

Comparability of degree standards?

Roger Brown

Introduction

1. In August 2009 the House of Commons Innovation, Universities, Science and Skills Committee (IUSSC) concluded that the Quality Assurance Agency for Higher Education (QAA) should be responsible for “maintaining consistent, national standards in higher education institutions in England and for monitoring and reporting on standards” (IUSSC, 2009: 148¹). The Committee made a number of recommendations to give effect to this. While their recommendations did not refer to comparability of standards, much of their questioning of witnesses had focused upon this.

2. This report discusses the issues involved in comparability of degree standards. It is in two parts. Part 1 begins by outlining the means by which individual universities and colleges and the academic community collectively protect the standards of UK degrees. It then describes the historical attachment to comparability and the pressures which have led to questions being raised about it. Part 2 considers whether genuine comparability is still feasible, and what options may be open to UK higher education if it were found to be impracticable. It should be noted that some other organisations, such as many professional and statutory bodies, set down minimum standards for professional practice. But it is with the academic standards that universities and colleges establish for their awards that this report is concerned.

Definitions

3. For the purposes of this paper, “standards” are described as the levels of achievement required to gain a specific university degree award. As regards “comparable”, the Concise Oxford Dictionary (Eighth edition, 1990) lists a number of meanings. The one that is used here is: “4. the equal or equivalent to”. In other words, for the purposes of this discussion, “comparability” means that there is genuine equivalence in the standards of learning required of, and achieved by, students following any two or more different programmes of study at one or more institutions in the same or different subjects, and leading to the same or a cognate award (this of course assumes that they are able to be compared). Issues related to standards differ from those related to quality. For the purpose of this discussion, matters related to quality are taken to refer to the *process* of learning, those concerned with standards refer to learning *outcomes*.

¹ For full references and bibliography, please refer to the full version of this report on www.hepi.ac.uk.

Part 1: Institutional mechanisms to control quality and standards

4. By international standards, UK universities and colleges have quite elaborate internal controls over quality and standards. The chief ones are:

- Admissions policies, so that only students capable of benefitting from particular programmes are enrolled (though, crucially, these vary considerably between institutions, as well as between subjects within institutions);
- Course approval, monitoring and review, so that only programmes that are fit to lead to an institution's award are offered;
- Assessment regulations and mechanisms, so that only students who reach the required level of attainment receive an award (again, these vary substantially between institutions);
- Monitoring and feedback processes, so that opportunities are taken to improve the quality of what is offered;
- Staff selection and development, so that only suitably qualified and trained staff teach students;
- Staff appraisal, so that staff receive regular structured feedback on their performance.

5. Within assessment, a key role has traditionally been played by external examiners. These are employed by, and answerable to, the institution concerned. Their job is to report on:

- Whether the standards set for awards at the institution concerned are appropriate;
- The extent to which assessment processes are rigorous, ensure equity of treatment for students, and have been fairly conducted within institutional regulations and guidance;
- The standards of student performance in the programmes which they have been appointed to examine;
- (Where appropriate) the comparability of the standards and student achievements with those in some other higher education institutions;
- Good practice they have identified (QAA, 2004).

6. External examiners are one of the chief means of achieving what David Watson has termed "a controlled reputational range"; the practice is virtually unique to the UK (Denmark and Malta also have them). Also peculiar to the UK

is the system whereby the degree awards that students in most subjects receive are classified e.g., First Class, Upper Second (2:1), Lower Second (2:2) etc. (Australia also has degree classification though its degree structures are closer to those of Scotland than England). Finally, professional and statutory bodies play an important role in protecting standards by accrediting programmes that lead to professional practice.

External quality assurance

7. Whilst in law UK institutions have complete autonomy as regards the standards associated with their awards, they work in practice within - what is again by international standards - a fairly extensive set of frameworks or "reference points":

- A Code of Practice covering all aspects of quality management, including assessment and course approval and review as well as external examining;
- A Framework for Higher Education Qualifications containing a broad description of the academic expectations associated with each level of award, together with more detailed descriptors of the skills and competences associated with award holders;
- Subject benchmark statements outlining what can be expected of a graduate in terms of the abilities and skills needed to develop understanding or competence in a particular subject;
- Guidelines for programme specifications setting out the intended aims and learning outcomes of each programme of study.

8. Although mainly related to outputs, not processes, together these are known, somewhat confusingly, as the "academic infrastructure". Institutions' use of the infrastructure is evaluated through periodic institutional reviews covering all aspects of quality management. These reviews, conducted by academic peers, may lead to judgements of "confidence", "limited confidence" or "no confidence" in all or a part of an institution's provision. These judgements in turn may cause a loss of reputation and/or funding. The reports are published by the QAA. What all this means is that the UK almost certainly gives more systematic attention to academic quality and standards than any other comparable system.

The principle of comparability

"Since the Council was established with the purpose of enabling colleges to plan their own courses and to admit and examine their own students, it will impose only such basic requirements as are necessary to ensure that its degrees are comparable in standards to those of the universities."
(Council for National Academic Awards, Statement no 2, April 1965)

9. Between 1965 and 1992, the Council for National Academic Awards (CNAA) was responsible for the standards of the awards offered in the polytechnics and other institutions in what was then called "the public sector" of higher education: they were indeed the Council's awards. The main way in which comparability was established was through the use of academic staff in the existing universities in course approval and review (validation). Subject panels visited institutions to see that curriculum proposals were soundly constructed and that the standards proposed were appropriate to the award. The CNAA's use of staff from existing university institutions established an important principle: that ultimately the only judges of the appropriateness of standards are academic peers in the discipline concerned, and that the way in which these judgements are formed and refined is through a collective process of peer group review, where tacit values and assumptions may be as or more important than open and explicit ones.

10. In the then-university sector, the issue of comparability was underscored by the review of external examining carried out by the Reynolds and Sutherland Committees, under the aegis of what was then the Committee of Vice Chancellors and Principals (now Universities UK) in the mid-1980s. The code of practice that emerged stated:

"The purposes of the external examiner system are to ensure, first and most important, that degrees awarded in similar subjects are comparable in standard in different universities in the UK...and secondly, that the assessment system is fair and is fairly operated in the classification of students." (Committee of Vice Chancellors and Principals, 1986)

11. This historical introduction has been provided to show how longstanding is the British attachment to comparability of degree standards. Within the UK, students, employers and others value consistency, which is also reflected in common undergraduate fee limits (and, generally, levels). Externally, the UK's success in attracting international students, partners and staff has depended very largely on the continuing currency and standing of, and some degree of consistency between, institutions, subjects and programmes.

Pressures on comparability

12. The CNAA was abolished in 1992 following the Government's decision to allow the polytechnics and certain colleges to obtain university title. The Higher Education Quality Council (HEQC) was established as a sector-owned body to monitor and advise on institutions' academic standards. At the Higher Education Funding Council for England (HEFCE) Annual Conference in April 1994, the then Secretary of State for Education and Science, John Patten MP, asked the Council to give greater attention to "broad comparability" of standards between institutions. HEQC's main response was to propose the academic infrastructure that has already been described. The Council also gave greater attention to academic standards within the institutional quality audit process.

13. Three sets of factors have now combined to raise further question marks over comparability.

14. First, the substantial evidence that has emerged over many years about insufficient professionalism by institutions, departments and academic staff in the practice of assessment leading, *inter alia*, to significant variations in the levels of achievement aimed at and realised by students – that is to say, inconsistent standards.

15. The QAA summarised some of these concerns in a 2008 publication 'Outcomes from institutional audit; Assessment of students; Second series':

"Worries include doubts in some cases about the double-marking and/or moderation of students' summative assessment; continuing difficulties with degree classification; departures from institutional practice in the way staff in departments and schools work with external examiners; and generally weak use of statistical data to monitor and quality assure the assessments of all students and degree classifications. The papers also find weaknesses in the arrangements of some institutions for detecting and dealing with plagiarism and for providing feedback on students' assessed work, including feedback to international students."

16. In a recent and comprehensive review, Yorke (2008) identified five main problem areas: variations in regulations and practices between and within institutions; lack of technical robustness, especially reliability; concerns about grading including the inappropriate use of arithmetic manipulations to produce an overall grade for a student's achievements; lack of clarity about expected performance even where learning outcomes are specified in some detail; and the communication of assessment outcomes, including insufficient appreciation of the "fuzziness" of assessment judgements and the limited reliance that can be placed on them. The fundamental problem is the complexity of knowledge and the difficulty of grading complex learning achievements at a time when because of wider changes in the system (see below), there is increased pressure for warrants of outcomes. Because a university education is not designed simply to impart knowledge but, in the words of the Robbins Report, to "develop the general powers of the mind", and more generally to develop intellectual powers and analytical thinking, assessing the extent to which different students achieve this is a particular challenge. Nevertheless, weak assessment practice exposes institutions to challenges from aggrieved students, as well as creating unfairness. It also undermines external examining as a vehicle for assuring comparability: how can external moderation be effective if internal assessments are insufficiently robust?

17. The second factor threatening comparability is the enormous expansion and diversification of the system since the mid-1980s. As well as the increase in the numbers of institutions awarding degrees and the number and range of subjects and students in the system, three developments are of particular

significance: (a) the increase in the categories of work being examined (invigilated exams, coursework, portfolios, projects, placements, etc.) and an associated reduction in the breadth of the knowledge and understanding actually being assessed at any one time; (b) the growth of joint, inter- and multi-disciplinary, and modular courses (modularity in particular places considerable demands on external examiners historically recruited mainly for their subject standing, knowledge and expertise); and (c) the increased importance of such concepts as "enterprise", "employability" and "transferable skills" to which conventional assessment methods, concerned as they mainly are with testing mastery of subject matter, may not be well suited. The net result is that, as the "organising principle" of assessment, subject/discipline has given way to institutional regulations and exam rules.

18. The third factor affecting comparability is the increased competitive pressure on institutions associated with variable fees and bursaries, league tables, resourcing and greater commercialisation. These pressures, and particularly the evidence (in a small number of cases) of management intervention in academic judgments on standards, were one of the main causes of the Select Committee inquiry. Increased "marketisation" is almost certain to lead to greater variations in standards at the same time as challenging the peer review structures that are the ultimate guarantor of quality (in a commercial marketplace, one legitimate response to competition is to reduce quality at the same time as maintaining price).

19. Together, these factors are a response to national policies to expand the system, diversify provision, increase efficiency, make the curriculum more responsive to perceived economic needs, and enlarge student choice at the same time as resources have been constrained and successful performance in research has continued to take priority in many institutions over pedagogical effectiveness. What they mean in a standards context is that whilst the desire for some degree of comparability has remained (and even in some respects increased, as the system has expanded and become more diverse), the ability to achieve it has diminished, ipso facto. How feasible is it to retain (or regain) real comparability?

Part 2: Responses to these pressures

20. In principle, genuine comparability of degree standards would require each of the following conditions to be met:

- There would need to be comparability of standards within all the components of a degree programme (including options) within an institution;
- There would need to be comparability in the standards required in the degree programme followed over several years;

- There would need to be comparability in the standards aimed at and achieved in similar programmes in the same subject in different institutions. The programmes would need to be taught by all the relevant institutions to a similar level and the work marked to similar standards. A student emerging with a degree in a particular subject from any institution would need to have achieved the same level of knowledge and understanding as students in that subject in all other institutions ;
- There would need to be comparability in the standards aimed at and achieved in different subjects both within an institution and across the sector.

21. Comparability within a single degree programme in a single institution should in principle be achievable. However, the extensive evidence about internal variability of assessments makes it seem unlikely that it is often achieved in practice.

22. Comparability of standards as between successive student cohorts should also be achievable in theory. However it is rare for assessed work to be judged against the standards of work of candidates on the programme in earlier years: it is far more common for examiners to carry the standards in their memories from year to year (there are several studies casting doubt on the efficacy of this).

23. The assurance of comparability within subjects taught at different institutions is one of the main purposes of external examining. Yet there remain significant inter-institutional variations in the distribution of degree classes that go beyond those that may be explicable in terms of students' entering achievements and characteristics. In fact, as the QAA pointed out in 2007, there are few opportunities for external examiners (or others) to make comparisons of the academic performance of students in specific subjects across – i.e., between – institutions (QAA, 2007a). The QAA document adds "Hence it is not surprising... that adverse comments by external examiners regarding the operation of institutional rules for classifying degrees are relatively uncommon: secure evidence for such adverse comments would be hard to identify" (paragraph 16).

24. Finally, even in the mid-1980s, consistency in standards between subjects was recognised to be difficult, as the earlier extract from the CVCP code on external examining indicates. This would be true even if there were more opportunities for assessors from different disciplines to work together. Yet the system we now have was conceived in a time when there were far fewer institutions, subjects and students. That said, there can be little doubt that, within an institution, having people from different disciplines commenting on work in other disciplines is a powerful stimulus to quality enhancement.

25. This discussion suggests that it would be wrong to aim for an unrealistic degree of comparability. Indeed, it actually suggests that the very question of whether the standards of degrees are comparable may be misguided.

26. At a time when only a very small proportion of the population went to university, and the student population was broadly equivalent in terms of background and ability – and when degree courses were considerably more uniform in terms of their nature and intended outcomes than they are now – it may have been a reasonable expectation that the outcomes of degree courses should be broadly comparable, and that there should be mechanisms available to police this (hence, external examiners). Today, the environment is radically different. Nearly half of the young population now participate in higher education, the range of ability of those students is much wider, and the purpose, nature and intended outcomes of programmes all vary considerably. It makes little sense to seek comparability of outcomes, and indeed it would actually be wrong to do so. Given the extraordinarily high previous educational attainment of students attending, say, Oxford or Cambridge, the substantially greater resources devoted to them, the greater intensity of study that they undergo, and other factors, it would in fact be a surprise if the outcomes of students from those universities were no higher than those of students from other universities who have far lower prior attainment, resources devoted to them, and so on. But, self-evident as this might seem, there are actually no instruments available to demonstrate it.

27. So, there are almost inevitably differences in the standard of outcomes of different universities. And it is right that that should be so. But there are no objective measures. What is perhaps more important is that there should be confidence that students graduating from all higher education institutions in all subjects have attained a minimum level of achievement. There must be confidence that the degree represents some minimum level of skills and knowledge. This does not of course guarantee comparability of standards; but the guarantee of a minimum level of achievement would nevertheless represent a readily understood and worthwhile feature of our higher education system.

28. Beyond that, having accepted that there are bound to be – and should be – differences in standards, it would be good if a mechanism were available to differentiate the knowledge, skills and ability of students from different universities (as well as courses). However this has eluded those searching for it so far, and the search seems almost certainly doomed to failure. Whereas the mechanisms of degree classification and external examiners to police these may have worked well in a small, elite, system where a small number of people could monitor these things, that no longer applies. Yet the same expectations exist (repeated by the Select Committee), and the same instruments remain to try to deliver the expectations. So what is to be done?

29. Rather than aim for an unrealistic degree of comparability, on the one hand, or abandon the notion, on the other, it is suggested that we should pursue

a number of approaches aimed at improving the assurance of standards and constraining the amount of variability in the system. The following paragraphs outline what these might be.

30. First, institutions should publish more information about the aims and outcomes of individual courses and programmes, what it is that students need to do to be able to obtain them, and how the institution will provide them with the necessary facilities and opportunities. This is already an important plank of national policy, in part because of previous work by HEPI. Although they have been criticised as reductionist, the learning outcomes now expected for every course should enable everyone connected with it to see what the course is about and in particular what kinds of achievement are aimed at.

31. Second, although they vary considerably in coverage and specificity, the subject benchmark statements do provide authoritative guidance about the aims and outcomes that the academic community generally considers appropriate for degrees in particular subjects. The current review of such statements might consider whether they can be sharpened up for this purpose without detracting from institutional responsibility or academic discretion. It should be understood though that these are concerned only with minimum levels of knowledge and understanding required to achieve a degree, and not with comparability of degree standards.

32. Third, one of the biggest problems with external examining is that institutions tend to rely too much upon external examiners whenever questions are raised about the appropriateness of their standards. QAA and the Higher Education Academy could between them sponsor the creation of networks of staff, generally but not exclusively on a subject basis, concerned with different aspects of assessment. The aim would be to strengthen the ways in which staff determine the appropriateness of the courses and awards they design by reference to practice elsewhere. This would be done by comparing the quality of student work and the assessment judgements made in relation to it. In this way the sector could build shared understandings of assessment standards in a way that is still uncommon at present (this was in fact one of the principal recommendations of the Graduate Standards Programme). This is the direction in which external examining might move.

33. Fourth, the sector should make a determined effort both to improve the quality and consistency of assessment practice and to communicate the limitations of any assessment method, even ones more valid and reliable than what we currently have. The QAA's 2009 report 'Thematic Enquiries into Concerns about Academic Quality and Standards in Higher Education in England' recommended a review of assessment practices "supported by developmental activities aimed at improving the robustness and consistency of assessment and classification practices within and between institutions", together with clarification and explanation of the reasons for, and meaning of, variation in particular approaches to assessment. This would seem appropriate, if not long

overdue. However, it is not mentioned in the consultative document on future arrangements for quality assurance recently published by HEFCE on behalf of the funding agencies and the representative bodies and it is not clear what priority this now has.

34. Fifth, we should press on with alternatives to honours degree classification even if a wholly or mainly formative assessment regime (strongly preferable, educationally) is unrealistic in a market-led system.

35. Sixth, whilst valuing the way in which external examining helps to keep the sector "honest", we should be more realistic about what we can expect from external examiners. Phasing out degree classification would certainly help here, but there is much more that needs to be done.

36. Finally, we should try as a sector to be much clearer about what we mean by such terms as "comparability" and "consistency". If we are muddled in our thinking, in our practices, and in the claims we make for them, how can we expect others to be clear about them?

Conclusion

37. Historically, the UK has had a strong attachment to the principle of comparability of degree standards. However, it is increasingly doubtful whether, in a diverse mass system incorporating significant student choice, a real degree of comparability, in the sense of equivalent levels of student learning achievement across all institutions and subjects, is practicable or even desirable. However, this need not mean a dilution of standards – minimum standards need to be maintained, and as far as possible differences in standards should be recognised and described.

38. There are a number of things that the sector could do to achieve this. Many of these are already in train. Perhaps the most fundamental is to improve the quality of assessment, with much greater professionalism and much more systematic comparisons of student performance on similar programmes across a range of institutions. This could be done either through a reformed external examiner system or through the creation of new subject-based networks, or some combination of them. Contrariwise, if a consensus is not achieved on comparability then the current moves to reform quality assurance, including the review of external examining and the evaluation of the academic infrastructure, may prove nugatory.