**Towards a scholarship of assessment – an opinion piece**

Chris Rust
Oxford Brookes University

*Abstract*
*This conceptual paper argues that given the centrality of assessment to the teaching and learning process, the scholarship of teaching and learning needs to develop a scholarship of assessment. Evidenced by the outcomes of QAA institutional audits, and the National Student Satisfaction survey, it is also argued that it is probably the aspect of teaching and learning most in need of improvement. Building on previous work (Rust, 2002; Gibbs & Simpson, 2002; Rust et al, 2005) the paper attempts to identify what that scholarship should include. Further, it argues why much current practice in the use of marks and arrival at degree classification decisions is intellectually and morally indefensible, statistically invalid, and must be firmly rejected if assessment is to be a scholarly activity.*

## Introduction

As part of the Teaching Quality Enhancement Fund (TQEF) initiative started in 1999, Universities in England were required to write learning and teaching strategies. The result was that most did just that, making no reference at all to assessment. In order to address this, over time, most strategies were explicitly renamed learning, teaching and assessment (LTA) strategies, and included sections on assessment. When Boyer argued (1990, p16) there were "four separate, yet overlapping functions" in the work of an academic, he identified them as the scholarships of discovery, integration, application and teaching. I am concerned that there is a similar danger that, as with those early TQEF strategies, because it does not have assessment in the name, the scholarship of teaching and learning (SoTL) community and movement might make the same oversight. Given what we know about the central importance of assessment, and its power to effect student learning, for good or bad, (see Ramsden, 1992; Gibbs, 1992; Boud, 1995; Brown et al., 1997; Rust, 2002; et al), I want to add to the argument already started by people such as Holroyd (2000) and Stefani (Stefani, 1998; Stefani, 2004-5) that it is vital that we explicitly articulate and establish a scholarship of assessment which should be at the very heart of our scholarship of teaching and learning.

There is a further reason why we should do this, namely, that while there undoubtedly may be room for improvement in many teaching and learning practices, as Boud said (Ibid, p35), while "Students can, with difficulty, escape from the effects of poor teaching, they cannot (by definition, if they want to graduate) escape the effects of poor assessment." And assessment seems to be the part of the learning and teaching process that has been especially criticised. "In the UK, assessment practices have consistently been one of the weakest features identified by the Quality Assurance Agency in subject reviews across the disciplines. Knight has described assessment as 'the Achilles' heel of quality' (2002a, p. 107) and accused summative assessment practices of being 'in disarray' (2002b, p. 275)" (Rust et al, 2005, p231) while Race (2003, p. 5) has described our practice simply as "broken". Yorke (2000, p7) has argued that "there is considerable

scope for professional development in the area of assessment" but in order for there to be such development, there will need to be a clearly articulated scholarship to support it.

**So what should this scholarship of assessment include?**

In fact, there is a considerable amount that we already know about assessment to be distilled from the learning and teaching literature, as I have argued previously (Rust, 2002), that easily lead to the following good-practice principles, namely, that all courses should:

- Be required to be designed according to the principles of constructive alignment, namely that there should be clear and explicit linkage (alignment) "between teaching method and assessment to the learning activities stated in the objectives so that all aspects of this system are in accord in supporting appropriate student learning" (Biggs, 1999, p11). Or as the Quality Assurance Agency succinctly recommend, there should be "effective and appropriate measurement of the achievement by students of the intended learning outcomes" (QAA, General principle 6).
- Ensure that the workload is realistic and the assessment is non-threatening and non-anxiety provoking (Gibbs, 1992 et al).
- Engender intrinsic motivation through relevant 'real world' assessment tasks, tasks which require active engagement by the student, and by providing a choice of tasks (e.g. Brown et al, 1994)
- Pace student learning and ensure there are sufficient formative tasks (Brown et al, Ibid).
- Structure skills development (e.g. Gibbs, 1981).
- Allow for 'slow' learning and early failure (Yorke, 2001).
- Include explicit guidelines on giving effective and prompt feedback (QAA, code of practice).

The power of assessment to influence student approaches to learning has already been cited. Any scholarship of assessment must therefore be predicated on the value that good assessment supports and positively influences student learning. In their meta-analysis of the research literature, Gibbs and Simpson (2002) identify the following eleven conditions (inexplicably reduced to ten in a later publication, 2004-5) under which assessment has been established as supporting learning:

Deleted: 4-5

1. Sufficient assessed tasks are provided for students to capture sufficient study time
2. These tasks are engaged with by students, orienting them to allocate appropriate amounts of time and effort to the most important aspects of the course
3. Tackling the assessed task engages students in productive learning activity of an appropriate kind

4. Assessment communicates clear and high expectations

5. Sufficient feedback is provided, both often enough and in enough detail

6. The feedback focuses on students' performance, on their learning and on actions under the students' control, rather than on the students themselves and on their characteristics

7. The feedback is timely in that it is received by students while it still matters to them and in time for them to pay attention to further learning or receive further assistance

8. Feedback is appropriate to the purpose of the assignment and to its criteria for success.

9. Feedback is appropriate, in relation to students' understanding of what they are supposed to be doing.

10. Feedback is received and attended to.

11. Feedback is acted upon by the student

Although there is obviously no suggestion that these are equally weighted in terms of importance, nevertheless it is interesting to note that while 1, 2 & 4 are essentially about motivation, and 3 is about creating a learning opportunity through an appropriate activity, the other seven are all about aspects of feedback. And if the literature suggests we are bad at assessment generally, the evidence is that it is in the area of feedback that we are possibly worst of all. It was criticised in more than 10% of QAA audit reports - "More than one in ten of the institutional audit reports …. recommended to institutions the advisability or desirability of ensuring that feedback is provided to students in a consistent and timely fashion" (QAA, 2006, p12). Problems with feedback have also been clearly highlighted in the National Student Satisfaction survey (see http://www1.tqi.ac.uk/sites/tqi/home/index.cfm)

Deleted: are

Deleted:

In terms of guidance on good feedback practice, a useful and slightly different addition to Gibbs and Simpson's conditions has recently been suggested by Nicol and Macfarlane-Dick (2006) in the form of the following seven principles:

Good feedback practice:

1. helps clarify what good performance is (goals, criteria, expected standards)

2. facilitates the development of reflection and self-assessment in learning

Formatted: Bullets and Numbering

Formatted: Bullets and Numbering

3. delivers high-quality information to students about their learning

4. encourages teacher and peer dialogue around learning

5. encourages positive motivational beliefs and self-esteem

6. provides opportunities to close the gap between current and desired performance

7. provides information to teachers that can be used to help shape the teaching

Formatted: Bullets and Numbering

A further contribution to this emerging scholarship of assessment was made by myself and colleagues (Rust et al, 2005), with a consideration of the implications for practice, if we were to consistently apply a social-constructivist view of learning to our assessment processes. The result can be summarised in the following diagram, depicting two interrelated and dynamic cycles of desirable practice, one for staff and one for students.

*Diagram here*

As Clegg and Bryan (2006a, p225) have pointed out, there is also "a growing evidence base of innovative practice" (Schwartz and Webb, 2002; Clegg and Bryan, 2006b), and they suggest a new assessment culture is emerging, defined by the following characteristics:

Deleted: So while

- Active participation in authentic, real-life tasks that require the application of existing knowledge and skills
- Participation in a dialogue and conversation between learners (including tutors)
- Engagement with and development of criteria and self-regulation of one's own work
- Employment of a range of diverse assessment modes and methods adapted from different subject disciplines
- Opportunity to develop and apply attributes such as reflection, resilience, resourcefulness and professional judgement and conduct in relation to problems
- Acceptance of a limitation of judgement and the value of dialogue in developing new ways of working

Formatted: Bullets and Numbering

So while, across the sector, there may frequently still be a significant gap between theory and practice, a start has clearly been made in identifying a scholarship of assessment, and

there is also a growing body of case study evidence of effective innovation and changes in practice being made, which is informed by that scholarship.

Deleted: assessment.

But the area of assessment where I would argue there is some of the worst practice, seems to continue largely unchallenged. Another underpinning value for the scholarship of assessment must surely be that of equity – as far as is humanly possible, the results and outcomes need to be fair. Yet much current practice in the use of marks and the arrival at degree classification decisions is not only unfair, but is intellectually and morally indefensible, and statistically invalid – despite the often disproportionate amount of staff time that is spent on so called quality assurance procedures to do with assessment. Let us consider just some of these bad practices.

1. Belief that it is possible to distinguish the quality of work to a precision of one percentage point

Although the reality of using percentages for much marking of student work does not actually mean the use of a one hundred point scale, because students rarely are given more than 70 or less than 35 (with some disciplinary differences which we will return to below), so it tends to be roughly a 35 point scale, this marking still implicitly suggests that it is possible to distinguish between individual pieces of work to a precision of one thirty-fifth of difference. And of course in doing this there will be numerous aggregations having to take place between how well different learning outcomes and assessment criteria have been met. Theories of judgement analysis would suggest, as Elander and Hardman have pointed out, citing Einhorn (2000), that this is just not possible. “It is the integration of information about multiple cues that research has shown human experts to have the most difficulty with (2002, p 304). “People are bad at integrating information” (Dawes, 1982, p 395).

2. Belief that double-marking will ensure fairness and reliability

Just because two markers arrive at the same or a similar mark does not mean that the system is reliable. It is quite possible that they have reached the mark for significantly different reasons. And where double-markers disagree, depending on the hierarchical and power relationship between them, the resolution may have little or nothing to do with the objective merits of the piece of work. The senior member of staff’s view may simply override the other’s, or in other cases, just because it is easier and saves time, a simple average between the two may be chosen.

3. The fact that most marks lack meaning unless they are stated in terms of norms, group summaries (the mean or median) or the objectives mastered

This is true from the question of, “What does the fact a student got 54% for a particular piece of work actually mean?” all the way up to the question of “What does an upper second degree classification tell anyone about a graduate from a particular course?” In isolation, neither piece of data conveys any real meaning either to another tutor, or to an employer, about the strengths and weaknesses, knowledge and skills of the student.

4. The combination of scores, which obscures the different types of learning outcome represented by the separate scores

Let us consider a module where there may be a piece of coursework explicitly designed to test the application of one aspect of theory in depth, and an exam designed to assess primarily a breadth of knowledge gained. When the two results from these assessments are simply turned into numbers and combined, the detail of what has been assessed is completely lost.

5. The combination of scores where the variation (standard deviation) for each component is different

This would be unacceptable in the practice of a first year statistics student, but university assessment systems do this all the time, both within modules, and in combining the total marks from different modules or units of study.

6. The distortion of marks by the type of assessment (e.g. coursework c.f. examination) and the actual subject discipline/s studied

It is well known in the literature that students are more likely to score highly on coursework rather examinations (Yorke et al, 2000; Bridges et al, 2002), and also on certain modules where things like fieldwork or individual projects are involved and motivation is high. It is also well established in the literature that marks will vary simply depending on the discipline being assessed, with much higher marks likely to be found in mathematics and statistics, for example, than in a subject like English (Yorke et al, 1997). But in modular degree programmes, where different subjects may well be studied in combination, marks are still likely to be just added together despite these differences. And where single disciplines are studied there is evident distortion in the resulting degree classification achieved. As Yorke et al point out (2000) from HESA data from 1999, 21.1% of Mathematical Science graduates get first but only 3.7% in Law. Why has there not been a national outcry at this blatant inequity?!

7. The distortion of resulting degree classifications by the application of idiosyncratic institutional rules

Several studies (Woolf and Turner, 1997; Armstrong et al, 1998) have also pointed out that the application of different institutional rules on how marks are combined, etc. can make considerable differences to the final degree classification obtained. With the same module results, different degree classifications could be obtained simply depending on which institutions rules are applied. In 2000, Yorke et al (p 230) said “there is a need for a deep inquiry into the fundamental nature of degree award algorithms, and a study of percentage-scale marking and grading,” but to my knowledge there has been no such inquiry. The Quality Assurance Agency (QAA) has just revised its code of practice on assessment, slightly weakening them if anything, in my opinion, and these issues are not even mentioned.

**Summary**

There is already a growing literature on assessment and an emergent scholarship that we in the SoTL community need to both promote and to build on if assessment is to be a scholarly activity. Awareness alone, however, is unlikely to be enough. The poor practice highlighted above is not simply explained by lack of knowledge. For whatever reason – conservatism, lack of time, lack of incentive, lack of interest – we know there is generally considerable reluctance to change learning, teaching and assessment practices. For that reason, if a scholarship of assessment is to be more than just espoused theory, in order to ensure assessment practices are scholarly, we need to campaign vigorously, and at every opportunity, especially with policy and decision-makers at every level. As Clegg and Bryan implore, we must join in a "search for excellence and … take the moral high ground, stand up for what research tells us is right and commit to better assessment." (2000a, p225) And we might do best, perhaps, to start with the two major issues of improving feedback and equity.

**Note**
An earlier version of this paper was presented at the Scholarship of Teaching and Learning (SOTL) 6th International Conference, London, UK, May, 2006

**Bibliography**


Armstrong, M. Clarkson, P. & Noble, M. (1998) *Modularity and credit frameworks: the NUCCAT survey and 1998 conference report,* (Newcastle-upon-Tyne, Northern Universities Consortium for Credit accumulation and Transfer)

Biggs, J. (1999) *Teaching for quality learning at university* (Buckingham, SRHE & Open University Press).

Boud, D. (1995) Assessment and learning: Contradictory or complementary? In P. Knight (Ed*) Assessment for Learning in Higher Education*, 35-48. (London, Kogan Page)

Boyer, E.L. (1990) *Scholarship reconsidered: priorities of the professoriate.* (Princeton, The Carnegie Foundation for the Advancement of Teaching)

Bridges, P. et al (2002) Coursework marks high, examination marks low: discuss. *Assessment and Evaluation in Higher Education*, 27 (1), 35-48

Brown, G., Bull, J. & Pendlebury, M. (1997) *Assessing student learning in higher education* (London, Routledge).

Brown, S., Rust. C. & Gibbs, G. (1994) *Strategies for diversifying assessment in higher education, Oxford*, Oxford Centre for Staff and Learning Development

Bryan, C & Clegg, K (2006a) Reflections, rationales and realities, in Bryan, C & Clegg, K (Eds) *Innovative Assessment in Higher Education*, 216-227, (Abingdon, Routledge)

Bryan, C & Clegg, K (Eds) (2006b) *Innovative Assessment in Higher Education*, (Abingdon, Routledge)

Dawes, R.M. (1982) The robust beauty of improper linear models in decision making, in D. Kahneman, P. Slovic & A. Tversky (Eds), *Judgement under uncertainty: heuristics and biases*, 331-407 (Cambridge, Cambridge University Press)

Einhorn, H.J. (2000) expert judgement: some necessary conditions and an example, in T. Connolly, H.R. Arkes & K.R. Hammond (Eds), *Judgement and decision making; an interdisciplinary reader* (2nd Ed), 324-335 (Cambridge, Cambridge University Press)

Elander, J. & Hardman, D. (2002) An application of judgment analysis to examination marking in psychology, *British Journal of Psychology*, 93, 303-328

Gibbs, G. (1981) *Teaching Students to Learn.* Milton Keynes, Open University Press

Gibbs, G. (1992) *Improving the quality of student learning*, Bristol, TES

Gibbs, G. & Simpson, C. (2002) *Does your assessment support your students' learning* available at: http://www.brookes.ac.uk/services/ocsd/1_ocsld/lunchtime_gibbs.html (accessed November 2002)

Gibbs, G. & Simpson, C. (2004-05). Conditions under which assessment supports students' learning, *Learning and Teaching in Higher Education*, 1, 3-31.

Holroyd, C. (2000) Are assessors professional? Student assessment and the professionalism of academics, *Active Learning in Higher Education*, 1 (1) 28-44

Knight, P. T. (2002a) The Achilles' heel of quality: the assessment of student learning, *Quality in Higher Education*, 8(1), 107–115.

Knight, P. T. (2002b) Summative assessment in higher education: practices in disarray, *Studies in Higher Education*, 27(3), 275–286.

National Student Satisfaction Survey, available at: http://www1.tqi.ac.uk/sites/tqi/home/index.cfm (accessed March, 2006)

Nicol, D. & Macfarlane-Dick, D. (2006) Formative assessment and self-regulated learning: a model and seven principles of good feedback, *Studies in Higher Education* 31 (2) 199-218

Quality Assurance Agency, *Code of practice: assessment of students general principles – quality and standards In HE*, http://www.qaa.ac.uk/public/cop/copaosfinal/genprin.htm (accessed November 2001)

Quality Assurance Agency (2006) *Outcomes from institutional audit: assessment of students*, (Mansfield, QAA/Linney Direct)

Race, P. (2003) Why fix assessment?, in: L. Cooke & P. Smith (Eds*) Seminar: reflections on learning and teaching in higher education* (High Wycombe, Buckinghamshire Chilterns University College).

Ramsden, P. (1992) *Learning to teach in higher education* (London, Routledge).

Rust, C. (2002) The impact of assessment on student learning-how can the research literature practically help to inform the development of departmental assessment strategies and learner-centred assessment practices, *Active Learning in Higher Education*, 3(2), 145–158.

Rust, C., O'Donovan, B. & Price, M. (2005) A social constructivist assessment process model: how the research literature shows us this could be best practice, *Assessment and Evaluation in Higher Education*, 30 (3), 231-240

Schwartz, P & Webb, G. (2002) *Assessment case studies, experience and practice from higher education* (London, Kogan Page)

Stefani, L.A.J. (1998) Assessment in partnership with learners, *Assessment and Evaluation in Higher Education*, 23 (4) 339-350

Stefani, L.A.J. (2004-5) Assessment of student learning: promoting a scholarly approach, *Learning and Teaching in Higher Education*, 1 51-66

Woolf, H., & Turner, D. (1997) Honours classifications: the need for transparency, *The New Academic*, Autumn, 10-12

Yorke, M. (1997) Module mark distribution in eight subject areas and some issues they raise, in N. Jackson (Ed), *Modular higher education in the UK*, 105-107 (London, Higher Education Quality Council)

Yorke, M., Bridges, P. & Woolf, H. (2000) Mark distributions and marking practices in UK higher education; some challenging issues, *Active Learning in Higher Education*, 1 (1), 7-27

Yorke, M. (2001) Address to ILT symposium *Widening participation and promoting student retention*, London, 27th September