

Towards Assessment As Learning*

Phil Race

Leeds Metropolitan University (Emeritus)

Abstract

It is widely accepted that in higher education internationally, assessment drives students' learning. As a consequence, students become ever more strategic, and only put energy into things that count towards their overall assessment. This article suggests a practical way of setting about reflecting on the assessment processes and instruments you use.

Many institutions have now adopted 'assessment for learning' approaches, to make better links between assessment and learning. In this paper, I argue that we can go even further and work towards 'assessment as learning' where all of our elements of assessment are designed with learning in mind. The paper provides you with a scoring grid, using which you can interrogate your own assessment elements, and determine how well they measure up to a combination of 'assessment for learning' and 'assessment as learning'.

Keyword(s): learning, assessment.

* Invited article. URL: <http://ojs.aishe.org/index.php/aishe-j/article/view/6>

1. Setting the scene

1.1 Assessment design: some key terms to address

Many scholarly sources on the design of assessment in higher education discuss a range of dimensions which need to be addressed in order to make assessment fit for purpose. These include:

- Validity
- Reliability
- Transparency
- Authenticity
- Manageability
- Inclusiveness

Too often, terms like these are explained in language which does not make addressing them any easier. Also, addressing them all at once is rarely possible, and is at the very least a highly complex balancing act. Later in this paper, I will provide straightforward explanations of these terms, so that you can interrogate your own assessment elements in terms of how well each of these variables is successfully addressed.

1.2 Learning: processes underpinning success

Linking assessment to learning is even more complex than addressing all of the assessment qualities referred to so far. Yet it is possible to think of learning in terms of seven underpinning factors, each of which can be linked separately to assessment (and to teaching, the design of learning resources, and curriculum design as a whole). I have argued elsewhere ([Race 2005](#)) that these factors are as follows:

1. *Wanting* to learn – which could be thought of as ‘intrinsic motivation’.
2. Taking ownership of the *need* to learn – which could be thought of as ‘extrinsic motivation’.
3. *Learning-by-doing*: practice, trial-and-error, experimenting, experience, repetition when appropriate.
4. Learning from *feedback*: praise, criticism, seeing the results of learning.
5. *Making sense* of what is being learned: ‘getting one’s head round it’, ‘light dawning’.
6. Deepening learning by *explaining* to others, *coaching* others, and *teaching* what has been learned.
7. Further deepening learning by *making informed judgements* on one’s own work, other people’s work, self- and peer-assessing.

Clearly, 6 and 7 above are in effect special contexts in which 1-5 can be focussed and enhanced, but it can be argued that these contexts are so necessary to cause learning to be sufficiently deepened that they can be regarded as every bit as essential in the overall picture of learning as are the first five factors. I have argued in 'Making Learning Happen' (Race 2005) that these factors all interact with each other in a way much more complex than a learning cycle. Their interaction can be likened to ripples reverberating backwards and forwards on a pond after a pebble has been dropped into it, all affecting each other. This argument will be developed further and extended particularly to address 6 and 7 in the second edition of 'Making Learning Happen' (scheduled for publication in 2010), but meanwhile 6 and 7 are included in the grid which follows to interrogate an assessment element.

1.3 Design of the 'interrogation' grid

Within the grid (see [Appendix](#)), there are 20 decisions to be made for each assessment element of a course or module. The grid contains space to interrogate four separate assessment elements, but it is suggested that readers start with a single element (Element A), and work through the discussion in this paper first, using this element (as rehearsal) before returning to further elements of assessment. In this discussion 'element of assessment' simply means any assessed task or activity which counts towards the overall award being studied for, and has marks or grades associated with the level of performance students show in it. Typical examples include those shown below, but any other assessment elements can be considered.

- Written exam
- Presentation
- MCQ (Multiple Choice Question) exam
- Essay
- Practical test
- Drawing
- Sculpture
- Dance
- OSCE (Objective Structured Clinical Exams)
- In-tray exam
- Oral exam
- Interview
- Short-answer exam
- Written reflection
- Report

For each of the 20 lines of the grid, a decision is to be made on a 1-5 scale. '5' is where the design of the assessment element fits very well indeed with 'assessment as learning' as a goal, or is of very high quality in the context of the variables of 'validity', 'reliability' and so on. In each case, this paper provides guidance on how decisions may be made on this 1-5 scale.

1.3.1 Why '1-5' rather than '0-5'?

Simply so that an assessment element which is very poor at addressing the assessment variables, and has very little to do with 'assessment as learning' can still have a score – the minimum total score of 20, which could be said to describe an assessment element which had little to do with anything other than assessment of learning, and was also poor regarding validity, reliability, and so on.

1.3.2 Why are all the 20 lines interrogated on the same scale of 1-5?

Simply as a first approximation towards balancing the 20 factors considered towards an overall judgement of how well the assessment element approaches the goal of 'assessment as learning'. It is likely that adjustments to this 'equal rating' position will be desirable in future. This grid is simply meant as a starting point towards addressing interrogating assessment on a complex and inter-related set of variables, in a relatively straightforward manner.

1.3.3 Why not just fill in the grid straight away, without reading further to find out how?

You may in fact wish to do this for a chosen element of assessment, and then revisit each item in turn on the basis of the discussion which follows for each line of the grid, to self-assess how well your instinct about your assessment element lines up with the thinking behind the scoring suggestions made in this paper. However, the meanings of such terms as 'validity', 'reliability' and so on need to be clarified to allow you to make informed judgements as you interrogate your own assessment elements, and for your scores to be compared to others' scores for their assessment elements, and so on. Furthermore, you would be wise to consult your students to help you make a scoring decision for several of the items in the grid.

Another reason not to fill the grid in straight away, is that relatively familiar variables like validity, reliability are interspersed with factors relating to the quality of learning associated with the assessment, and the final three lines of the grid refer to the place of this particular assessment element in the overall assessment pattern for the course or module. So this paper takes you through the agenda in a manner that is intentionally non-linear.

2. Step by step guidance on the 20 factors

2.1 Students love it (wanting to learn)

If the assessment element is one which students dread (possibly a traditional exam) this might warrant a score of '1' – the minimum. If it's the sort of assessment which students really look forward to, and enjoy doing while they are performing it, that would be nearer a '5'. Of course, different students will have different views about how much they enjoy any particular

assessment form, so you may like to ask 20 or more students, and average their views to help you make your 1-5 decision here.

2.2 Students learn by doing preparing for it

This line is about the kind of learning students perform leading up to the assessment (and to some extent also any learning they achieve while actually doing the assessment). The minimum score of '1' would be warranted by any assessment where students merely 'filled their heads up with information' ready to regurgitate it during the assessment itself. This might apply to some traditional written exams in some disciplines, and equally to some kinds of essay or report where the learning by doing is quickly forgotten after the task is completed.

Higher scores may be associated with useful practice, problem-solving, explaining things to each other (or to anyone else who would listen), learning by getting things wrong and finding out exactly why, and so on. Decide as honestly as you can the extent to which students' learning is active while preparing for this particular assessment element on the usual scale: 1-5.

2.3 Students make sense of their learning preparing for it

This is linked, but sufficiently different to '2' above to warrant a separate line on the grid. This is more about how permanently students 'get their heads round things' in the way that this assessment causes them to prepare for it. A low score might be warranted if students simply prepare themselves in a way where they are 'OK on the day' and then let their learning slip, perhaps quite intentionally, as they prepare for the next assessment on their schedule. A high score for this item would be where you know that most students prepare for this assessment in a way which really consolidates their learning (and not merely that you *intend* them to prepare in such a way!). You may indeed wish to check this out with a sample of students.

2.4 Students coach each other preparing for it

This time, the student activity represented by this line is quite explicit. If students are involved in coaching or teaching each other as they work towards this assessment, the score could be as high as '5'. If they prepare entirely on their own, in solemn isolation, it could be '1'. If they're involved in discussing, explaining to each other, and so on, the score could be nearer '5'. Remember, it's what students actually do that governs the score for this – not necessarily what you *hope* they do: ask them.

2.5 Students practise making informed judgements

If students' preparation for this assessment centres mainly around self-assessing their own learning, and peer-assessing the products of fellow-students' learning, this would score a '5'. If this assessment does not involve students making informed judgements in such ways, it's more likely to be a '1'. If students do at least some self- or peer-assessment, the score might be somewhere between '1' and '5'. Making informed judgements on material from the literature is of course still useful (perhaps warranting a score of '2' or '3'), but almost certainly less intense an experience as self-assessing their own work, or peer-assessing each others' work, where their judgements need to be able to be supported by feedback to each other.

2.6 Students design the criteria

This relates particularly to the extent to which students 'take ownership of the need to learn'. When they have worked together to establish the criteria, the assessment is much more 'owned' by them. When they are using self- and peer-assessment, using criteria whose design they have shaped, the ownership is at its best, and a '5' might be warranted for this line. If students have merely been given 'the official criteria' the score might only be '2' or so – and if students have no idea of the exact nature of the criteria, the minimum of '1' might be appropriate here.

2.7 Students own the weighting of the criteria

While this might at first sight seem to overlap with '6' above, the process of students sorting out what is important and what isn't in their design of the criteria has so much to do with their 'making sense' of the topic that this criterion deserves a line of its own. If students have collectively worked out the marking scheme, a '5' may be warranted for this line, especially if this was done in the context of something they self-assessed or peer-assessed.

2.8 Validity

The crux here is the answer to the question 'is this assessment really measuring what it is intended to measure?' rather than 'simply measuring what happens to be easily measurable'. For example, a traditional unseen written exam might score a '1' here, as it is just measuring what students can do on their own, in a quiet room, with what comes out of their heads and gets through in handwriting to their answer scripts legibly enough to score them marks. That said, a written maths exam measures quite well whether students can do maths, and a problem-solving kind of exam does measure whether students can solve some kinds of problems, and the scores could be higher for this line.

However, higher scores may be more readily warranted on the grounds of validity for assessments such as an OSCE (objective structured clinical exam) as used by medical students, where what is measured is essentially what practitioners are intended to be able to do, such as interpreting a set of X-rays, scanning a patient's case notes to arrive at a prescription, talking to a patient (an actor in practice) to work out what is wrong, and so on. Similarly, a presentation may have high validity as an assessment format, if students need not only to master some learning but also present it authoritatively and clearly to others. An oral exam (viva) may also score highly on validity, if it is felt that this is the most effective way of determining the extent of students' learning. Furthermore, an exam of the 'in-tray' variety may also score highly, as this depends on students making a series of informed decisions based on information supplied to them over the time of the exam.

2.9 Reliability

This would attract the minimum score of '1' for an assessment element where there are known to be problems regarding two or more people agreeing on the mark or grade for students' work (not least, essays!), but also in forms of assessment which may be more valid than essays, notably portfolios, but where different assessors can often come to quite different awards. Even dissertations fare quite badly regarding reliability, as different assessors often look for different things while assessing them, and their overall marks can be influenced

disproportionately by the presence or lack of such things. A high score for this line might be associated with (for example) a multiple-choice test or exam, where the scoring is no longer likely to be influenced by human frailty (though the question design might still be!).

2.10 Transparency

This is about the extent to which students know how the assessment works, and how exactly it is marked, and scores or grades are reached. If, as far as the students are concerned, it is a 'black box' assessment – they do their best, then find out if that was good enough – the assessment element probably merits the minimum score of '1' here. If it's something where students have had practice at marking examples you've given them, or better still their own or each other's work, the score may be a '5', if they feel they know exactly how the assessment works. It is important to distinguish between 'transparency' and 'familiarity'. For example, students can be quite familiar with traditional exams, while still not knowing exactly how they are marked. The same often goes for essays, dissertations and reports. 'Transparency' here is about how well students have a grasp of what will be going on in the minds of their assessors as they come to assessment judgements about their work.

2.11 Self-authenticity

This is in part about how well the assessment avoids plagiarism. Here, a traditional exam might score a '5', if precautions are sufficient to ensure that no-one can substitute for candidates. Similarly, a solo presentation or an oral exam (viva) may score '5' as an assessment element. Assessment formats where plagiarism is possible score much lower on this line, including essays, reports, dissertations, and other written work, where it is possible for students to copy other people's work, buy or download work from the internet, and so on. In rating your assessment element on 'self-authenticity', it is perhaps wisest to step back from any feelings of 'I'm sure none of my students would do this', or 'the anti-plagiarism software makes this highly unlikely', as the most skilled plagiarists are never caught! For this item, it is best to consider the *possibility* rather than the *probability* of plagiarism occurring when deciding your 1-5 score.

2.12 Real-world authenticity

This is about how well the assessment element links to the real-world professions students may be qualifying to enter. For example, doctors, lawyers, accountants and managers hardly spend their working lives *writing* about medicine, law, accountancy or management – they *do* it, not write about it. So essays are likely to score a '1' in disciplines such as these – and several others.

OSCEs (objective structured clinical exams) in medical education are likely to deserve a '5' here, if they are designed to be what doctors need to be able to show that they can do, not just write about. The practical part of a driving test would be high on real-world authenticity, whereas the 'theory' part which accompanies such a test in some countries is more of a memory test, and less well-linked to authenticity. (Do you still remember the stopping distance at 50 miles per hour on a wet road?)

2.13 Manageability – efficiency for students

This is essentially about the value of the time spent by students preparing for the assessment. You may need to ask a sample of your students about your particular assessment element to help you towards your 1-5 rating for this item. How much time do they know they *waste* in their preparations for this assessment? How do they see their time-efficiency relating to this item of assessment in the overall context of the bigger picture of their total assessment menu? Both of their answers to these questions need to be informed by how much this assessment element *counts* overall. The score for this item could be as low as ‘1’ if students feel it takes them forever to prepare for this assessment element, compared to others.

2.14 Manageability – efficiency for you

Whereas in many of the other items on the grid you may need to consult students to help you decide how assessment links to their experience of learning, this time you will know only too well how much time and energy the assessment element takes from you. Perhaps one factor to help you decide your 1-5 score for this item is how well you think the time you spend marking this element is well spent, considering the contribution of the element to students’ overall award. For example, if the element involves you in marking a large pile of essays or reports, but only contributes 5% or less to the overall award for students, the score will probably need to be a ‘1’! If it’s a computer-based multiple-choice exam for a large cohort of students, the design time might be very significant, but the marking is automated, and the score may be nearer a ‘5’.

2.15 Inclusiveness

This is a very complex issue. Its significance may depend a great deal on the composition of the student group, and to some extent on the size. Factors which may need to be considered here include:

How well the assessment provides a level playing field for:

- Students learning in a second language;
- The extent to which the assessment may disadvantage students with particular needs, such as dyslexia, visual impairments, hearing impairments, and so on.
- Students who for whatever reason are less successful than their optimum in this particular kind of assessment.

For this item, therefore, you will need to bear all manner of factors in mind when deciding your rating for ‘inclusiveness’ on the 1-5 scale. Sometimes you will have a very clear idea of how well the assessment concerned provides a level playing field for the particular student cohort, and at other times you may need to make judgements as best you can on the basis of what you know about the students.

2.16 Students get *and use* feedback as a result of it

The key words here are ‘and use feedback’; we all know how common it is for students to get feedback and fail to use it. For summative assessment elements, students often get little feedback (perhaps just a pass/fail award, or a score or a grade), and for this item the score

may only be '1'. Then there are the cases where students get quite a lot of feedback, but the feedback comes too late for them to put it to any real use – that too may warrant a '1' score. Or there may be cases where students don't seem to take any notice of the feedback, or don't even pick up their marked work containing feedback – that too could warrant a score of '1'.

Of course all students are different, and some may be making good use of the feedback they get, while others make much less use, so for the assessment element concerned you may need to consider an averaged score for this item.

Some forms of assessment are much richer in feedback than others (including student peer-assessment, and student presentations to an audience), and you will need to take this into account when working out the score for this item.

2.17 Alignment: how well it links visibly to learning outcomes

This links both to the perceived quality of the design of the assessment, and also to how well students have information about the targets they are meant to attain, as can be expressed through well-used intended learning outcomes. When curriculum is validated or reviewed, either internally or externally (for example by professional bodies), the alignment to learning outcomes is often required to be made more explicit.

If students are not aware of the intended learning outcomes, or don't realise that such outcomes reflect their attainment targets, the score for this item is likely to be '1'. If students are fully aware of the links, the score could be more towards '5'. The score you decide could be regarded as a measure of the 'constructive alignment' of the assessment element in the context of students' learning, and the design of the curriculum as a whole.

2.18 Students use several ways of communicating and explaining

This item (along with 19 and 20 below) refers to the overall picture of Assessment elements A to D (or more). A single assessment may well focus on just one method of students communicating their learning, for example in writing, or orally, or in group contexts, or online, and so on. And even 'in writing' can take many forms, for example unseen written exams, coursework essays and reports, written reviews, written reflections and so on. If all of the assessment elements use very similar ways of students communicating their learning, the score for this item might be as low as '1' for each of the elements involved. If overall there is a rich mix in how students communicate their learning, the score could be '5'.

2.19 Diversity: overall range of assessment types

This item also refers to the overall picture of the complete set of assessment elements for the course or module. Every assessment format disadvantages some students. Therefore, the more different forms of assessment making up the overall picture, the less likely that the same students are likely to be repeatedly disadvantaged by any one format. The score for this item needs also to be considered in terms of the extent to which any particular assessment format dominates the overall picture. For example, if a written exam counts for 80% or more of the overall award, the score for diversity should be as low as '1'. If there is a mix of four quite different forms of assessment, each counting for 25% overall, a score of '5' may be justified.

2.20 The 'wow' factor as gained from student feedback

This too refers to the overall picture of the assessment of the course or module, but this time in terms of students' *feelings* about the assessment element concerned. Of all the 'measures' in this grid, this is necessarily the most subjective one! However, it is linked to the value of the assessment as a positive driving force for students' learning, and links in its own way to the enhancement of their 'want' to learn the subject matter linked to the assessment element, but even more to the actual learning payoff they derive while preparing for, and then undertaking the assessment element. It is likely that only the occasional assessment element will in practice attract a 'wow' factor score from students, but when it does, it is important to recognise.

3. Conclusions

At six of my workshops in 2009, participants have been talked through each item of the grid in turn (and in different orders), and then invited to add up their scores for the assessment element they have chosen. Quite often, traditional forms of assessment have scored considerably lower (for example scores in the 20s) than more innovative forms which have sometimes scored as high as the mid 80s. Putting the grid to work with staff in higher education is showing that this attempt to interrogate assessment design can be a valuable prelude to working systematically towards 'assessment as learning' causes staff to reflect very deeply on the design of their assessment elements.

Linking assessment firmly to learning can be regarded as one of the most complex of our tasks in higher education. It is hoped that this step-by-step process of trying to analyse twenty separate aspects of assessment design, may contribute towards making assessment a more efficient and effective driver leading to better student learning.

4. References

Race, P., 2005. *Making Learning Happen: A Guide for Post-Compulsory Education*, Sage Publications Ltd. (Second edition commissioned to appear in 2010.)

5. About the author

Phil Race retired in June 2009 as Visiting Professor: Assessment, Learning and Teaching at Leeds Metropolitan University, but continues as Emeritus Professor there. He can be contacted through his website¹. He would like to acknowledge Royce Sadler, David Boud and Gordon Joughin, all from Australia, who visited Leeds Met in 2008, and with whom discussions of various aspects of assessment and feedback proved invaluable in shaping the thinking behind this paper.

1 <http://www.phil-race.co.uk/>

6. Appendix: Interrogating Assessment Elements: Scoring Grid

Your name:				
Course or module:				
Assessment element A:				
Assessment element B:				
Assessment element C:				
Assessment element D:				
Assessment element				
	A	B	C	D
Factors relating to the particular assessment element in isolation				
1. Students love it (wanting to learn)				
2. Students learn by doing preparing for it				
3. Students make sense of their learning preparing for it				
4. Students coach each other preparing for it				
5. Students practise making informed judgements				
6. Students design the criteria				
7. Students own the weighting of the criteria				
8. Validity				
9. Reliability				
10. Transparency				
11. Self-authenticity				
12. Real-world authenticity				
13. Manageability – efficiency for students				
14. Manageability – efficiency for you				
15. Inclusiveness				
16. Students get and use feedback as a result of it				
17. Alignment: how well it links visibly to learning outcomes				
Factors relating to the particular assessment element in the overall context				
18. Students use several ways of communicating and explaining				
19. Diversity: overall range of assessment types				
20. The 'wow' factor, as gained from student feedback				
Total/100				