

¿Why do we need to “repair” our assessment processes?

A discussion paper

¿Por qué necesitamos “reparar” nuestros procesos de evaluación?

Artículo de discusión

Phil Race

Senior Academic Staff Development Officer , University of Leeds, UK

e-mail: w.p.race@adm.leeds.ac.uk

Abstract

In the issue N 1, Vol. 4, 2003 (pp 50-51) of this Journal there was published a review of the interesting book by Phil Race “Designing Assessment to Improve Physical Sciences Learning” (Learning and Teaching Support Network, LTSN, University of Hull, 2001). In this article the author provides a provocative personal view of the problems of assessment of students within higher education in England. Many will find this controversial, but the issues raised are important in all countries and at all levels of education. Assessment of students is vitally important but it tends to dominate and to distort the whole educational process. In all systems one important objective for both teachers and students is that students should do *their* best in the assessments and hopefully gain appropriate qualifications. However, the objectives of, and competences developed through, learning and teaching must reach well beyond merely ‘passing the test’. Phil Race argues that the highly developed examination system in England is not sufficiently valid, reliable, transparent or authentic. A particular challenge is a response to the Special Educational Needs Discrimination Act (2002) that requires that the necessary discrimination between abilities by assessment should not *unfairly* discriminate against those with disabilities. There are no easy answers to the issues raised, however it is not just teaching and learning that need to be updated and continually reviewed, but teaching, learning *and* assessment. We all need to be involved in getting our assessment processes in good repair (fixing them) and keeping them ‘fixed’.

Key Words. Assessment, Learning, Science education.

Resumen

En el Volumen 4, Número 1, 2003 (pp 5-51) de esta revista, se publicó una reseña del interesante libro de Phil Race “Diseño de evaluaciones para mejorar el aprendizaje de las Ciencias Naturales” (Learning and Teaching Support Network, LTSN, University of Hull, 2001). En este artículo, el autor entrega un interesante punto de vista personal de los problemas de calificación de los estudiantes de educación superior en Inglaterra. Muchos lo encontrarán controversial, pero los problemas tratados son importantes en todos los países y en todos los campos de la educación. La evaluación de los estudiantes es de vital importancia, pero tiende a dominar y a distorsionar todo el proceso educativo. En todos los sistemas, un objetivo importante tanto para los profesores como para los estudiantes, es que estos últimos logren su mejor evaluación y por lo tanto su mejor calificación. Sin embargo, los objetivos del conocimiento, la enseñanza y las competencias desarrolladas a través del tiempo, deben ir más allá de merecer la “aprobación del test”. Phil Race argumenta que el muy desarrollado sistema de examen en Inglaterra no es suficientemente válido, confiable, transparente o auténtico. Un reto particular es la respuesta al acto de discriminación de las necesidades especiales en educación (2000), que requiere que la discriminación necesaria entre habilidades por evaluación, no debería discriminar injustamente contra las inhabilidades. No hay respuestas simples a estos problemas, sin embargo no solamente se trata de la necesidad de actualizar la enseñanza y el aprendizaje, se trata de enseñar, aprender y evaluar. Todos necesitamos estar involucrados en construir nuestro proceso de evaluación en favor de la buena calidad de evaluación y la enseñanza.

Palabras Clave. Evaluación, Aprendizaje, Educación en ciencias.

Context

“If it ain’t broke, don’t fix it”. “Don’t throw out the baby with the bathwater”. “You don’t fatten pigs by weighing them”. That’s almost enough of clichés! These were around long before assessment became as broken as it is today. But I will argue in this article that assessment is indeed broken in higher education nowadays, and needs fixing. In this article, I’d like to take you through some tough thinking about assessment, and encourage you to play your part in working towards fixing it. I’d like to challenge the status quo of assessment in higher education. In other publications I have tried to do this in the conventional scholarly manner, but now I think it is time to appeal to hearts and minds about the action, which needs to be taken, and not just to air intellectual reservations.

However, I would like to assert at the outset that the vast majority of assessors whom I know in higher education approach assessment with commendable professionalism, and bring to bear upon it all of the integrity, patience and care that they can. They spend a long time adjusting the wording of assessment tasks, and designing criteria with which to measure the evidence, which students deliver to them. Moreover, the decisions they make on the basis of this evidence are made carefully and painstakingly. Their good intentions are unbounded. But – the final cliché – the way to hell is paved with such intentions. Perhaps because assessors tend to grow gradually into the assessment culture of higher education, it is not surprising that they can be unaware of some of the prevailing problems that dominate the scene.

How is assessment broken?

Assessment should be valid, reliable and transparent. Anyone who cares about the quality of the assessment they design for students will say how they strive to make it so. We are also *required* to make

assessment valid, reliable and transparent by the Quality Assurance Agency. Most institutional teaching and learning strategies embrace these three qualities in the aspirations of universities. 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, that 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.

“Come on, Phil” you may argue. “We spend half of our lives on assessment. We have assessment boards, exam boards and external examiners approving our assessment instruments and practices and moderating our implementation of assessment. We’ve spent ages fine-tuning the assessment regulations. We’ve got years of experience at making assessment better. What more could we possibly be asked to do?”

“Assessment is the engine which drives student learning” (John Cowan). “And our feedback is the oil which can lubricate this engine” (Phil Race). But sometimes we’re too busy assessing to give really useful feedback. And students are too busy getting ready for their *next* assessment to take any notice of our feedback on their *previous* one. And when we come to the most important assessments (summative exams, and so on) feedback isn’t even on the agenda all too often. And what do we measure in these important assessments? ‘That which we can measure’ – not always what we *should* be trying to measure. It’s far easier to measure students’ achievement of relatively routine objectives, and much harder to measure their achievement of really important objectives. This led me to write over ten years ago ‘if you *can* measure it, it probably isn’t *it!*’.

“So it’s still broken.” I continue to argue. I’d better explain a bit more. Let’s go back to ‘valid, reliable, and transparent’ for a while. Let’s just clear up the meanings of these three little words.

Validity?

Valid assessment: this 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 students. 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 lectures) 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?” Ask students; 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 very hung up on measuring what students *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 students’ writing. What about those who aren’t good at writing? Or to be more critical, what about those students who have at least some measure of *disability* when it comes to writing?

The writing is on the wall for us regarding any tendency for our assessment instruments and processes to

discriminate against students with disabilities. ‘SENDA’ (Special Educational Needs Discrimination Act) is likely to cause us to have to make far-reaching changes to our assessment just to keep it within the law. This is a tricky one, as in one sense the purpose of assessment is to *discriminate* between students, and to find which students have mastered the syllabus best, and least, and so on. If we’re honestly discriminating in terms of ability, that might be legal. But if we’re discriminating in terms of disability it won’t be legal. 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 students aren’t discriminated against time and time again because they don’t happen to be skilled at those forms of assessment which 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 just beginning to work out the effects of dyscalculia, disgraphia, dyspraxia, and so on. Many of us are beginning to realise for the first time that in that packed lecture theatre, we do indeed have students with disabilities, not just the occasional student in a wheelchair, but perhaps a quarter or a third of our students 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 to be satisfied with the levels of validity to which we aspire?

So we’re not really in a position to be self-satisfied regarding the validity of even our most-used, and most practised assessment instruments and processes. But this isn’t new – we’ve used them forever it seems. That doesn’t make them more valid. But we’re 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?

Reliability?

For many, this word is synonymous with ‘fairness’ and ‘consistency’. This one is easier 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. Not just moderation, of course. Reliability 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? No we don’t, in many disciplines. There are some honourable exceptions. ‘Hard’ subjects such as areas of maths and science lend themselves to better measures of agreement 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. “But multiple marking just causes regression to the mean” can be the reply. “And after all, the purpose of assessment is to sort students 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”.

So why is reliability so important anyhow? Not least, because assessing students’ work is the single most important thing we ever do for them. Many lecturers in higher 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 lecturers it would help? And perhaps better, if we all regarded ourselves as researchers into assessment, alongside anything else we were researching into? “Students can escape bad teaching, but they can’t escape bad assessment,” says David Boud. Our assessments can end up with students getting first class degrees, or thirds. This affects the rest of their lives. Now if our assessment were really fair (reliable), we could sleep easily about who got firsts or thirds. The students who worked hardest would get better degrees, and the students who lazed around wouldn’t. This indeed is often the case, but most of us can

think of exceptions, where students got good degrees but didn't really deserve them, or students who seemed worthy of good degrees didn't come up with the goods, so we couldn't award them to them. So perhaps it's not just that our assessment isn't too reliable, it's our discrimination that's sometimes faulty too.

And transparency?

One way of putting 'transparency' is the extent to which students 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 by students, and also specify the forms in which students will present evidence of their achievement of the outcomes. There's a nice sense of closure matching up assessment criteria to intended learning outcomes. It's almost a shame that there's yet another problem: some of the *real* learning outcomes go beyond the intended learning outcomes. Patrick Smith (Buckinghamshire Chilterns University College) argues that these are the *emergent* learning outcomes. Some of them are unanticipated learning outcomes. And it could be further extrapolated that there is some tendency for the 'you know it when you see it' extra qualities which get the best students the best degrees are firmly embedded in their achievement of *emergent* learning outcomes, and their evidencing of these outcomes within our assessment frameworks.

Leave aside this additional factor, and go back to the links between intended outcomes and assessment criteria. How well do students themselves appreciate these links? How well, indeed, do assessors themselves consciously exercise their assessment-decision judgements to consolidate these links? Students 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 students during our teaching the exact nature of the associated assessment criteria. In other words, students often find it hard to get their heads inside our assessment culture – the very culture that will determine their degree classifications.

The students 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 students who manage this? Is this the ultimate assessment criterion? Is this the difference between a 1st and a 3rd? And is this the *real* learning outcome, the achievement of which we're measuring? And if so, is this stated transparently in the course handbook?

So, 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 – validity, reliability 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?

Time to fix assessment?

OK, there's a problem, but we've just not got enough time to fix it? *Why* haven't we got time to fix it? Because we're so busy doing, to the best of our ability, and with integrity and professionalism, the work which spins off from our existing patterns of assessment, so busy indeed that we haven't left ourselves time to face up to the weaknesses of what we're doing? Perhaps, we simply *dare not* face up to the possibility that we may be making such a mess of such an important area of our work? It can help to pause and reflect about just how we got into this mess in the first place.

A couple of decades ago, the proportion of the 18-21 year old population of the UK participating in higher education was in single figures, now it's approaching 40%, and Government waxes lyrical about increasing it to 50%. When there was only 5%, it could be argued that the average ability of those students who participated in higher education was higher, and they were better able to fend for themselves in the various assessment formats they experienced. Indeed, they usually got into higher education in the first place because they'd already shown to some extent that they'd got at least a vestigial mastery of the assessment culture. Now, there are far more students who haven't yet made it in understanding our assessment culture, let alone gearing themselves up to demonstrate their achievement within it.

At the same time, when we were busy assessing just a few per cent of the population, we had time to try to do it well, using the time-honoured traditional assessment devices at our disposal. Trying to do the same for five or ten times as many students is just not on. We can't do it. We can't do it well enough. We're assessing far too much to do it reliably, for a start.

And what about the students? Their lives are dominated by assessment. The intelligent response to this (thank goodness our students remain intelligent) is to become strategic. In other words, if there aren't any marks associated with some learning, strategic students will skip that bit of learning. If it counts, they'll do it. It's easy to go with the flow, and make everything important 'count' so that students will try to do all of it. But in the end this just leads to surface learning, quickly forgotten as the next instalment of assessment looms up. We're in danger of using assessment to *stop* learning instead of to start learning. It's no use us bemoaning the increased extent to which students have become strategic, when our assessment is the cause of this.

Who owns the problem of fixing assessment?

We can only ever really *solve* problems, which we own. But the assessment problem is so widely owned. It's dangerously easy to feel there's just nothing that we can do about it. It's easy enough to identify scapegoats, including:

- Professional bodies, in whose name we feel we need to stick to the status quo;
- Pre-university education systems, which cast the die and train pupils into particular expectations of learning and assessment;
- Institutional, faculty and departmental assessment regulations, which limit our room for manoeuvre;
- Teaching and learning strategies, which are so all-encompassing that we can't suspend belief

and start afresh again;

- Heads of department or school, who are often seen (sometimes seen wrongly) to be content with the status quo;
- External examiners who would have to be convinced when radical changes may need to be made;
- Students themselves who could or would complain about rapid changes to the level of the playing field or the position of the goalposts (even if the whole is enveloped in thick fog at present);
- The world outside academe, where there's a view about what a graduate should be, and so on;
- Journalists, broadcasters and editors who would give us a hard time if anything were to be found wrong in the way we did the job we are paid to do;
- Politicians and policy-makers who got to where they are by succeeding in the system of assessment we already have, and dare not admit that it might have been flawed;
- Parents, employers, taxpayers and others who foot the bill for higher education.

However, if we're perfectly frank about it, each assessment judgement is almost always initially made in the mind of one assessor in the first instance. True, it may well then be tempered by comparisons with judgements made in other people's minds, but to a large extent assessment remains dominated by single acts of decision-making in single minds, just as the evidence which is assessed is usually that arising from the product of a single mind at a given time within a given brief. Living on a crowded planet may be a collaborative game, but we tend to play the assessment game in predominantly singular circumstances, and competitive ones at that.

The fact of the matter is that to fix assessment will require individuals to change what they do, but that won't be enough to change the culture. Teams of individuals with a shared realisation of the problem will need to be the first step.

How can we fix assessment?

We need to work out a strategy. But any strategy has to be made up of a suitably chosen array of tactics. Sometimes it's easier to start thinking of the tactics first. What could be a shopping list of tactics to play with for starters in this mission. They include:

- Getting students into our assessment culture, by using peer-assessment and self-assessment more, so that they are better within our culture when we assess them;
- Reducing the quantity of assessment (say by a factor of three) so that we have time to do it well, and students have time for their learning not to be completely driven by assessment;
- Increasing the quality of assessment, so that it is fit for purpose, and more valid, more reliable and more transparent;

- Increasing the diversity of assessment instruments and processes, so that student casualties (where particular students are discriminated against repeatedly by the same old assessment formats) are not so frequent;
- Training (yes, training, not educating) our students to be better-able to play the game of working out where the goalposts are, and practising how to demonstrate their achievement of our intended learning outcomes.

But turning such tactics into a strategy is a big job, and beyond the scope of a short provocative article such as this. However, that big job won't even get started unless people are convinced that it needs to be done, and that was the purpose of this article. My aim was not on this occasion to write a scholarly article repeating what wise people have already written about in the literature (for years and years now). My intention was to employ challenging language to convince you that you've got a problem. What are you going to do about it?

Publications on assessment and feedback, for lecturers and for students

Race, P (1992) *500 Tips for Students* Blackwell, Oxford. (Russian edition: 1996).

Race, P (1993) *Never Mind the Teaching – Feel the Learning* SEDA Paper 80: SEDA Publications, Birmingham.

Brown, S, Race, P and Smith, B (1996) *500 Tips on Assessment* Kogan Page, London.

Race, P and McDowell, S (1996, 2nd edition 2000) *500 Computing Tips for Lecturers and Teachers* Kogan Page, London.

Brown, S and Race, P (1997) *Educational Development in Action* SEDA Paper 100, SEDA Publications, Birmingham, UK.

Race, P (1998) *Changing Assessment to Improve Chemistry Learning* FDTL Improve Chemistry Project, University of Hull and RSC.

Race, P (1999) *How to get a Good Degree* Open University Press, Buckingham.

Race, P (ed.) (1999) *2000 Tips for Lecturers* Kogan Page, London.

Race, P (1999) *Enhancing Student Learning* SEDA Special 10, SEDA Publications, Birmingham.

Brown, S, Race, P and Bull, J (eds.) (1999) *Computer Assisted Learning in Higher Education* Kogan Page, London.

Race, P (2000) *How to Win as a Final Year Student* Open University Press, Buckingham.

Race, P (2001) *The Lecturer's Toolkit: 2nd Edition* Kogan Page, London.

Race, P and Brown S (2001) *The ILTA Guide* Institute for Learning and Teaching, York, UK and available online at www.Education.Guardian.co.uk by searching for 'ILTA'

Race, P (2001) *Self, Peer and Group Assessment* LTSN Generic Centre publications.

Race, P (2001) *Students' Guide to Assessment* LTSN Generic Centre publications.

About Phil Race

BSc, PhD, FCIPD, PGCE, ILTM

My original training was as a scientist, but over the years I became progressively more interested in teaching, learning and assessment, and gradually became an educational developer. My principal interests span assessment design, lecturing, small-group teaching, group learning, open learning, study-skills development, learning resource design and trainer-training. I am particularly keen to de-mystify these areas, and to get ideas across without recourse to some of the complex jargon too often encountered in the literature in these fields.

For 24 years, I was at the University of Glamorgan in Wales, where I started as a lecturer in physical chemistry and ended up as Professor of Educational Development, and was granted Emeritus Professor status there when I took early retirement in 1995. I now work two days per week in the Staff and Departmental Development Unit at the University of Leeds, and for the rest of my time I run training workshops for staff and students in universities, colleges and other organisations throughout the UK, and giving keynotes and workshops at conferences on teaching and learning. I also work abroad, and have visited Canada, Australia, New Zealand, Ireland, Czech Republic, Slovakia, Denmark, Holland, Ukraine, Hungary, Greece, Israel, Sweden and Singapore in recent years.

My mission is to improve and enhance the quality of students' learning, by helping teaching staff to develop their methods and approaches, and by helping students to develop their own learning skills. I am also particularly interested in the design of assessment instruments and processes, and I am keen that both assessment and feedback should play positive and motivating roles in student learning. I design and lead highly interactive training workshops for staff in Further and Higher Education, and in commerce and industry. I am a Member of the Institute for Learning and Teaching, and also an ILT Accreditor, and a Fellow of the Chartered Institute of Personnel and Development.

I can be contacted through my website at www.Phil-Race.net or emailed at w.p.race@adm.leeds.ac.uk