

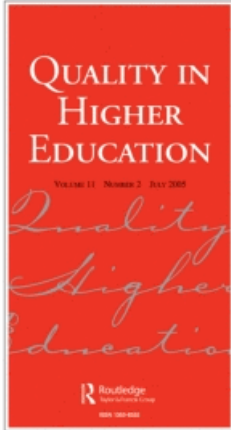
This article was downloaded by: [PERI Ethiopia]

On: 10 April 2011

Access details: Access Details: [subscription number 789104926]

Publisher Routledge

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



## Quality in Higher Education

Publication details, including instructions for authors and subscription information:

<http://www.informaworld.com/smpp/title~content=t713443244>

### Benchmarking Academic Standards in History: an empirical exercise

Harvey Woolf<sup>a</sup>; Angela Cooper<sup>b</sup>; Bernard Bourdillon<sup>b</sup>; Paul Bridges<sup>b</sup>; Debbie Collymore<sup>b</sup>; Chris Haines<sup>b</sup>; David Turner<sup>b</sup>; Mantz Yorke<sup>b</sup>

<sup>a</sup> University of Wolverhampton, Wolverhampton WV1 1SB, UK <sup>b</sup> Anglia Polytechnic University, UK

Online publication date: 05 November 2010

**To cite this Article** Woolf, Harvey , Cooper, Angela , Bourdillon, Bernard , Bridges, Paul , Collymore, Debbie , Haines, Chris , Turner, David and Yorke, Mantz(1999) 'Benchmarking Academic Standards in History: an empirical exercise', *Quality in Higher Education*, 5: 2, 145 – 154

**To link to this Article: DOI:** 10.1080/1353832990050205

**URL:** <http://dx.doi.org/10.1080/1353832990050205>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.informaworld.com/terms-and-conditions-of-access.pdf>

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

# Benchmarking Academic Standards in History: an empirical exercise

HARVEY WOOLF<sup>1</sup> & ANGELA COOPER<sup>1</sup> WITH BERNARD BOURDILLON<sup>2</sup>, PAUL BRIDGES<sup>3</sup>, DEBBIE COLLYMORE<sup>4</sup>, CHRIS HAINES<sup>5</sup>, DAVID TURNER<sup>6</sup> & MANTZ YORKE<sup>7</sup> [1]

<sup>1</sup>University of Wolverhampton, Wulfruna Street, Wolverhampton WV1 1SB, UK, <sup>2</sup>Anglia Polytechnic University, UK, <sup>3</sup>University of Derby, UK, <sup>4</sup>University of Wolverhampton, UK, <sup>5</sup>Middlesex University, UK, <sup>6</sup>Oxford Brookes University, UK, <sup>7</sup>Liverpool John Moores University, UK

**ABSTRACT** *The article reports on a small-scale project that investigated the potential of using a benchmarking club on assessment practices in History as a means of establishing and comparing academic standards. The context of the project is described and the benchmarking club's key findings are presented. Issues addressed are the subject goals, levels of study, assessment methods, assessment criteria, marking practices, feedback mechanisms, and the monitoring and dissemination strategies used in the participants' universities. A number of preliminary conclusions aimed at national policy, institutions, subjects and individual modules suggest that benchmarking as a process for comparing academic standards across departments to identify best practice could be a valuable tool of quality assurance. However, there must be doubt about the worth of creating academic benchmarks as baselines against which to measure the academic standards of subjects.*

## Introduction

The potential for benchmarking academic standards is evaluated through a small comparative investigation into the assessment practices employed in the History discipline in four British universities. The study was part of a broader examination carried out by the Student Assessment and Classification Working Group (SACWG) [2] of how academic standards are established in a number of disciplines. The work was sponsored by the Higher Education Quality Council (HEQC) and its successor, the Quality Assurance Agency for Higher Education (QAA), as part of a much wider investigation into benchmarking academic standards. The first section outlines the background to the project. Section two summarises the approach SACWG adopted for and the key outcomes of the History benchmarking club. The final section highlights some of the preliminary conclusions about benchmarking academic standards in higher education in the UK that the group has drawn from the exercise.

## Background

The growth in the diversity and complexity of the provision of higher education in the UK over the past two decades has raised questions about the extent to which academic standards are consistent across time and between different institutions (Chapman, 1996; NCIHE, 1997). As part of its work on graduateness between 1995 and 1997, the HEQC had

started to explore the possibilities for using benchmarking as a tool to obtain a greater understanding of how academic standards are established and how standards can be compared across a diverse sector (Jackson, 1999).

Industry, commerce and public services such as hospitals in the UK have used benchmarking for many years. Companies and organisations involved in similar activities have formed themselves into 'benchmarking clubs' to provide a structural framework within which to benchmark their practices. When applied to higher education, benchmarking has been limited to quantifiable comparisons of, for example, financial systems (CVCP, 1962; Fielden, 1997). Extending the principles and practices of benchmarking to something as conceptually complex as academic standards was therefore entering new and largely uncharted territory. Part of the challenge was, and is, that there is no one notion or definition of benchmarking. Benchmarking may be undertaken to identify best practice in the field to enable other institutions to adopt that practice for their own needs. It might seek to establish a threshold for a minimum level of acceptable performance. It might be used to compare an organisation's performance with others in the same field and to determine whether that performance is 'good enough' for the organisation's continued well-being (Price, 1994; Jackson, 1999; Yorke *et al.*, 1999).

SACWG has been investigating differences in subject and degree performances in the member institutions since 1994 (Yorke *et al.*, 1996; Woolf & Turner, 1997). Although the group had not hitherto used the term, it had effectively been engaged in a series of benchmarking exercises and had generated many of the components of a benchmarking club. The HEQC project extended SACWG's work to a more systematic analysis of assessment methods and practices and to a more self-conscious reflection on the conditions under which benchmarking clubs can be successful. The historians who participated in the project were afforded an opportunity to compare their approaches to assessment with colleagues from other universities in a structured way and to identify good practice in those institutions. There was no intention to define or design either threshold standards or level descriptors for SACWG History departments. The project thus focused on two of the several components of benchmarking: identifying best practice and comparing performance.

### *The History Benchmarking Club*

History was selected as an example of a traditional Arts discipline. From previous work, it was known that historians tended to use a relatively narrow spread of marks and that grade distributions were broadly similar in all the member institutions. This suggested that there were common assumptions underlying assessment in History that could be usefully explored in a SACWG benchmarking club. Eventually, four of the seven SACWG universities responded to the invitation to take part in the project. All were post-1992 universities with a commitment to modular structures, programme flexibility and widening participation in higher education. All had experience of the Council for National Academic Awards' (CNAA) processes and procedures. The four departments all offer History as a single subject degree and as part of a modular combined honours programme. Each department provides modules that cover a wide temporal and geographical range of topics and all include historical skills units in their programmes.

The project was launched in March 1997 with a general briefing for representatives from all the SACWG institutions on the nature and purpose of benchmarking. This briefing was an important stage in the project as it permitted the potential participants to clarify issues and to resolve any concerns, for example about confidentiality and ownership of the

process. Subsequent contacts were used to elaborate the purpose of the exercise and distribute the first draft of a questionnaire designed to identify the departments' assessment philosophies, policies and régimes. These communications also served as a helpful ice-breaking prelude to a formal meeting of the benchmarking group in May 1997. A key element of the meeting was the time spent developing a sense of openness and ownership. Common problems, anxieties and achievements were shared from the outset. This forming stage helped to underline the facilitators' statement that there was no hidden agenda driving the project. As well as creating an appropriate dynamic for the group, the meeting confirmed the more specific goals of the project and finalised the content of the questionnaire. The goals were:

- to identify what the assessment tasks and practices actually were for each module surveyed and for the study of History as a subject in each of the four universities participating in the project;
- to explore the rationales underpinning the various assessment practices in operation in each module, and within the four History departments;
- to compare the student results on similar modules at the same level to try to understand the implications for academic standards in the different universities;
- to explore the significance for standards of differences in assessment practices.

The questionnaire covered the departments' overall assessment policies and the specific assessment régime and methods for eight modules selected by the departments and for two common modules, 'Introduction to History' and the third-level (final year) 'Honours Dissertation', chosen by the group as a whole. The results derived from the questionnaire were discussed at a further meeting in September 1997. Subsequently, all the participants were invited to comment by correspondence on a draft report based on the meetings and analyses of the questionnaires. Ideally, there should have been more face-to-face meetings but the timescale and the funding for the project precluded the organisation of additional meetings.

### **Extent of Commonality**

The broad conclusion of the project was that there is considerable commonality in approaches to, and outcomes of, assessment across the participating departments. The following are among the findings that illustrate this conclusion.

#### *Subject Goals*

The aims and outcomes of the sample of modules used for the survey demonstrated a high level of consistency among the departments. Although there were some differences in terminology, all the participants thought that there needed to be a balance between subject knowledge and personal transferable skills.

In particular, there was a shared understanding of the importance of independent research as a key skill that history students should acquire. This was reflected in the significance placed on the third-level honours dissertation, which was seen as an essential component of all History programmes.

Key skills or personal transferable skills or their equivalent were regarded as an important constituent of the History curriculum. Although these skills were not always explicitly stated in departments' documentation, they were present in all the History assessment regimes considered by the group.

*Levels of Study*

All the members of the group found difficulty in specifying *levels of study* in a coherent and meaningful way. None the less, all the universities had working definitions of levels. There was agreement that the same assessment instrument could be used at different levels.

An attempt was made to compare the intellectual demands being placed on students at different levels using Bloom's Taxonomy (Bloom, 1956) to analyse a sample of coursework and examination questions. Whilst the participants recognised Bloom's limitations, there was a consensus among the four departments that the same sort of intellectual and subject-specific skills should be assessed at each level. As might be anticipated, the departments emphasised the higher-level cognitive skills of analysis, synthesis and evaluation. What differed at each level were the expectations of the staff marking the assignments.

*Pressures on the Assessment Process*

There was a widespread concern that external pressures (increased student numbers, diminishing resources and heavier reliance on part-time staff in the teaching and assessing of students) were determining the selection of assessment strategies.

Additionally, there were growing concerns about over-assessment of students, the instrumentality of (some) students in their attitudes towards learning and a possible lack of parity in workloads across different modules and different assessment régimes.

*Assessment Methods*

The essay, the dissertation and the formal, timed, unseen examination remained the staple methods of assessment for History. Essays were of various lengths and allowed varying degrees of autonomy to students in the selection of their own subject matter. Occasionally, the departments experimented with variants of seen and open-book examinations. There was only limited assessment of the formal presentation of seminar papers, other types of presentations or seminar contributions. Group projects, critical bibliographies, 'editing' document collections, peer- and self-assessment were among the other methods adopted. However, these approaches represented a small minority of the ways in which History modules were assessed.

*Assessment Criteria*

While there may be university or faculty guidelines for assessment criteria, these are invariably and necessarily translated into subject and then module assignment criteria by the subject teams. All the participants have developed or were developing grading criteria for either the subject as a whole or for the different levels in the subject or for individual modules and assignments.

In three of the four universities more work appears to have been done on criteria for oral work than for written work. There was no clear explanation for this. It may be that because oral assessment in history is relatively new it had prompted staff to think about the matter from first principles. However, written work could be perceived as 'traditional' and, as such, staff might have been satisfied that there was vast experience of assessing the standard of individual pieces of work. Hence, perhaps there had not been the same sense of a need to publish criteria for written work as there had been for oral work.

Whereas there was evidence of considerable thought having been given to developing a variety of assessment tasks or instruments within modules, less attention appeared to have been given to the development of grading criteria for individual assessment tasks. There was certainly a feeling that grading criteria can be perceived as unhelpfully rigid in evaluating the quality and standards of individual assignments.

### *Marking Practices and Moderation*

Staff in each of the departments used several 'marking models' when determining levels of performance. These included:

- *the platonic model*, in which the marker has a clear idea in mind about the content, style, and presentation expected and assesses against that model;
- *intuitive approaches*, which derive from the marker's experience of grading at undergraduate level;
- *criteria-referencing*, where the marker produces a set of criteria (for example, relevance, structure, hypothesis formulation) and then grades against these criteria.

Some of the participants chose a sample of scripts to help them determine finer distinctions within broad categories of a grade band before finally grading a whole batch of assignments. No one ranked all scripts before either ascribing or confirming grades for an assignment. Within the group the specific model used tended to be left to individual markers.

Sample double-marking of coursework routinely occurred in three of the four departments. The subject team usually determined the sample sizes. No department used blind double-marking for coursework (blind double-marking is where the second marker is unaware of the comments or graders of the first marker). The reasons for this were more pragmatic than matters of principle: staff were eager to return assignments to students, as quickly and with as much comment as possible. Blind double-marking was seen as delaying feedback to students, and being cumbersome to manage. All four departments double-marked every dissertation submitted, which underscored the importance attached to the dissertation in a history programme. Two departments have opted for blind double-marking of dissertations. However, there may also be a pragmatic reason for this: normally two copies of a project are submitted simultaneously and thus there need be no additional delays in returning work with comments to students.

Double-marking of examination scripts was a standard practice in all the departments. Whether all the scripts or a sample are second-marked derived from a combination of university and departmental policies. Three of the four departments used blind double-marking for examinations. All but one of the departments had adopted anonymous marking for examinations. The use of student numbers and other devices, such as 'stick down corners' on examination answer papers, allowed anonymity of examination scripts to be logistically feasible. No one yet appeared to have solved the conundrum of how to mark coursework completely anonymously. Samples of both coursework and examination scripts were sent to external examiners by all the departments. As with other aspects of moderation, the size and nature of the sample were dependent on a mixture of university and departmental policies and practices. The external examiner was regarded very much as a partner in the confirmation of standards. Although the external examiner's recommendations on changes to marks may ultimately take precedence, the process was perceived as one of negotiation.

*Feedback on Assessment and Documentation for Students*

All departments stressed the importance of feeding back to students on the results of assessments. There was an increasing use of cover sheets to help standardise the nature of the feedback given for coursework across the range of a department's modules [3]. In contrast, there tended to be little feedback given to students about their examination performance.

The return of coursework assignments and, when it occurred, examination scripts was in some cases accompanied by tutorials. However, it was becoming rarer for these tutorials to be an integral part of the assessment process. At the same time, there was a perception among staff that, as external pressures on students increased, many students were becoming more instrumental about assessment: it was reported that students often did not attend feedback tutorials. 'Piles of uncollected assignments in the departmental office' was how one participant described the outcome of this attitude.

In keeping with the practice of making information readily available, all the departments produced subject guides and other supplementary information. This documentation was distributed to students, part-time and full-time staff and external examiners. In addition to other material, these guides provided guidance on assessment methods, assessment criteria and how students could obtain feedback on their assessments.

*Student Performance and Assessment Tasks*

Although more detailed work remains to be done on the relationship between different types of assessment tools and individual module results, a preliminary survey of results indicated that there was no discernible difference in overall module results by type of assessment used. One representative's experience, though, was that where essay titles were negotiated, a higher grade was more likely to be achieved.

*Monitoring Standards*

Individual module teams together with the subject leader checked the reliability and validity of assessment instruments at the start and at the end of a module, once the final grades had been ratified. At the earlier stage, either a small group or the subject leader reviewed assessments to ensure that there was no overlap between examination and coursework questions and that the questions were comprehensible. External examiners were usually involved in this exercise. After a module had been completed, module grades were interrogated to help determine the reliability and consistency of modules' assessment processes. The findings of all these checks then informed the assessment practices of the next iteration of the module and were fed in to the subject's annual monitoring report.

While all the departments had a battery of monitoring mechanisms in place, the participants believed that colleagues could be trusted to know how to assess and that it was this trust which provided the ultimate reliability of the departments' assessment processes. Yet it was also noted that many of the traditional mechanisms for socialisation into the departmental culture had disappeared or were disappearing rapidly. Particular concern was expressed about the increasing reliance and demands being placed on part-time staff. Some institutions did involve part-time staff in staff development activities but this was not universal. Even though more formal systems were being put in place to counteract the fragmentation of subject teams, these systems, by themselves, could not recreate the traditional forms of subject acculturation.

### *Disseminating Standards and Benchmarks Among Staff*

All the universities had developed a variety of mechanisms and methods for helping new staff understand the academic standards established by the departments. The mechanisms included a combination of formal mentoring schemes; informal discussions with more experienced colleagues; being part of a module teaching team; and (accredited) institutional staff development programmes.

Whatever methods were used, there was a strong feeling in the group that 'the whole process [of establishing standards] is subtle, organic and opaque' and that it is not easily reducible to a series of workshops on the techniques of assessment. Moreover, there was a recognition, and concern, that many of the informal methods of setting commonly-agreed standards that were used when group sizes were smaller were no longer either viable or sufficient to meet the needs of mass higher education.

Disseminating information about standards within departments occurred in a variety of ways. Major vehicles in all the departments were external examiners' observations and reports and the annual monitoring of modules, subjects, and awards. Annual staff-development activities and review and validation events were another important means of informing colleagues about matters relating to quality and standards. Other conduits and sources were colleagues who acted as external examiners in other institutions or as QAA assessors, departmental teaching and learning sub-committees, the formal appraisal of teaching, national reports, such as QAA Teaching Quality Assessments, and informal monitoring discussions with colleagues.

### **Preliminary Conclusions**

A small-scale study inevitably raises more questions than it can possibly answer. Nonetheless, the exercise has pointed the way to a number of conclusions about the nature of benchmarking academic standards that further work will either support or refute. These conclusions relate to the operation of academic benchmarking clubs and to the significance of benchmarking for national policy for higher education institutions, for subjects and for individual modules.

Although these conclusions relate to the UK, benchmarking may have implications for higher education systems in other parts of the world. For example, in countries where there is no national framework for quality assurance, benchmarking clubs could serve as a means of providing inter-institutional comparability. Even where national frameworks are in place, a benchmarking exercise of the kind described here could provide additional evidence about the equivalence of degrees in the system. Certainly international benchmarking clubs can aid credit transfer and accreditation by creating forums within which to develop mutual understanding of different higher education cultures and processes.

### *Academic Benchmarking Clubs*

Benchmarking has been likened to an academic research project. As with a research project, it is important that all participants are absolutely clear from the start about the purpose of the undertaking. Equally important is the need to test the data collection instruments in order to ensure that ambiguities and redundancies are removed before the collection of data 'goes live'. Perhaps the key prerequisite for the establishment of a successful group is the building of trust and confidence among the participants. Face-to-face meetings help to build trust relatively quickly, generate high levels of common understanding and provide



opportunities for the exchange of good practice. While work can be carried out by correspondence, this inevitably generates its own difficulties. Not least of these is the protracted nature of the dialogues needed to refine and modify issues. Telephone calls, faxes and e-mails certainly help to accelerate the process but, even with these systems, there is an unavoidable time lag in arriving at a mutually accepted conclusion to a debate or discussion. Above all, correspondence does not easily admit the interaction and shared understanding that it is possible to achieve in face-to-face meetings. Holding meetings creates resource demands, not least of which is the call on participants' time. Benchmarking clubs have to be administered and financially resourced, if only at the minimal level of reimbursing participants for their out of pocket expenses.

Part of the sense of mutual trust should stem from the activities of the facilitators in creating a non-judgemental and supportive working environment. The role of the facilitator is crucial. In this project, one of the facilitators taught on one of the contributing programmes and the other had a broad social science background. This meant that there was a raft of matters that were generally understood and which therefore did not need to be articulated or discussed. However, it was all too easy for assumptions to go unquestioned or unchallenged. Whichever approach is adopted, it is essential that the facilitators should explore the group's presuppositions at as early a stage in the process as possible.

All the project's participants came from post-1992 universities. Certainly, the involvement of staff from similar institutions helped to create an initial confidence. A number of assumptions, both spoken and unspoken, could be taken for granted. There was a common institutional language, even if there were several dialects of that language. Histories and cultures, problems and opportunities were similar. However, the inclusion of participants from other types of institutions should not be an insuperable barrier to the creation of an effective benchmarking club. There would, though, probably have to be a longer preparatory phase. The precise length of this phase would likely be related to the size of the group. The experience of the History project and the sister projects in Business and Computing suggests that the optimum membership of a benchmarking group is between four and twelve. Four members are the minimum to allow for a sufficient variety of perspectives to be explored; more than twelve is probably too large a group to encourage an effective interchange of ideas. There is a parallel here with the group dynamics of academic seminars.

### *The National Dimension*

The project put into sharp focus the extent to which academic standards and assessment criteria are relative rather than fixed. They differ between institutions and subjects and, the group's discussions implied, over time. It does not follow, though, that one institution's or one subject's standards are higher or lower than another institution's or subject's standards simply because they are different. Benchmarking is a means of establishing and demonstrating the consistency of standards across institutions and subjects, while preserving diversity of provision. The QAA's decision to issue subject benchmarking statements (QAA, 1998) is a brave pioneering effort to articulate a series of threshold standards for higher education in the UK. However, the evidence of this project indicated that such statements are likely to be at such a level of generality that they will permit many different interpretations, rather as the four History departments have inflected their institutional level descriptors to meet the specific demands of the departments' individual curricula. This conclusion was supported by the publication of the Chemistry, History and Law draft pilot benchmarking statements in the last months of 1998 [4].

### *The Institutional Dimension*

There is value in institutions establishing subject benchmarking clubs across the institution and between institutions. To ensure that like is being compared with like, there would probably be more value in institutions joining self-selecting clubs. Both intra- and inter-institutional benchmarking are likely to create a greater and systematic understanding of academic similarities and differences in, for example, courses and programmes, modules, level descriptors, assessment strategies, methods and tasks, marking and grading systems, and regulatory frameworks. Such understanding would, in its turn, encourage self-improvement in the performances of institutions.

### *The Subject Dimension*

The project indicated that benchmarking can be a powerful tool for helping subject groups to clarify and confirm the validity of their rationale for the assessment principles and strategies they have adopted. Such confirmation would reinforce the trust colleagues have in each other's marking and would serve as another means of ensuring that there is a common understanding of academic standards at the different levels of the subject's provision across the entire department. Benchmarking could, for example, lead the subject team to review the evidence it requires to allow the team and their external examiner(s) to make meaningful judgements about its academic standards, to re-consider the data presented to the formal assessment boards or, indeed, to revise its whole conception of academic standards.

### *The Module Dimension*

At module level, the project highlighted once again that it is essential to consider the extent to which a module's assessment tasks and the weightings attached to those tasks are appropriate for the module's stated outcomes. In particular, there is a need to ensure that key or personal transferable skills developed in the module are clearly articulated and suitably assessed. The work in the project also underlined how important it is to question the appropriateness of the assessment criteria for the types of assessment employed in a module and to review the module team's approaches to marking assignments to enhance consistency, especially when several members of staff teach the module.

### **A Final Comment**

The emphasis in the project was to examine the potential of using an assessment benchmarking club as a means of investigating how academic standards were established, of exploring the extent to which such standards were consistent across a number of institutions and of identifying good practice in the participating universities. The project was not concerned to establish a set of (minimum) benchmarked standards against which to measure a department's performance as teachers and assessors of History degrees. Nor did the History club set out to devise a set of common level descriptors for History in the four universities. However, the project threw doubts on the feasibility of producing such prescribed performance measures or generic descriptors. Against that, the History group provided a considerable amount of positive evidence of the value of academic benchmarking *as a process* to enhance understanding of academic standards within a discipline and to compare examples of good practice across the sector. It is the authors' view that it is this

approach to benchmarking, rather than the generation of benchmarking statements which specify (threshold) standards, that will make more productive use of time when academics are assuring quality in their disciplines.

## Notes

- [1] Collectively, the Student Assessment and Classification Working Group (SACWG).
- [2] SACWG is an informal group of individuals from the Anglia Polytechnic, Derby, Liverpool John Moores, London Guildhall, Middlesex, Oxford Brookes and Wolverhampton Universities who have a common interest in assessment matters. SACWG acknowledges the contributions of the Heads of History who participated in the project and the support of HEQC/QAA, both for this project and for the Group's four national workshops. The views expressed in this article should not be taken as representative of either the former HEQC or QAA. The two other subject areas involved were Business and Computing.
- [3] An aspect of feedback which would have been explored in a longer term project is students' evaluations and expectations of good quality feedback.
- [4] Subject groups, convened by the QAA, prepared the draft statements and consulted with the subject communities in various ways. These drafts and subsequent versions of the statements are available from the QAA.

## References

- BLOOM, B.S., 1956, *Taxonomy of Educational Objectives: I cognitive domain*, 1st edn (London, Longmans).
- CHAPMAN, K., 1996, *Inter-Institutional Variability of Degree Results: an analysis in selected subjects* (London, Higher Education Quality Council).
- COMMITTEE OF VICE CHANCELLORS AND PRINCIPALS OF THE UNIVERSITIES OF THE UNITED KINGDOM AND THE UNIVERSITIES FUNDING COUNCIL (CVCP/UFC), 1962, *University Management Statistics and Performance Indicators* (London, CVCP/UFC).
- FIELDEN, J., 1997, 'Benchmarking university performance', *Association of Commonwealth Universities Bulletin of Current Documentation*, 131 (October), pp. 20–25.
- JACKSON, N., 1999, 'Introduction to benchmarking assessment practice in UK higher education', in: N. JACKSON (Ed.) *Pilot Studies in Benchmarking Assessment Practice* (Gloucester, Quality Assurance Agency).
- NATIONAL COMMITTEE OF INQUIRY INTO HIGHER EDUCATION (NCIHE), 1997, *Higher Education in the Learning Society: report of the National Committee of Inquiry into Higher Education* (Norwich, HMSO).
- PRICE, I., 1994, *A Plain Person's Guide to Benchmarking* (Sheffield, Unit for Facilities Management Research, Sheffield Hallam University).
- QUALITY ASSURANCE AGENCY (QAA), 1998, 'Subject benchmarks', *Higher Quality*, 4 (October), pp. 11, 18–19.
- WOOLF, H. & TURNER, D., 1997, 'Honours classifications: the need for transparency', *New Academic*, (Autumn), pp. 10–12.
- YORKE, M. et al., 1996, 'Module mark distributions in eight subject areas and some issues they raise', in: N. JACKSON (Ed.) *Modular Higher Education in the UK in Focus* (London, Higher Education Quality Council).
- YORKE, M. et al., 1999, 'Benchmarking academic standards: a pilot investigation', in: N. JACKSON (Ed.) *Pilot Studies in Benchmarking Assessment Practice* (Gloucester, Quality Assurance Agency).