sources of technology which do not have any deleterious effects on our environment. For example, hydroelectricity, solar energy, wind power, tidal power and wave power are all ‘free’ forms of energy which do not cause toxins to be released into our ecosystem, and potentially allow us to continue to be dependent on various electrical technologies. Other renewable resources can allow us to continue using paper, some liquid fuels, and medicines. Often, the cost of using renewable resources is greater than that of consuming the finite reserves of our planet, but if large populations are to continue to depend on technology, that cost – sooner or later – will have to be accepted.

In conclusion, it is certainly true that technology played a major part in allowing large populations to develop. To sustain such populations, technology will indeed be indispensable – but only if the wisdom of choosing *appropriate* technologies is brought to bear on preserving our ecosystem. Otherwise, technology will eventually make large populations untenable – or large populations will make the continued use of some technologies impossible.

**Sources: Race, P. (2014) *Making Learning Happen: 3rd edition,* London: Sage.**

**Race, P. (2015) *The Lecturer’s Toolkit: 4th edition:* Abingdon, Routledge.**

**Contact: phil@phil-race.co.uk http://phil-race.co.uk Twitter: @RacePhil**