

Memorandum

Submission from ASKe to the Innovation, Universities, Science and Skills Select Committee inquiry into “Students and Universities”.

Rust, C., Carroll, J., Handley, K., O’Donovan, B., & Price, M.

1.0 Executive summary

ASKe believes that if the UK degree is to continue to be seen as an academic benchmark standard, and is to maintain its reputation internationally, and with employers, parents and students, there is a need for a complete ‘root and branch’ change in assessment processes and practices. In addition to providing evidence for this assertion, ASKe makes a number of practical recommendations as to how the sector could start to bring this change about.

2.0 ASKe

2.1 ASKe is the Centre for Excellence in Teaching and Learning based at Oxford Brookes University Business School. It was set up in summer 2005 with a £4.5 million award (spread over five years) from HEFCE in recognition of good practice based on pedagogic research into aspects of assessment carried out by staff in the Business School and the Oxford Centre for Staff and Learning Development. ASKe's work focuses on ways of helping staff and students develop a common understanding of academic standards, and it builds on and promulgates established good practice. Last year we funded the bringing together of 40 national and international experts in assessment which met in November 2007 and has become known as the Weston Manor Group. The outcome of that group’s discussions was the production of a six tenet manifesto for change to assessment practice related to standards (attached).

2.2 This response is focused solely on the section of questions regarding degree classification.

3.0 Information for the committee

3.1 We would argue that there are numerous and significant methodological flaws in current assessment practice at both the macro level of degree classification, and at the micro level of the assessment of individual students, which means that there should be growing concern about the integrity of the degree as a qualification and what it means to be a graduate.

3.1.1 This has effectively been accepted by both the Burgess report’s

Memorandum

(2007) admission that the degree classification system is “no longer fit for purpose” (p5) and the QAA’s admission that currently:

“ (a) it cannot be assumed students graduating with the same classified degree from *different* institutions, having studied different subjects, will have achieved similar standards; (b) it cannot be assumed students graduating with the same classified degree from a *particular* institution, having studied different subjects, will have achieved similar standards; and (c) it cannot be assumed students graduating with the same classified degree from different institutions, having studied the *same* subject, will have achieved similar standards.” (QAA, 2006, emphasis added)

3.1.2 The learning outcomes of a degree are complex and address a range of cognitive and practical skills but “A potential employer wants to know one thing: is a degree from the university of X creditable? If so, how does it compare with one from the university of Y? Yet these are questions the QAA cannot answer” (Kealey, 2008).

3.2 And before these recent admissions, we already knew from a number of studies that traditional reliance on the external examiner system to mediate standards within the system was misplaced (e.g Newstead and Dennis, 1994).

3.3 Let us consider some of the major questionable beliefs and bad practices in the system (Rust, 2007):

3.3.1 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).

3.3.2 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

Memorandum

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.

3.3.3 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 the student, another tutor, or to an employer, about the strengths and weaknesses, knowledge and skills of the student.

3.3.4 The practice of combining scores, which obscures the different types of learning outcomes 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.

3.3.5 The practice of combining 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.

3.3.6 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 than examinations (Yorke et al, 2000; Bridges et al, 2002). 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

Memorandum

Yorke et al point out (2000) from HESA data from 1999, 21.1% of Mathematical Science graduates get firsts but only 3.7% in Law.

3.3.7 The distortion of generating 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 institution's 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 there has been no such inquiry.

3.4 We would also argue that, ironically, a number of these bad practices have been as a result of a (failed) desire to create greater reliability which has been at the cost to the other, vitally important function of assessment – formative feedback and the improvement of learning.

3.4.1 "The types of assessment we currently use do not promote conceptual understanding and do not encourage a deep approach to learning.....Our means of assessing [students] seems to do little to encourage them to adopt anything other than a strategic or mechanical approach to their studies." (Newstead 2002, p3)

3.4.2 "Many research findings indicate a declining use of deep and contextual approaches to study as students progress through their degree programmes". (Watkins & Hattie, 1985; Kember et al, 1997; Richardson, 2000; Zhang & Watkins, 2001)

3.4.3 "This quest for reliability tends to skew assessment towards the assessment of simple and unambiguous achievements, and considerations of cost add to the skew away from judgements of complex learning". (Knight 2002 p278)

3.5 As for advantages and disadvantages of the classification system itself, it is arcane and peculiar to the UK, and to undergraduate courses. Even on postgraduate UK courses the differentiation of outcomes is much simpler and easier to understand being, either, just Pass or Distinction, or Pass, Merit, or Distinction. The planned introduction of the Higher Education Academic Record (HEAR) is a welcome development given that it should set student achievement in the context of the mission and values of the course studied which will vary in relation to the academic, employability or professional focus. However, in addition to the problems of educating employers to HEAR's usefulness, we need to ensure that it is useful. And this will require the recording of much more than

Memorandum

simply the almost meaningless marks or grades achieved by a student on individual modules.

3.6 Regarding plagiarism, it is a problem; and the concern about student plagiarism is an even greater problem. There is evidence to show it is rising and, in particular, that deliberate attempts to deceive assessors are rising sharply from a relatively low base of (a generally agreed assumed level of) 10 - 15 cases per 1000 submissions. Statistics about levels of plagiarism are contradictory and hard to evaluate as they ask very different questions of different groups of students. Surveys that show "almost all students cheat" are frequent but irrelevant since they usually refer to one-off or pragmatic decisions with little or no impact on students' overall skills /learning or on the credibility of their final award. Cheating and plagiarism does not threaten important graduate skills which are tested in other ways such as nurses knowing how to care for patients or engineers knowing how to build bridges. There is much useless scaremongering in this area, implying that UK graduates are not reliably assessed on discipline specific skills.

3.6.1 The opportunities for plagiarism have risen exponentially since 2003, both in terms of available internet resources and via bespoke writing 'services' (sic). It is estimated that the latter are available via more than 250 sites in the UK alone. In 2005, the Guardian stated such 'services' attracted spending of more than 200 million pounds per year. These opportunities and evidence of their use do now present a threat to generic, coursework-assessed courses. Copying and faking work is likely to be a regular practice in large, generic courses in some disciplines. Business, Computing, and Law are most often mentioned though concern in all disciplines is widespread. In some cases, studies show up to 50% of students say they submit others' work, at least for some of the assessment, in large, generic courses assessed by coursework.

3.6.2 There is a significant issue of plagiarism with students who lack sufficient skill to succeed, including but not exclusively international students (IS). ISs are over-represented in institution's punishment statistics because they are much more likely to be identified as plagiarists, both because of change in language and because of the way in which text-matching software works.

3.6.3 Text matching software can help to identify work that warrants extra attention by markers but will not solve the problem as plagiarism is a pedagogic issue requiring an integrated pedagogic response. All universities should use text-matching software as an adjunct to other measures.

3.6.4 Simplistic reactions to the problems of plagiarism, like a retreat to exams or reliance on technology are not the solution. Addressing plagiarism is well within the capacity of university pedagogic and administrative processes and there are examples of it being handled with creativity and good effect across the UK. There are also many examples of universities who have yet to

Memorandum

address the issue systematically and, in those cases, a significant issue remains.

3.7 The fundamental premise, on which our recommendations for change are based, is that “meaningful understanding of standards requires both tacit and explicit knowledge” (O’Donovan et al. 2004) and while the provision of explicit knowledge has been addressed through learning outcomes, benchmarks and assessment criteria, the role of tacit knowledge is largely ignored because “tacit knowledge is experience-based and can only be revealed through the sharing of experience – socialisation processes involving observation, imitation and practice” (Nonaka, 1991). To establish standards at both local and national levels therefore also requires the implementation of such processes, both nationally and locally, for both staff and students.

4.0 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

Burgess Group (2007) *Beyond the honours degree classification: The Burgess Group final report*, London: Universities UK

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*, Cambridge: Cambridge University Press, 331-407

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), Cambridge: Cambridge University Press, 324-335

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

Kealey, T. (2008) Degrees won’t be trusted until regulation changes, *Education Guardian*, 11th November, p10

Kember, D., Charlesworth, M., Dabies, H., MacKay, J., & Stott, V. (1997). Evaluating the effectiveness of educational innovations: using the study process questionnaire to show that meaningful learning occurs, *Studies in Educational Evaluation*, 23 (2), 141-157.

Memorandum

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

Newstead, S. (2002) Examining the examiners: why are we so bad at assessing students? *Psychology Learning and Teaching*, 2 (2), 70-75

Newstead, S.E. & Dennis, I. (1994) Examiners examined: the reliability of exam marking in psychology, *The Psychologist*, 7, 216-219.

Nonaka, I. (1991) The Knowledge-Creating Company, *The Harvard Business Review*, November-December, 96-104.

O'Donovan, B., Price, M. & Rust, C. (2004) Know what I mean? Enhancing student understanding of assessment standards and criteria, *Teaching in Higher Education* 9, 325-335

Quality Assurance Agency (2006) *Background briefing note: - The classification of degree awards*, London: QAA, available at: http://www.qaa.ac.uk/education/briefings/classification_20Sept06.asp (accessed 13th November, 2008)

Richardson, J. T. E. (2000). *Researching Student Learning: Approaches to Studying in Campus-Based and Distance Education*, Buckingham, U.K.: Society for Research into Higher Education & Open University Press.

Rust, C. (2007) "Towards a scholarship of assessment" *Assessment and Evaluation in Higher Education*, Vol. 32, No. 2, 229-237

Watkins, D., & Hattie, J. (1985). A longitudinal study of the approaches to learning of Australian tertiary students, *Human Learning*, 4, 127-141

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*, London: Higher Education Quality Council, 105-107

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

Zhang, L. F. & Watkins, D. (2001). Cognitive development and student approaches to learning: an investigation of Perry's theory with Chinese and US university students, *Higher Education*, 41, 236-261.

Memorandum

5.0 Recommendations for inclusion in the committee's report

5.1 To establish national standards in any given discipline requires the establishment of a disciplinary community of assessment practice across the sector. This requires bringing together members of the discipline from different institutions to compare the quality of their students' work and their marking judgements. [Much could be achieved by emulating the assessment practices used in schools in the 80s to standardise the marking of what was called Mode 3 work where staff from all the schools in a region came together and moderated their marking in this way.] The Subject Centres would be ideally placed to organise this, and it could sensibly replace the current external examiner system as an extended and much more efficient form of peer review.

5.2 To establish national standards for a degree across disciplines it is necessary to reopen the discussions of the 90s into what does 'graduateness' mean? And it would be sensible to start this discussion by looking at the work on graduate attributes that is on-going in Australia. Only once it has been identified as to what it is that should be common to the notion of a graduate can any systems of comparison be put in place. This is a task that the HEA should be well placed to lead on.

5.3 The QAA should be charged to completely rewrite their good practice guidance on assessment, mindful of the many criticisms of current practice identified in the literature, and summarised above, informed by the ASKe/Weston Manor "assessment manifesto", and starting with a consideration of the abolition of numerical systems.

5.4 Regarding plagiarism, it should be recommended that simplistic solutions (e.g. 'return to invigilated exams') are NOT the answer; neither should university managers hold unrealistic expectations about text-matching software. Instead, it should be recommended that all universities adopt the integrated set of actions (the 'holistic approach') that some have already adopted.