Conformity or Improvement?
Subject Benchmarking in Some Subject Areas
in Some Universities

Doctorate of Education
University of Leicester
School of Education

2004

Steve Pidcock
Acknowledgements

This study could not have been completed without the generous participation of those individuals who kindly consented to take part in the fieldwork. The promise of anonymity makes it impossible to thank them by name, but the expression of thanks is no less sincere for that. Warm thanks are also due to the Programme Committee of the Doctorate of Education at the University of Leicester under the chairmanship of Professor Les Bell for its flexibility and understanding which enabled me to complete my studies in spite of severe personal difficulties extraneous to the programme. Particular thanks go to my research supervisor, Dr Hugh Busher, for his patience, his persistence and his good humour.
Abstract

Steve Pidcock

Conformity or Improvement? Subject Benchmarking in Some Subject Areas in Some Universities

This study takes as its starting point the recommendation of the National Committee into Higher Education (1997) that the Quality Assurance Agency should establish a system providing benchmark information on standards operating within the qualifications framework in UK higher education. The introduction of subject benchmarking led to fears of increased external intervention in the activities of universities and a more restrictive view of institutional autonomy, accompanied by an undermining of the academic profession, particularly through the perceived threat of the introduction of a national curriculum for higher education.

The study situates subject benchmarking in the context of a potentially changing identity for UK higher education, which some observers see as dominated by a growing lack of trust in universities and university professionals on the part of government and articulated through the increasingly prevalent use of performance indicators requiring institutions and individuals to conform to externally imposed norms. The various requirements of the Quality Assurance Agency, including, therefore, subject benchmarks, are seen by some as threatening diversity by enforcing conformity.

After a consideration of the literature on various aspects of this context, and in particular of the benchmarking principle, the findings of an empirical investigation are presented, in which higher education professionals in chemistry, history and quality assurance were asked about their perceptions of subject benchmarking and its impact. The investigation did not bear out the fears articulated at the inception of subject benchmarking. Furthermore, the investigation showed that subject benchmarking was perceived as having none of the characteristics normally associated with a benchmarking system; in particular, it was not perceived as leading to improvement.

Finally, observations are offered on the way forward for subject benchmarking, and areas are suggested for further research.
Contents

1. Introduction 1

1.1 The context of the study 1
1.2 Theoretical perspectives on benchmarking 8
1.3 The purpose of the study 13
1.4 The importance of the study 15
1.5 The scope of the study 17
1.6 Thesis overview 18

2. Literature Review 20

2.1 Universities and the Evaluative State 20
2.1.i The purposes of universities and university education 20
2.1.ii Universities and accountability 23
2.1.iii Academic identity in universities 26
2.1.iv The evaluative state 29
2.1.v Synthesis 31
2.2 Standards 32
2.2.i The academic standards debate 32
2.2.ii Performance indicators 36
2.2.iii Synthesis 39
2.3 Quality 39
2.3.i Quality and quality systems 39
2.3.ii Quality and standards 47
2.3.iii Synthesis 49
2.4 Benchmarking 49
2.4.i Definitions and typologies 49
2.4.ii Benchmarking and quality enhancement 55
2.4.iii Benchmarking and conformity 60
2.4.iv The appropriateness of benchmarking 67
2.4.v Synthesis 71

3. Methodology 72

3.1 Research design 72
3.2 Validity and reliability 74
3.3 The sample 80
3.4 Ethical issues 84
3.5 Research instruments 88
3.5.i Interviews 88
3.5.ii Data gathering 91
3.5.iii Documents 93
3.6 Analysis of data 95
3.7 The study 99
4. Findings

4.1 The respondents

4.1.i Identifier codes
4.1.ii Academic departments
4.1.iii Administrative units

4.2 Documents

4.2.i Departmental documents
4.2.ii Institutional documents

4.3 The interviews

4.4 Synthesis

5. Discussion

5.1 The introduction of subject benchmarking

5.1.i Subject benchmarking
5.1.ii Threshold standards

5.2 Subject benchmarking and improvement

5.2.i Subject benchmarking and best practice
5.2.ii Subject benchmarking and dynamism
5.2.iii Subject benchmarking and quality enhancement

5.3 Consequences of subject benchmarking

5.3.i Benchmarking and academic identity
5.3.ii Benchmarking and universities

5.4 Benchmarking and conformity

5.4.i Cultures of conformity
5.4.ii Benchmarking and national curriculum

5.5 Synthesis

6. Conclusion

6.1 The research questions
6.2 The significance of the study
6.3 Limitations of the research
6.4 Subject benchmarking: the way forward
6.5 Future research

Tables

Table 4.1: Experience of Academic Staff
Table 4.2: How do you feel about the introduction of benchmarking?
Table 4.3: How do you react to the performance criteria and/or threshold standard?
Table 4.4: Do you think that benchmarking will encourage the dissemination of best practice?
Table 4.5: Do you think that it was necessary to introduce subject benchmarking?
**Abbreviations**

The following abbreviations are used both in the body of the text and in the references.

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>AUT</td>
<td>Association of University Teachers</td>
</tr>
<tr>
<td>BRTF</td>
<td>Better Regulation Task Force</td>
</tr>
<tr>
<td>CBI</td>
<td>Confederation of British Industry</td>
</tr>
<tr>
<td>CHE</td>
<td>Committee on Higher Education</td>
</tr>
<tr>
<td>CHEMS</td>
<td>Commonwealth Higher Education Management Service</td>
</tr>
<tr>
<td>CHES</td>
<td>Centre for Higher Education Studies</td>
</tr>
<tr>
<td>CNAA</td>
<td>Council for National Academic Awards</td>
</tr>
<tr>
<td>CVCP</td>
<td>Committee of Vice-Chancellors and Principals</td>
</tr>
<tr>
<td>DES</td>
<td>Department for Education and Science</td>
</tr>
<tr>
<td>DfEE</td>
<td>Department for Education and Employment</td>
</tr>
<tr>
<td>DfES</td>
<td>Department for Education and Skills</td>
</tr>
<tr>
<td>FEU</td>
<td>Further Education Unit</td>
</tr>
<tr>
<td>GSP</td>
<td>Graduate Standards Programme</td>
</tr>
<tr>
<td>HE</td>
<td>Higher Education</td>
</tr>
<tr>
<td>HEFCE</td>
<td>Higher Education Funding Council for England</td>
</tr>
<tr>
<td>HEQC</td>
<td>Higher Education Quality Council</td>
</tr>
<tr>
<td>HUDG</td>
<td>History in Universities Defence Group</td>
</tr>
<tr>
<td>LTSN</td>
<td>Learning and Teaching Support Network</td>
</tr>
<tr>
<td>NATFHE</td>
<td>National Association of Teachers in Further and Higher Education</td>
</tr>
<tr>
<td>NCE</td>
<td>National Commission on Education</td>
</tr>
<tr>
<td>NCIHE</td>
<td>National Committee of Inquiry into Higher Education</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>PI</td>
<td>Performance Indicator</td>
</tr>
<tr>
<td>QAA</td>
<td>Quality Assurance Agency for Higher Education</td>
</tr>
<tr>
<td>RSC</td>
<td>Royal Society of Chemistry</td>
</tr>
<tr>
<td>SCOP</td>
<td>Standing Conference of Principals</td>
</tr>
<tr>
<td>SRHE</td>
<td>Society for Research into Higher Education</td>
</tr>
<tr>
<td>THES</td>
<td>Times Higher Education Supplement</td>
</tr>
<tr>
<td>TQM</td>
<td>Total Quality Management</td>
</tr>
<tr>
<td>UUK</td>
<td>Universities UK</td>
</tr>
</tbody>
</table>
1.1 The context of the study

In 1997, the National Committee of Inquiry into Higher Education (NCIHE), under the chairmanship of Sir Ron Dearing, published its report *Higher Education in the learning society* (NCIHE, 1997). The members of the NCIHE were appointed 'with bipartisan support by the Secretaries of State for Education and Employment, Wales, Scotland, and Northern Ireland on 10 May 1996' (NCIHE, *op. cit.*: 1), the purpose of the committee being: 'To make recommendations on how the purposes, shape, structure, size and funding of higher education, including support for students, should develop to meet the needs of the United Kingdom over the next 20 years, recognising that higher education embraces teaching, learning, scholarship and research' (NCIHE, *op. cit.*: 3). Dearing himself (NCIHE, *op. cit.*: 2) observes: 'Much of our report is concerned with material things and with the central role of higher education in the economy. It would be surprising were it not so.' Nevertheless, in the committee's (NCIHE, *op. cit.*: para.5.11) view, the four main purposes
of higher education are:

- to inspire and enable individuals to develop their capacities to the highest potential levels throughout life, so that they grow intellectually, are well-equipped for work, can contribute effectively to society and achieve personal fulfilment;
- to increase knowledge and understanding for their own sake and to foster their application to the benefit of the economy and society;
- to serve the needs of an adaptable, sustainable, knowledge-based economy at local, regional and national levels;
- to play a major role in shaping a democratic, civilised, inclusive society.

These purposes are enunciated within the context of the committee’s (NCIHE, op. cit.: 3) terms of reference:

The Committee should have regard, within the constraints of the Government’s other spending priorities and affordability, to the following principles:

- there should be maximum participation in initial higher education by young and mature students and in lifetime learning by adults, having regard to the needs of individuals, the nation and the future labour market;
- students should be able to choose between a diverse range of courses, institutions, modes and locations of study;
- standards of degrees and other higher education qualifications should be at least maintained, and assured;
- the effectiveness of teaching and learning should be enhanced;
- learning should be increasingly responsive to employment needs and include the development of general skills, widely valued in employment;
- higher education’s contribution to basic, strategic and applied research should be maintained and enhanced, particularly in subjects where UK research has attained international standards of excellence or in Technology Foresight priority areas;
- arrangements for student support should be fair and transparent, and support the principles above;
- higher education should also be able to recruit, retain and motivate staff of the appropriate calibre;
- value for money and cost-effectiveness should be obtained in the use of resources.

Thus, immediately after the priorities of widening participation and student choice, comes a concern for the standard of higher education (HE) awards, particularly degrees. Aldrich (1996:43) observes:

it must be noted that assessment in higher education is ... causing considerable
concern. The rapid transition from an elite to a mass system has raised fears that standards have been reduced. In the late nineteenth and early years of the twentieth century, some university departments refused, on principle, in the first ten years to give a first-class degree, so that a reputation for rigour might be established. In spite of the massive increase in numbers of universities and of university students in the 1990s, such self-denying ordinances no longer obtain, and first-class degrees proliferate.

In a similar spirit, the NCIHE, after proposing a national framework for HE qualifications (NICHE, 1997: paras. 10.9-10.45 and Recommendation 22) recommends (NCIHE, op. cit.: Recommendation 24):

that the representative bodies and Funding Bodies amend the remit of the Quality Assurance Agency to include:

- quality assurance and public information;
- standards verification;
- the maintenance of the qualifications framework;
- a requirement that the arrangements for these are encompassed in a code of practice which every institution should be required formally to adopt, by 2001/02, as a condition of public funding.

The NCIHE goes on to recommend (NCIHE, op. cit.: Recommendation 25) that the early work of the Quality Assurance Agency (QAA) should include:

- to work with institutions to establish small, expert teams to provide benchmark information on standards, in particular threshold standards, operating within the framework of qualifications, and completing the task by 2000;
- to work with universities and other degree-awarding institutions to create, within three years, a UK-wide pool of academic staff recognised by the Quality Assurance Agency, from which institutions must select external examiners;
- to develop a fair and robust system for complaints relating to educational provision;
- to review the arrangements in place for granting degree-awarding powers.

Armstrong (1999: 21) explains a consequence of the NCIHE's recommendations on quality: 'The assurance that will be sought in the future is that local standards comply with external criteria.' Previously, there was a tacit assumption that everybody knew what a degree from a British university was worth (Brown, 1999a: 35): additionally, in a small,
elite HE system, standards could be implicit (Armstrong, 1999: 11). Any move towards an explicit codification of standards thus represents a significant policy shift in HE. To this is to be added, moreover, the fact that, in the more recent mass HE, governments are seeking to guarantee value for money and also provide public accountability (Armstrong, op.cit.: 8).

However reasonable such developments might appear at first sight, they do nevertheless have their detractors. Cobban (1999: 216), for instance, argues: 'Where such amorphous and elusive subjects as education and scholarship are concerned, the value for money concept is rendered meaningless when translated crudely into rigidly defined objectives or targets.' Similarly, the recently imposed exposure to external scrutiny might be seen not only as breaking into the traditional territory of professional groups (Middlehurst, 1992: 24), but also as threatening to undermine the 'ideology of professionalism' (Kogan, 1986: 152), explained (Kogan, op.cit., p.152) as the assumption that, in schools, for instance, 'teachers are doing a good job'. Tapper and Salter (1998: 22) raise further objections, identifying three very significant consequences of the Dearing Report: it represents another move towards the creation of a national curriculum in HE; it offers a more restrictive interpretation of university autonomy; and it represents a threat to the growing diversity of British HE.

Nevertheless, British HE institutions have no choice but to come to terms with the requirements on quality and standards arising from the NCIHE recommendations. John Randall (2000a: 2), the then Chief Executive of the QAA, states unequivocally: 'The academic review method embraces firmly the standards agenda set by the Dearing Report.' When Randall (op.cit.: 2) continues with the reassuring statement that academic reviewers,
'will be sensitive to the challenges that institutions will face in working with benchmark statements for the first time', any hint of flexibility in the system is quickly dispelled:

In the early stages of the new review cycle, reviewers will not necessarily expect that institutions will have adjusted their provision in the light of benchmark statements. However, they will expect to see evidence of the approach that institutions intend to take to addressing benchmark statements. As with the Code of practice, benchmark statements reflect existing good practice, and feedback on the draft statements suggests that institutions already find them a valuable point of reference.

(Randall, op. cit.: 2)

The clear implication of this is that, after 'the early stages of the new review cycle', academic reviewers will 'necessarily expect' institutions to have adjusted their provision 'in the light of benchmark statements'. If it is unclear how the boundary is to be set for 'the early stages' of the new review cycle, it seems abundantly clear that adjusting provision 'in the light of' benchmark statements is intended to imply that institutions will be expected, indeed required, to demonstrate a close correlation between the QAA's benchmark statements and their own subject provision. This impression is further reinforced by what Randall (op. cit.: 2) has to say about the qualifications framework:

Any changes arising from the qualifications framework will take time to introduce. If a programme leading to a named award does not meet the expectations for that award, it will be necessary either to upgrade the course content, or to award instead a qualification appropriate to the actual level of the programme. Such changes cannot be made overnight. In particular, the award for successful completion of a programme cannot be reduced for students already accepted to study on that programme. Accordingly, the Agency believes that any changes that are needed should be introduced progressively over the period from 2000 to 2006. Where a necessary change can be introduced reasonably quickly, this should be done, but if adjustments need to be made to longer programmes, most or all of the period suggested may be required for this.

In the light of such an approach, Randall's (op. cit: 2) assertion that institutions find benchmark statements 'a valuable point of reference' requires interpretation. On the one
hand, it is clearly likely to be of interest to HE institutions to know what similar institutions are doing in comparable areas. On the other hand, however, it may be that institutions are keen to know what agenda is being set by the QAA, in order subsequently to avoid critical QAA reports, with the attendant risks of unfavourable publicity, recruitment difficulties and loss of funding. Presenting the then ‘new approach’ (QAA, 1998b: 2) to quality assurance, the QAA (op.cit.: 3) begins, for instance, with a scarcely veiled threat: ‘Bodies that award qualifications in respect of secondary, further and vocational education are subject to statutory regulation. Higher education, which is largely self-regulating, must now be able to demonstrate no less rigour in safeguarding its own standards.’ For Diamantopoulos (2001: 16), this demonstration probably leads to academic departments feeling ‘forced to issue ... oppressive guidelines to their staff in order to avoid falling foul of the thought police (aka the Quality Assurance Agency)’.

In a subsequent presentation of the principles underlying the new approach, the QAA (1999c: paras.40-41) indicates its intention to apply a ‘lighter touch’ and ‘varying intensity’ (QAA, op.cit.: paras.42-51) to its review of subject provision. Nevertheless, in March 2001, Baty (2001b: 1) reports: ‘Education secretary David Blunkett confirmed that because of concerns that the [‘new approach’] framework will fail to deliver the promised “lighter touch”, he had told the QAA to cut inspections.’ A sense of the resentment of the QAA felt in some quarters at this stage is well conveyed by MacLeod and Elliot Major (2001: 11):

Chief executive John Randall was (coincidentally?) A few thousand miles away in India when the education secretary was persuaded to order a 40% cut in inspections.

... The question of who should be exempt from future visits is by no means settled, it seems. In the meantime, Randall has taken to quoting remarks about Marks and Spencer in the Observer. “They ... had been left for too long under the control of smug, arrogant managements who thought they could peddle any old tat as long as
it bore their illustrious names.” Remind you of anyone? He asks innocently.

The consultation document HEFCE 01/45 (HEFCE, 2001) issued jointly in July 2001 by the HEFCE, the QAA, UUK and SCOP introducing ‘a hands-off, audit-based approach to external scrutiny’ (Baty, 2001d: 3) is seen by Baty (op.cit.: 3) as ‘nail in coffin for TQA [Teaching Quality Assessment]’ and, moreover, ‘a climb-down by the QAA’ (op.cit.: 3). Just how significant this change of direction was seen to be is indicated by Baty (2001e: 1): ‘While lecturers’ leaders rejoiced at Tuesday’s [21 August 2001] news that Mr Randall had quit as chief executive of the Quality Assurance Agency, Mr Randall and his supporters warned that universities were winning too many concessions in their war against prescriptive regulation.’ On 31 August 2001, the Times Higher Education Supplement headed its front page: ‘QAA: THE WAY FORWARD. Take the screw out of scrutiny.’ As expressed by Foster-Carter (2001: 15): ‘What universities are afraid of is simple. It is that the bureaucratic Leviathan built up by control freaks such as the unlamented Randall, under phoney populist slogans of accountability, has become a monster destroying what it so intrusively scrutinises.’

Similarly, in presenting the quality assessment process for higher education in England (hereafter referred to as ‘the external review process’) resulting from HEFCE 01/45 and the responses to it, the QAA (2002a: para.3) asserts: ‘the future approach must be rigorous and robust in securing the accountability of HEIs’. The principal mechanism for assuring this is no longer a programme of subject reviews from 2002 onwards, which is cancelled (QAA, 2002d: para.2), but instead ‘a six-year cycle of institutional audits’ (QAA, 2002a: para.12). Power (1997: 143) warns: ‘audits may turn organizations on their heads and generate excessive preoccupations with, often costly, auditable processes. At the extreme,
performance and quality are in danger of being defined largely in terms of conformity to such process.' Baty (2001c: 3) identifies a similar problem with the QAA institution-wide audits: 'These will introduce about 200 new rules in 11 quality assurance codes of practice, impose a strict new qualifications framework, and will make universities conform to programme specifications and subject benchmarks.'

The QAA's (2002c: para.12) Handbook for institutional audit: England specifies that one of the things on which audit teams will focus is: 'the use made of external reference points, including the Code of practice, The framework for higher education qualifications in England, Wales and Northern Ireland (the FHEQ), and Subject benchmark statements'. The external review process thus reaffirms the significance accorded to Subject Benchmarking. In situating subject benchmarking within the quality agenda, it is important to consider the underlying concept of benchmarking.

1.2 Theoretical perspectives on benchmarking

As expounded by Alstete (1996), benchmarking is used by HE institutions in order to provide information for quality improvement by searching for best practices: it is, moreover, a collaborative venture. However, a formative evaluation (QAA, 1999b: para.8.10) of the benchmarking process undertaken in Law, Chemistry and History between April and December 1998 raises as an issue the fact that, '[t]he Benchmark information is not at present in a form that would assist Academic Reviewers to grade the performance of programme providers.' Such a view regards benchmarking as part of a system of performance indicators, thus undermining a collaborative paradigm. Indeed, as Smith et al. (1999: 1) explain, benchmarking information will help 'the new cadres of academic reviewers to make informed judgements in the setting of standards. These will enable
categorization, it is suggested, of UK departments into three categories: below minimum thresholds, exceeding minimum thresholds and achieving standards appropriate for the "typical" student.' Jackson (1999: xi) highlights a dichotomy to be found at the heart of benchmarking:

While the application of benchmarking methods in higher education might appear to be a recent phenomenon, the principles that underpin our desire to collaborate in order to learn about, evaluate and improve our own practice have been applied for a long time. What is new is the attempt to attach new meanings, develop new processes and reinterpret what has gone before in terms of two different notions of benchmarking, namely:

- developmental benchmarking - promoting best practice;
- regulatory benchmarking - assuring quality and standards.

Much of the conflict and confusion in discussions about the pros and cons of benchmarking for higher education relates to the inbuilt tension between these two different purposes.

The tension thus identified by Jackson is important, given the significance of subject benchmarking as one of the four elements (alongside the Qualifications Framework, the Code of Practice, and Programme Specifications) of the QAA's standards infrastructure (Wright and Williams, 2001: 11). Indeed, the QAA (Wright and Williams, op.cit.: 11) asserts of these four elements: 'They may be regarded as the most innovative and distinctive feature of the emerging UK quality assurance arrangements.' Particularly problematic is the improvement function of benchmarking when applied to academic disciplines. As Schofield (2000: 104) notes of benchmarking: 'To date most work has taken place in university management, but the greater test will be its use in enhancing the quality of teaching and learning and of research.' For Elton (1992: 5): 'Enhancement, when applied to quality, must be the objective of all in higher education'. Similarly, in a review of international practice, Harman (1998: 356) sums up: 'Quality assurance programmes often aim to serve a variety
of purposes but generally their primary purpose is a combination of public accountability, improvement and renewal.' It may seem unremarkable, then, when the Handbook for institutional audit: England (QAA, 2002c: para.1) begins with the statement: 'The mission of the Agency is to promote public confidence that the quality of provision and standards in higher education are being safeguarded and enhanced' (emphasis added). The QAA's (1999e: para.9.4) own advisory group on multidisciplinary and modular programmes, indicates how benchmarking can be used for improvement in HE: 'The enhancement model of benchmarking is one based on using internal and external points of reference related to the development of acknowledged best practice.' According to Jackson (2002: 139): 'QAA subject benchmarking provides a good example of systemic policy that is being introduced in good faith that it will ultimately enhance student learning because communities of subject-based practitioners will be clearer about the learning that is valued in and across subjects.' However, the contrast with the QAA's (2001a: 4) own rationale for subject benchmarking following the publication of the first 22 documents could hardly be starker: 'By describing the nature and standards of higher education programmes, benchmark statements provide the academic communities with a means of defining what can be expected of a graduate in a particular subject area.'

In its information for students, the QAA (2002e) explains that HE institutions will be using benchmark statements 'as a reference for writing programme specifications and for internal quality assurance purposes. The benchmark statements help to ensure that the standards of degree programmes across the UK meet an agreed level.' This minimalist view implies a predominantly control function for subject benchmarking. According to Randall (2000b: 11): 'there is general acceptance of the principle of scrutiny as a means of providing
accountability, generating public information and disseminating good practice.' Henkel (2000a: 98), however, reviewing quality assurance in HE 1983-1997, summarizes: 'Overall there was a strong sense among academics that while it was difficult to object to quality assurance policies, they were a product of major change in higher education and represented challenges to academic values, conceptions of higher education and self-esteem.' McLernon and House (2001: 15), for instance, complain that academics are now 'drowning in a sea of audits, assessments, reviews and feedback.' They (McLernon and House, op.cit.: 15) write of a new initiative, 'The Transparency Review', that future research will 'uncover how some cost-cutting civil servants and superannuated, pettyfogging bureaucrats devised an insane initiative', and wonder (op.cit.: 15) whether academics will now revolt. Their (McLernon and House, op.cit.: 15) conclusion is most revealing: 'Or perhaps, as in the past, we will just conform like reluctant sheep, while politely bleating our disapproval from the staff coffee rooms? - Bet you've already guessed the answer - BAAAAAAA!' (emphasis added). The despondent 'as in the past' suggests a feeling that academics and HE institutions in fact have no choice but to conform to externally imposed constraints.

Of the QAA and its 'labyrinthine torture' (Pratt, 2002: 25), Pratt (op.cit.: 25) notes: 'there are few incentives to innovate in the QAA's dour world of paper-based policing of procedures and programmes.' As has been seen, the augmented role for the QAA originates in the Dearing Report, of which Shattock (1999: 11) notes: 'It tended to emphasise control and bureaucratic tidiness over institutional vitality and creativity. This has particularly caused concern over matters to do with quality'. For Stevens (2002: 12): 'Dearing's report spelled the end of university autonomy.' In Alderman's (2002a: 15) view: 'The creeping
disempowerment of academic communities in Britain’s degree-awarding institutions is malignant in itself and in the consequences it has for teaching, learning, and the support of teaching and learning.' Such views may attributable to an ambiguity inherent in the QAA per se, Reuter (2001: 15), for instance, pointing to ‘the QAA’s inability to distinguish between regulation, regulatory enforcement and evaluation.’

For the QAA (2002a: para.15), one of the objectives of the external review process is: ‘to provide a means of securing accountability for the use of public funds received by HEIs.’ The introduction of benchmarking and threshold standards, together with associated elements of the NCIHE/QAA quality assurance model, such as programme specifications and the Code of Practice, can be seen to raise fundamental questions about the status and autonomy of UK universities and their academic staff, and about their accountability relationships both with society at large and with particular societal groups, such as students and employers. Alongside this, however, but intimately connected with it, are difficulties associated with using benchmarking as a device for quality enhancement. As Jackson and Lund (2000: 222) observe:

> It is difficult to promote trust in a methodology when an important motivation for state-inspired benchmarking is to target funding, reduce or redistribute resources, or produce league tables that affect the reputation, market position and profitability of an institution. The issue is generic to all HE systems that are attempting to use benchmarking in this way. It needs to be resolved in a positive way if benchmarking is to continue to flourish.

Moreover, this in turn highlights the significance of performance indicators for contemporary HE. Lund and Jackson (2000: 25) explain: ‘The creation and use of performance measures, capacities for self-evaluation and external review are features of the “evaluative state” ... which is essential for the management of public policy based on the idea of performance contracts ...’. As Kogan and Hanney (2000: 32) note, the changes
associated with the rise of the evaluative state 'move the state into a stronger regulatory position, and also close to boundaries of what had hitherto been academic mastery on (sic) the means of normative control.'

1.3 The purpose of the study
This study investigates reactions to subject benchmarking during the transitional phase from the initial promulgation (NCIHE, 1997) of the Dearing-inspired quality agenda for HE into the period introducing the first institutional audits (QAA, 2002a: para. 66). The purpose of the study is to gauge reactions both at the level of individual members of staff within academic disciplines and at institutional level.

Interest is focused on two broad areas, which are actually so intimately intertwined as to be practically inextricable from one another. One of these areas is concerned with the nature and outcomes of the subject benchmarking system itself. Brown (1999b: 278) points to the question 'concerning how far our efforts to benchmark students' achievements are focused on the potential benefits of improvement rather than just measurement.' A second area involves considerations of power, control and compliance. On sector fears about the possible imposition of a national curriculum for HE, Brown (1999a: 38), for instance, observes:

the Chief Executive of the QAA and the Director of Programme Review regularly refute the suggestion that the QAA has any wish whatever to introduce a national curriculum for higher education in the UK. Worries still remain that veiled hints about what will happen if the sector does not “sort itself out” on issues of standards and accountability might lead ultimately to yet tighter control, not only of how programmes are specified, but also what they should contain.

These interrelated elements provide the conceptual framework for the study which follows.

Within the broad framework of these two general areas, this study seeks to test the QAA's
(1999b: para.9) initial finding that the view taken of the benchmarking process by members of the academic community is 'broadly positive', to investigate the perception of links between benchmarking and the external review process, and to identify steps taken at subject and institutional level to accommodate the benchmarking and quality regime. It is possible that differing views will be taken in post-1992 universities, used to external scrutiny by, for instance, the Council for National Academic Awards (Pratt, 1997: 210), and in pre-1992 universities, with a long tradition of academic autonomy (NCIHE, 1997: para.3.62). Some may even feel that they are in a situation akin to that described by Cobban (1999: 65): 'From the early-fourteenth century newly qualified bachelors and masters, both in England and at Paris, had to declare publicly that they would not teach anything that was contrary to faith.'

A further consideration is the need for some subject areas to work within constraints imposed by external bodies for professional accreditation. It is possible that academic disciplines such as Law, which are used to seeking professional accreditation, will react differently to the introduction of subject benchmarking than will academic disciplines with no such tradition, such as History. If this is the case, then it may well be that academic disciplines such as Law will see the introduction of subject benchmarking and threshold standards as less of a threat to autonomy than will academic disciplines such as History.

Allied to such matters are questions of perceptions of academic status and institutional purpose. Cobban (1999: 64) remarks acerbically that the masters' guild in medieval English universities 'did not experience the humiliation of having to produce banal and disingenuous mission statements or of having thrust upon it the inappropriate verbiage of a mercantile society to name only two of the albatrosses that so depress modern British universities.'
Notwithstanding this, it is possible that a different, perhaps more instrumental view may be
taken of the new quality assurance model at the institutional level, where successful
benchmarking and conformity to the post-Dearing quality agenda may be seen more in the
context of published statements of performance indicators, and of continued public funding.

1.4 The importance of the study

makes politicians feel safe, but is a chimera of standardisation – which doesn’t make the rest
of us feel safe.’ Alongside misgivings about the pressure for conformity and the possible
steer towards a national curriculum for HE are considerations of the extent to which the
Dearing Report intervenes in “‘the private life’ of higher education’ (Trow, 1997: 26). In
spite of its (NCIHE, 1997: para.15.4) axiomatic respect for ‘academic freedom within the
law’, the report is intent on establishing mechanisms for the regulation and control of
academic activity, and ‘is written from outside the system, looking in’ (Trow, 1997: 26),
thus not reflecting ‘on the whole ..., a practitioner’s analysis of the “new higher education”
(Jary and Parker, 1998: 10). Considerations such as these lead Trow (1998: 116) to
comment on the Dearing Report’s ‘continuing resistance to an awareness of the diversity
and complexity of higher education, and its consequent commitment to dirigisme, to tight
controls from the centre’.

The post-Dearing quality agenda may also be seen as calling into doubt the perceived
identities of HE academics and institutions. As Henkel (2000b: 2) points out: ‘The idea of
identity is, arguably, central to individual academic lives and to the workings of the whole
profession.’ The new quality regime may, then, be viewed as occasioning a further ‘decline
of donnish dominion’ (Halsey, 1992). Similarly, the new quality agenda may be contributing
to a reassessment of universities and their position in contemporary society. The Dearing Report (NCIHE, 1997: para.5.10) declares: 'The aim of higher education is to enable society to make progress through an understanding of itself and its world: in short, to sustain a learning society’. Jary and Parker (1998: 10) comment: ‘In taking a wide-ranging, and generally humanist and progressive, view of the benefits of a further expansion of UK HE, the Dearing Report is very much to be commended’; but Barnett (1998: 15) objects that ‘there is no serious discussion of the learning society as such.’ Possibly more fundamentally, however, there is no attempt to examine ‘some of [the] core assumptions about what universities are for in the first place’ (Jary and Parker, 1998: 21). Such gaps are perhaps symptomatic of what Williams (1998: 1) sees as ‘the central weakness of Higher Education in the Learning Society - superficiality.’

Certainly, the Dearing Report has prompted the widest possible spectrum of reactions. In 1998, the Department for Education and Employment (DfEE, 1998:1) described Dearing as ‘a seminal report providing us with the basis for moving forward into a new era of funding and development for higher education and its relationship with lifelong learning as a whole.’ By 2000, however, Watson and Bowden (2000: 13) were noting: ‘At the mid-term stage, the prospects for a harmonious and effective sectoral re-organisation remain bleak.’ A more radical view is taken by Maskell (2001: 18): ‘No country whose standards were capable of being raised would ever have commissioned the Charter for Higher Education of 1993 or the Dearing Report of 1997. Or, if they had, they would have burned them (and their authors with them) as soon as they saw the results.’

Subject benchmarking is a radical new departure with potentially profound and far-reaching consequences for UK HE, raising questions about autonomy and control, the nature of
universities, and their relationships with society and the state. The importance of these issues is reflected in the literature, as is the significance of quality and quality management, particularly the pursuit of quality enhancement. Brennan and Shah (2000: 4) note: ‘Much of the literature on quality in higher education appears to lack any significant empirical basis.’ This study adopts an empirical approach in order to investigate how subject benchmarking and its impact are perceived and evaluated by those who are required to put it into practice.

As has already been seen, such a requirement is referred to in terms of compliance (Armstrong, 1999) or conformity (Power, 1997). Newby (1999: 263) suggests that compliance, if widespread, becomes conformity. No attempt is made here to establish such a hierarchy of terms: compliance and conformity are taken to be essentially equivalent, referring both to localised and to generalised occurrences of the phenomenon.

1.5 The scope of the study

The first three academic disciplines in which subject benchmarking was undertaken were Chemistry, History and Law (QAA, 1998a: Part IV, para.5). The QAA (op. cit.: Part IV, para.5) explains this choice as follows: ‘These three subjects have been identified as going some way towards representing the variety of higher education provision; they are also sufficiently discrete as subjects to provide a starting point for developing a means of identifying benchmark information and threshold standards.’ As material constraints on the study preclude covering all three subjects, Chemistry and History are selected in order to provide exemplars of arts-based and science-based subjects.

In this study, and following common parlance, pre-1992 universities are referred to as ‘old’ universities, and post-1992 universities as ‘new’ universities. Material constraints make it
necessary to limit the study to two English universities, one old (OldU) and one new (NewU).

For each institution, the primary source of data is a series of semi-structured interviews with academic members of staff and with representatives of central quality services. A total of 30 academics were interviewed, corresponding to approximately 30% of the full complement. For each department, this included an individual with particular responsibility for quality: with only one exception (OldU Chemistry), each Head of Department was interviewed. Additionally, interviews were conducted with two members of each institution’s quality services, one representing quality assurance, and one representing quality enhancement. Alongside these 34 interviews, access was also gained to a very limited amount of internal documentation.

Such a limited sample can, of course, make no claim to representativity; but it can provide valuable insights into the actualities of four departments currently operating within the national quality model. The study investigates reactions to subject benchmarking and a range of associated issues purely from the perspective of academic and certain administrative staff. Future research could profitably seek to elucidate the student perspective.

1.6 Thesis overview

After a discussion of methodological considerations, there is a review of the literature on the major themes to be investigated. Of particular importance is an analysis of benchmarking theory and its applicability to the academic activities of universities. The findings from the project fieldwork are then presented in detail and discussed within the
conceptual framework outlined above. This allows certain conclusions to be drawn, reflections on the outcomes of the project to be made, and areas for further research to be identified.
Chapter 2

Literature Review

2.1 Universities and the evaluative state

2.1.i The purposes of universities and university education

According to The Compact Edition of The Oxford English Dictionary (1971) a university is:

The whole body of teachers and scholars engaged, at a particular place, in giving and receiving instruction in the higher branches of learning; such persons associated together as a society or corporate body, with a definite organization and acknowledged powers and privileges (esp. that of conferring degrees), and forming an institution for the promotion of education in the higher or more important branches of learning; also, the colleges, buildings, etc., belonging to such a body.

Newman (1982: 115), writing in 1852, explains:

This process of training, by which the intellect, instead of being formed or sacrificed to some particular or accidental purpose, some specific trade or profession, or study or science, is disciplined for its own sake, for the perception of its own proper object, and for its own highest culture, is called Liberal Education; ... And to set forth the right standard, and to train according to it, and to help forward all students towards it according to their various capacities, this I conceive to be the business of a University.
For Flexner (1930: 23): ‘Intensive study of phenomena under the most favourable possible conditions – the phenomena of the physical world, of the social world, of the aesthetic world, and the ceaseless struggle to see things in relation – these I conceive to be the most important functions of the modern university.’ Not belonging to the university are (Flexner, op. cit.: 27): ‘Assuredly neither secondary, technical, vocational, nor popular education.’

A change in outlook by the second half of the twentieth century is identified by O’Hear (1988: 9): ‘It is no doubt unfashionable to see the aim of any sort of education as the cultivation in the minds of the initiated of a sort of wisdom, as opposed to skills, techniques, information or money-earning capacities.’ Nevertheless, for Halsey (1992: 17): ‘In the most general terms the university is the realization of a single idea - the idea of a social institution to ensure the continuity of intellectual work.’ In Evans’s (1999: 2) view: ‘The training of minds is, perhaps, the supreme distinctive task of universities.’ Maskell and Robinson (2001: 171) add a further dimension: ‘Education has to be a form of the pursuit of the good.’ It may seem obvious to state that ‘the primary purpose of universities is advanced education’ (Ball, 2001: 13); but that raises questions about the constitution, organization and objectives of that education, and thus also of universities.

Writing in 1963, Kerr (2001: 54) notes: ‘there are only about thirty universities in Great Britain and it is clear what is and what is not a university.’ Thirty years later, King (1995: 16) refers to ‘established cultural notions as to what real universities do’, going on to list as characteristics of ‘the traditional idea of a university’ (King, op.cit.: 17):

- social, democratic, individual, and (not just) economic purposes;
- the notion of academic freedom and the ability to test and criticize existing and new ideas;
- comparability if not identicality of degree standards, and assurance through peer
review;

the view that academic issues, such as academic standards, the academic acceptability of courses, and the quality of research output, are best dealt with by academics; administrators (now increasingly managers) simply match resources and ensure that other institutional goals are compatible with basic academic objectives (Sutherland, 1993 [recte, 1994]).

The HE environment is changing, however, as Kogan (2000: 208) observes: 'Universities – originally organised so as to assert minimal influence over the disciplines - [now] have a whole range of institutional roles concerned with the advancement of what is essentially social and public policy - new employment legislation, collective research work directed to economic development, quality assurance, equal rights, internationalisation, and relations with the community.'

A feature of these new preoccupations is the shift from elite to mass higher education (Trow, 1973), about which more will be said later. Additionally, as Jackson (R.) (1999: 98) explains: 'The universities have moved from a position in which they were private, grant-aided institutions trusted to develop their own idea of what they should be, to a position in which they are seen as public agencies paid to deliver the services wanted by government through processes increasingly prescribed by the state.' In Ozga's (2000: 24) view: 'in the 1990s, the main function of education is the service of the economy.' The overall development is summed up trenchantly by Smith (D.) (1999: 148): 'Changes to universities have been faster and more overwhelming in the second half of the twentieth century than at any time previously.'

And the turmoil continues into the twenty-first century, MacLeod and Elliot Major (2002: 9) reporting: '[Higher Education Minister Margaret] Hodge said [in April 2002] there would be a different pattern of higher education in 10 years' time, with a blurring of the
boundaries between universities, further education colleges and even schools.' Only one week after this pronouncement, Sir Howard Newby (quoted in Goddard and Thomson, 2002: 1), Chief Executive of HEFCE, stated: 'The role of higher education institutions has changed strikingly. They are not just doing teaching and research but knowledge transfer, tackling social exclusion and contributing to local regeneration.' Small wonder, perhaps, that Barnett (1997: 1) declares: 'We have lost any clear sense as to what a university is for in the modern age', or that MacFarlane (1999: 145) sees the name 'university' persisting as 'an increasingly meaningless label'. For Education Secretary Charles Clarke (quoted in THES reporters, 2002: 1): 'For 35 years, there has been a sense of drift in the definition of a university.' In an open letter to the new Minister for Higher Education, Alan Johnson, Bassnett (2003: 13) advises: 'you’ll need to do a bit of thinking about what universities are in 2003. Those of us who work in them get so many conflicting messages, we have trouble facing ourselves in the mirror.'

Smith and Langslow (1999: 7) point out: 'The expectations of what universities will provide ... vary with the different interested parties.' For Williams (2002: 14):

There needs to be open public discussion about what universities are for, the proper relationship between higher education and society, what kind of accountability is needed, whether academic freedom and autonomy matter and how far the paymaster can, or should, call the piper's tune. Dearing did not begin to tackle these questions.

2.1.ii  Universities and accountability

Kogan (1986: 25) defines accountability as: 'a condition in which individual role holders are liable to review and the application of sanctions if their actions fail to satisfy those with whom they are in an accountability relationship' (emphases removed). Henkel (2000a: 69) demonstrates that 'by the early 1980s government was making it clear that universities
should not be excluded from its drive to tighten accountability in all public services.'

The accountability relationship between universities and the State inevitably has consequences for institutional autonomy, a characteristic of universities acknowledged by Robbins (Robbins Report, 1963: para.15), Jarratt (CVCP, 1985: para.2.4) and Dearing (NCIHE, 1997: para.15.4). As Tapper and Salter (1995: 69) remind: 'university autonomy has always been conditional upon the political context.' Robbins (Robbins Report, 1963: para.725) observes: 'there may be some who still believe that is an outrage if the role of the State or its organs is anything but completely passive; but they are not many. Nevertheless the whole question of control is understandably the subject of widespread apprehension.' It is to be expected that this apprehension is now even greater, given that, '[o]ver the past twenty years, the balance of power between the state and the institutions of higher education has been progressively redefined in favour of the former' (Tapper and Salter, 1998: 22).’ Bassnett (1998: v), for instance, complains of ‘the seemingly endless process of centralising higher education and restricting the freedom of universities.' Similarly, Henkel (2000a: 62) writes of ‘[n]ational policy change, with its emphasis upon accountability and demonstrable performance rather than trust in professional self-regulation'.

In a review of accountability prepared for the HEFCE, PA Consulting (HEFCE, 2000: 12) observe: ‘An effective accountability regime is one that assures public interests in the probity and performance of publicly funded activities, while minimising the burden on service providers’ (emphasis added). PA Consulting’s (HEFCE, op.cit.: 9) starting point is stated as follows: ‘Universities and other higher education institutions have expressed concerns that they face a growing “burden” of accountability - manifest in terms of external data requests, inspections, audits, funding bids and submissions – which is imposing costs
they would not otherwise incur and which is not justified by the benefits achieved’ (emphasis added). The conclusion is that ‘institutions’ perceptions of significant and growing accountability demands are substantiated’ (HEFCE, op.cit.: 3). Scott (2001: 10), for instance, records that, after ‘the ideological revanche of Thatcherism ... the remorseless bureaucracies of academic accountability - the Research Assessment Exercise ... and Quality Assurance Agency - further eroded those crucial margins of irresponsibility and creativity.’

The demands of external quality assessments are thus a particularly burdensome aspect of the accountability relationship. Even if, under its new Acting Chief Executive, Peter Williams, the QAA began to adopt a more conciliatory tone (Williams, 2001), the emphasis on accountability remains unchanged. Williams (op.cit.: 2) explains the role of external quality assurance as ‘first and foremost, to check that institutions are running their academic affairs in a way that can command public confidence. This is the accountability role, whose integrity cannot be jeopardised.’

According to Randall (quoted in Canovan, 2001: 8): ‘The public in this country is in no mood to take on trust the services provided to it by professional people. Accountability is high on the agenda for us at the moment’. For Russell (2001: 14), however: ‘Accountability, like liberty, is a hurrah-word, and the mere use of it tends to convince that the thing described is good.’ Finch (1997: 155) argues that ‘where intellectual life is vested in institutions which are in receipt of public funds, there is a perfectly proper interest in public accountability.’ On the other hand, Alderman (1996: 192) warns: ‘Academic autonomy is the lifeblood of higher education. It must not be stifled in the name of university politics or sacrificed on the altar of public accountability.’ On the basis of
empirical research carried out in England between 1994 and 1997 and covering seven disciplines, including chemistry and history (Henkel, 2000a: 23), Henkel (op. cit.: 97) reports broad acceptance of public accountability by academics, without there being, however, wide support for quality assurance policies: ‘Views about the context in which they should be understood were often critical, and objections to the way in which the policies were implemented revealed that academic acceptance of the principle of public accountability was more equivocal than at first appeared’ (Henkel, op.cit.: 97). Similarly, Morley (2003: 55), in a study covering a wide range of subjects, but not including chemistry or history (Morley, op.cit.: xi), notes: ‘academics and managers tended to support the concept of accountability but found the cumbersome bureaucracy requirements unacceptable.’ For Shore and Wright (2000), however, the impact of audit procedures on university culture is ‘to engender a coercive type of accountability’, in which ‘non-compliance is not an option’ (Shore and Wright, op.cit.: 76).

Jackson (2000b: 32) states that academic benchmarking is ‘a product of this more dynamic, competitive and publicly accountable environment.’ In the context of the present study, it is important to note that, as one of the four elements of the standards infrastructure (Wright and Williams, 2001: 11), subject benchmarking is a key component in the QAA’s conception of its accountability role (Williams, 2001: 2). Perceptions of subject benchmarking and its impact on higher education are thus likely to play a significant part in shaping the internal policies of institutions and the self-image of individual academics.

2.1.iii Academic identity in universities

Henkel (2000a: 98), in the research quoted in the last section, notes: ‘Overall there was a strong sense among academics that while it was difficult to object to quality assurance
policies, they were a product of major change in higher education and represented challenges to academic values, conceptions of higher education and self-esteem.' Various sources of these challenges can be identified. The PA Consulting (HEFCE, 2000: para.5.2.4) study of accountability in higher education, for instance, having considered the financial costs of accountability arrangements, also identifies behavioural costs, such as the need to work long hours to fit in accountability-related work, this in turn leading to an impairment of teaching and, in particular, research: 'At the same time, teaching and research assessments put academics’ professional reputation and even their careers at risk. It is not surprising then that quality assessments are known to affect staff morale '.

Furedi (2001: 4) identifies a further source of challenges to academic identity: ‘Through its activities and its advocacy of a managerial ethos, oriented towards a skills-focused system of education, the [QAA] affected every aspect of university life.’ The collegiality-managerialism debate is long-standing and has generated an extensive literature (Dopson and McNay, 1996). At its heart is the idea originating in the Middle Ages of the university as ‘a self-regulating guild’ (Cobban, 1999: 231). Halsey (1992: 54) acknowledges ‘the power of inherited ideology’; similarly Barnett (1990: 100): ‘the idea of an academic community, with its connotations of sharing and common commitment, dies hard.’ Kogan (2000: 209), however, warns of the use of the word ‘community’ in discussions of higher education: ‘it is a warm glow word which implies good relations when they may not exist or may not even be necessary for good working.’ Nevertheless, ‘managerialism is inimical to collegiality, the traditional ideal of management of higher education institutions’ (Pratt, 1991: 4). Kogan and Hanney (2000: 199) draw attention to ‘the issue of the extent to which the essentials of higher education work make it difficult to yield to managerialist values'.
Moreover, as Kogan (2000: 210) observes: ‘The constituents of professionalism sit at odds ... with the kind of behavioural steers and prescriptive frames laid down by recent government-inspired actions and policies.’

Blake et al. (1998: 32) add a further dimension when they argue that the managerial practices of quality control inform quality assurance systems which ‘have fostered our propensity to police ourselves: they are part of a more pervasive culture of surveillance.’ More will be said on this in the section on standards.

An additional challenge to academic identity comes from the shift from elite to mass education. Halsey (1992: 13) is forthright: ‘Mass higher education still strikes the British academic ear as a self-contradictory absurdity.’ For all that, only five years later, Henkel (1997: 142) stresses: ‘Higher education in Britain is going through a process of fundamental change, in that it is now making a decisive transition from an elite to a mass system, at the same time as relationships between the state and public services are being radically reappraised throughout the Western world.’ Henkel (op. cit.: 142) observes: ‘Academics ... on the face of it accept the necessities of massification and of public accountability but not many of the consequences as they are at present being defined. They are struggling to hold onto [sic] the values and modes of working that belonged to an elite system’.

Evans (2002: 14) warns that with the culture of accountability in public life goes another shift: ‘the growth of a culture of compliance’, which has generated (Evans, op. cit.: 14) an ‘Identikit model of the academic’. If there has been a ‘shift in expectations about the role and performance of academics’ (Evans, op. cit.: 14), the worst effect has been ‘the increasing difficulty of dissent from these expectations’ (Evans, op. cit.: 14).
The dilemma of academic identity is encapsulated by Court (2001: 12): ‘Changes in higher education and society have dramatically altered the way the world sees academics, and the way academics see themselves.’ This study seeks to assess the part played in this by the introduction of subject benchmarking.

2.1.iv  The evaluative state

Neave (1988: 7) observes: ‘The cultivation of quality, efficiency and enterprise - the present-day version of governmental ‘New Theology’ - are not in themselves matters on which any right-minded person would dispute.’ Controversy arises, however, in ‘their operational, legislative and financial consequences for the individual establishments that go to make up the nation’s system of higher education’ (Neave, op.cit.: 7). According to Neave (op.cit.: 7): ‘policies which were first developed as an empirical, short-term response to financial difficulties at the start of the decade have now assumed a long-term strategic thrust’, which, in terms of the relationship between higher education and government, can be interpreted as the ‘Rise of the Evaluative State’ (Neave, op.cit.: 7). Neave (op.cit.: 8) reminds that the notion of the Evaluative State is not entirely new: ‘From the time that governments assumed the task of underwriting the finance of higher education as well as defining the legal and administrative framework within which that institution evolved, evaluation has been an important though often implicit aspect of that responsibility.’ As Kogan and Hanney (2000: 32) explain, however: ‘with the rise of the Evaluative State ..., academic judgements are incorporated into a rigorously administered public evaluative regime which affects reputations and the allocation of research monies and includes the publication of teaching quality gradings.’ In a reconsideration of the Evaluative State, Neave (1998b: 278) points out: ‘The ability to modify the rules of the game and, at the
same time, to require compliance – or to confer trust – are very real expressions of power.’

Kogan and Hanney (2000: 31) argue that the last two decades of the twentieth century were characterised by ‘the attempts known as New Public Management ... to both decentralise and to substitute more direct means of control by the package of instruments associated with the rise of the Evaluative State. In the context of British higher education, the total effect was, however, more centralisation.’ Essentially: ‘The logic of the Evaluative State emphasises the importance of output measurement as a tool for improving public services’ (Dill, 1998: 372). The implications for higher education are explained by Kogan and Hanney (2000: 32): ‘These changes move the state into a stronger regulatory position, and also close to boundaries of what had hitherto been academic mastery on the means of normative control.’

Neave (1998a: 132) draws attention also to the presence of ‘parastatals’, including the QAA, which ‘are presented as bodies independent from government’ (Neave, op. cit.: 132), and the presence of which ‘allows a distinction to be drawn between state administration - Ministries and Departments of State, taken up with issues of strategy – and “independent” agencies of execution, verification and control, the home of the technocracy of provable efficiency’ (Neave, op. cit.: 132).

An additional manifestation of the Evaluative State may be seen in the QAA Code of Practice, which was for Dearing (NCIHE, 1997: para.10.68) a pre-condition of a lighter touch to quality assessment, and publication of which began in January 1999 (Williams, 1999: 9). In the QAA’s (1998a: Part VI, para.5) own presentation, the codes ‘will indicate what an institution will be expected to be able to demonstrate through its own relevant
quality assurance mechanisms.' On initial reactions to the Code, the QAA (1998c: 21) concedes: 'Some institutions thought that a formal requirement to adopt the QAA Code posed a threat to their autonomy, and that it would be more appropriate if they were to be asked instead to 'agree to adhere' to it. The Agency accepts this point of view.' The perceived threat is scarcely removed, however, as the QAA (op. cit.: 21) continues: 'It will be for the institutional review process to come to a view on whether 'non-adherence' in any particular aspect or instance is or is not diminishing that security [of quality and standards] to an unacceptable extent.' As Baty (2001a: 8) reports on the publication of one of the later sections of the Code: 'Failure to adhere to any of the precepts could lead to public criticism, and is ultimately linked to a university's public funding.'

A new, and as yet untested, element in the relationship between higher education and the Evaluative State is the introduction of subject benchmarking: 'The progressive emphasis on performance measurement (through PIs), the growth in external review and institutional self-evaluation and now the expansion of benchmarking in UK HE is (sic) a direct response to [the new public management] and the creation of the evaluative state' (Jackson and Lund, 2000: 221). This study seeks to shed some light on perceptions of that relationship.

2.1.v  Synthesis

The introduction of subject benchmarking is set in a context of rapidly changing conceptions of the nature and purpose of universities, accompanied by a shift from elite to mass university education. Commentators note the increasing power of the state over universities and the growing emphasis on the accountability of universities rather than on trust in professionals. This realignment of the power relationship between universities and the state creates the potential for restricting the scope for creativity and autonomy in
universities. Studies suggest that academics accept the principle of accountability, but not
the burdensome bureaucracy associated with it. Moreover, the increasing importance of
quality assurance policies is seen as reflecting a fundamental change in HE, calling into
question the very nature of academic identity. The growth of managerialism in HE and the
emergence of a culture of accountability are seen as generating a culture of compliance in
universities. For some, these phenomena indicate the growing power of the evaluative state
over universities.

This study seeks to gain insights into the extent to which subject benchmarking is perceived
as symptomatic of and/or contributing to the debate surrounding, on the one hand, the
relationship between universities and society, and, on the other hand, the interpretation of
academic identity.

2.2 Standards

Whilst it is not the purpose of this study to engage in a statistical analysis of degree results
over time, between universities or across subjects, the issue of academic standards is clearly
central to the debate surrounding subject benchmarking and its introduction.

2.2.1 The academic standards debate

For Trow (1987: 268): ‘the issue of academic standards ... is one of the most important
issues that an institution has to confront.’ Sutherland (1994: 11) warns: ‘The concept of
standards is important in a variety of ways. The first is the formal responsibility which
universities have ... of setting the standards or output measures of what counts as degree-
level work. We neglect this at our peril.’ Difficulties arising in the late twentieth century are
identified by Silver and Williams (1996: 27): ‘When higher education was small enough, the
possibility and validity of comparisons were implicit in the external examiner process'. That process, through the interaction of internal and external examiners, led to a consensus on standards, which in turn was legitimised by the external examiner system: ‘and hence there was a notion that academic standards were national. However, higher education has grown and changed, and questions about academic standards and their comparability - the consensus - are being openly confronted’ (Silver and Williams, op.cit.: 28, emphasis in original). Middlehurst (1996: 72), too, stresses that, by the mid-1990s, recent expansion and diversification of the higher education system had called into question ‘the belief that all degrees in the UK reflect broadly comparable levels of intellectual development and achievement.’

Brown (R.) (1997: 273) identifies the then Secretary of State for Education and Science, John Patten, as the first person to voice publicly worries about standards, following visits to Malaysia and Singapore, ‘where he was asked whether quality had been maintained in the recent rapid expansion of student numbers. At a speech to the HEFCE conference in April 1994 Mr Patten asked HEQC to “place more emphasis on broad comparability in the standards of degrees offered by different institutions” and he made very clear the importance he attached to this (DES 1994).’ From this emerged HEQC’s Graduate Standards Programme (GSP) (Brown, op.cit.: 274).

The definition of academic standards adopted by the GSP (HEQC, 1997: para. 4.1) in 1994 is: ‘explicit levels of academic attainment that are used to describe and measure academic requirements and achievements of individual students and groups of students’ (emphases removed). In its Final Report, the GSP (HEQC, op.cit.: para. 5.5) states: ‘The GSP has shown that there is no general consensus, either within or outside higher education, that UK
degrees are broadly comparable with one another in terms of an equivalence of output standards.’ A similar conclusion is reached by Dearing (NCIHE, 1997: para.10.57), who also (NCIHE, op.cit.: para.3.66) identifies differing views on standards: ‘Although most institutions believe national standards should be maintained, there are some, particularly among the pre-1992 universities, who argue that standards differ across the sector and that this is an inevitable consequence of a mass system of higher education which should be formally recognised.’ By contrast, Scott (2003b: 14) asserts: ‘while all degrees are not the same, the range between the best and the worst is far narrower than in almost every other country.’ Lomas and Tomlinson (2000: 138) demonstrate that ‘there is a diverse range of opinion about standards, both within and between the different types of HEIs and it would appear that the “more means worse”/”more means different debate” will continue unabated.’

Brennan (1996: 12) notes for ‘the term “standards” in higher education … a bewildering diversity of usage.’ The QAA (2003a: 2) defines academic standards as: ‘predetermined and explicit levels of achievement that must be reached for a qualification to be awarded.’

For Lucas and Webster (1998: 105): ‘The maintenance of standards through time and throughout the system is axiomatic to British education.’ It is, however, ‘evident that confidence in the maintenance of standards and comparibility of awards between institutions is weak and probably weakening’ (Lucas and Webster, op.cit.: 105). Indeed, Thomson (2001: 8) reports: ‘More than half of academics think higher education expansion should be halted because academic standards are suffering.’ A note of caution is introduced by Elton (1998: 38), who draws attention to the distorting influence of coursework assessment: ‘To that extent[,] the upward drift [of degree results] is essentially an artefact,
resulting from the naive belief of academics that one can change the assessment instrument without changing both what it measures and also the degree boundaries.' Similarly, Haines (1999: 136) points out that recent changes in the degree regulations of many universities 'have sometimes increased the ease with which higher degree classifications can be obtained.' In the case of chemistry, however: 'The percentage increase (from 40.7 to 44.5 per cent) in first and upper second class degrees awarded [between 1987 and 1997] is probably much less than most people would have expected and certainly not in itself sufficient cause for alarm about falling standards' (Haines, op. cit.: 137). Nevertheless, the QAA (1998b: 11) asserts: 'Public concern that the transition to mass higher education may have resulted in erosion of degree standards needs to be addressed, and subject benchmarking has the potential to do that.' Additionally, Yorke (2000: 83) reminds: 'Ultimately, the standards that students achieve depend, of course, not only on the quality of the programme but also, and crucially, on their own commitment to their studies.'

Nevertheless, for Scott (2001: 10), British higher education 'still embodies a commitment to excellence and high standards which extends far beyond a dying nostalgia for elitist values - a higher education that acts mass but thinks elite.' Similarly, the Acting Chief Executive of the QAA (Williams, 2001: 1) states: 'Despite some claims to the contrary, our higher education system is among the most coherent and consistent, in terms of academic quality and standards, of any in the world.' The leader writer of the THES (anon., 2002: 16) complains that an announcement just made by the HE minister 'still contains phrases about raising standards in higher education that manage to imply that these are lacking, despite all the evidence to the contrary.'

An important feature of Dearing's (NCIHE, 1997: para.10.9) approach to standards is 'a
framework of qualifications broad enough to cover the whole range of achievement, consistent in terminology, and well understood within and outside higher education.' Hence: 'The introduction of a framework of qualifications underpins the development of robust and recognised standards' (NCIHE, *op.cit.*: para.10.54). Kenyon (1997: 19) comments: 'It is the link between qualifications and standards which arouses the sensitivities and, dare I say, brings out the prejudices. But we do have to make this connection; it is demanded of us.'

Haines (1999: 137) cautions that, whilst there has not been perfect calibration of standards within and between subjects in UK universities:

> it is worth considering how much the perception of standards ... has been affected negatively by the publication of data, including assessment of quality, that were subsequently used in inappropriate ways. The use of such data in the production of league tables and their publication – with little or no explanation or interpretation – was bound to suggest a wide variability in standards.

According to the QAA (1999a: 4), subject benchmarks 'should enable broadly comparable standards of attainment to be identified.'

2.2.ii Performance indicators

Cave *et al.* (1997: 24) define a performance indicator (PI) as:

> a measure - usually in quantitative form - of an aspect of the activity of a higher education institution. The measure may be either ordinal or cardinal, absolute or comparative. It thus includes the mechanical applications of formulae (where the latter are imbued with value or interpretative judgements) and can inform, and be derived from, such informal and subjective procedures as peer evaluations or reputational rankings.

According to Gaither *et al.* (1994: 52): 'The introduction of performance indicators to the lexicon of higher education management in Great Britain marked a major shift in power that welcomed a conservative ideology to government in 1979. Government’s perspective on higher education became one of “value for money”, emphasizing economy, efficiency, and effectiveness'. In the move from élite to mass HE, British governments 'are likely to
demand more information to measure value. The academy now finds itself competing for resources not only from government, but also from a marketplace filled with skeptical consumers' (Gaither et al., 1994: 59). Scott (2003a: 14) observes: 'If it moves, measure it—this seems to be the new mantra for the public services. Universities are no exception.'

Comparing institutions is not new. Lucas and Webster (1998: 105) observe: 'There has long been a rough and ready hierarchy of universities in the public mind, one which has readily translated into a presupposition that a Second from Oxford is superior to the same degree from Strathclyde'. A similar outlook is detected by the GSP (HEQC, 1996: para.10.12.2): 'Many [academics] argued that global comparability [in the standards of UK degrees] was a polite myth, and agreed with the comment that “those who were streetwise knew which universities had credibility in which subjects”.' For Day (1994: 27) 'it is essential for employers to create and operate dispassionately, over time, a rankings list of universities based on cumulative knowledge of the graduates hired.'

Halsey (1992: 58) notes: 'Of course, “wherever two or three are gathered together” there shall also be a sociological commonplace: invidious comparisons will emerge.' Comparing institutions, however, is not easy (Woolf, 1999: 101), and performance indicators need to be interpreted ‘with care and with an understanding of how modern higher education works' (Brooks, 2000: 16). Brown (1999b: 275) notes: 'It is becoming clear that funding bodies, academic leaders and quality managers (though not always academics themselves) are keen to seek measures of comparability between academic programmes and institutions, while wishing to avoid crude measures, such as league tables, that provide incomplete and often misleading pictures.' For Margham and Jackson (1999: 104), however, 'there is clearly a “reputational” hierarchy table of institutions made explicit by league tables and
such organizations as the Russell Group’; and, as King (2002: 14) observes, universities are ‘reputation-maximisers’.

PA Consulting (HEFCE, 2000: para. 5.2.4) record: ‘Excessive hours are, in part, a product of the “excessive” preparations for teaching assessments … The publication of league tables makes academic staff feel compelled to work for the highest possible score in terms of the assessment criteria, whether or not this leads to better quality teaching.’ In Belsey’s (2000: 16) view: ‘The early years of the 21st century were the great age of league tables, which were generally invested with magical powers.’

Strathern (2000: 310) identifies a particular dichotomy: ‘Higher education professionals at once accede to the idea of accountability and regard performance indicators as highly constructed and artificial means of measuring real output.’ Allied to this is the misgiving underlying Yorke’s (2000: 70) observation: ‘One of the problems with performance indicators is that they may be used for purposes for which they were not designed.’ Cave et al. (1997: 3) note of the process begun in the 1980s and leading to the explicit introduction of PIs into higher education in the UK: ‘At stake were issues of institutional autonomy, governance and control of the aims, objectives and evaluative criteria and mechanisms of higher education.’

It is against this background that the present study considers subject benchmarking. The QAA (1998b: 8) states that national subject benchmarking would make it possible ‘to form a view on whether satisfactory provision was at the level of the benchmark, above that level, or significantly exceeding it.’ This interpretation makes subject benchmarking a form of PI, and as Lund and Jackson (2000: 25) comment: ‘The creation and use of performance
measures, capacities for self-evaluation and external review are features of the "evaluative state".

2.2.iii Synthesis

The expansion and diversification of HE has raised doubts in some quarters about the comparability of standards on undergraduate degrees. Subject benchmarking was introduced as a result of Dearing's (NCIHE, 1997) concern about standards, and was established against a background of the growing use of PIs in HE. This study seeks to situate subject benchmarking within the standards debate.

2.3 Quality

2.3.1 Quality and quality systems

As Ball (1985: 96) notes: "'Quality in education' is a subject extraordinarily difficult to come to grips with, and full of pitfalls. There is no single final answer to the quality question, and we should not look for it. But the issue cannot be avoided.' Indeed, the importance of quality in HE is emphasized by Sallis (1993: 11): 'Quality is at the top of most agendas and improving quality is probably the most important task facing any institution. However, despite its importance many people find quality a most enigmatic concept.' There is broad agreement that the concept of quality is elusive (Green, 1994: 12), slippery (Walsh, 1994: 52), controversial (CHES, 1994: 18), problematic (Liston, 1999: 4), or yet to be defined (Houston and Studman, 2001: 476). Nevertheless, Goodlad (1995: 8) refers to 'the obsession with quality', and for Morley (2003: vii): 'Quality has become a universalizing metanarrative.' Harvey et al. (1992: 7) point out: 'Definitions of quality vary and, to some extent, reflect different perspectives of the individual and society. There appears to be no consensus about which definition should be used in higher education'. As
Watty (2003: 218) notes: 'where there are different conceptions [of quality] between stakeholders and within stakeholder groups, there is potential for conflict.' The QAA (2003a: 2) defines academic quality as: 'the effectiveness of procedures and provisions that enable students to achieve a qualification', a definition throwing no light on the nature of the 'procedures and provisions' in question.

Quality models and approaches originate in the world of business (FEU, 1995: 4; Liston, 1999: 8; Sallis, 1993: 45). Zairi (1996: 1) explains that '[t]he quality concept in its completeness as a philosophy of managing business organizations was first associated with the Japanese in the 1960s', not coming to Europe and the USA until the 1980s. For Srikanthan and Dalrymple (2002: 216), however, the attempt to implement across all the operations of a university quality management models as practised in industry 'is flawed in view of their tenuous fit with the core operation: education.' Three 'models of quality' (Sallis, 1993: 45) are provided by W.Edwards Deming, Joseph Juran and Philip B.Crosby.

Deming (1986: 23) proposes 14 points as 'the basis for transformation of American industry', the first point being (Deming, op.cit.: 23): 'Create constancy of purpose toward improvement of product and service, with the aim to become competitive and to stay in business, and to provide jobs.' A problem with this approach is that it is difficult to define the product of higher education (Sallis, 1993: 28). Moreover, as Bruneau and Savage (2002: 2) ask: 'How does one measure the intangibles of education - critical thinking, creativity, tolerance, wisdom?'

Juran (1989: vi) sets out to 'provide companies with the strategies needed to attain and hold quality leadership', including quality improvement, defined (Juran, op.cit.: 28) as: 'the
organized creation of beneficial change; the attainment of unprecedented levels of performance.' For Juran (Juran, *op.cit.*: 15), 'Reaching agreement on what is meant by *quality* is not simple. ... For managers, no short definition is really precise, but one such definition has received wide acceptance: quality is *fitness for use*’ (emphases in original). A definition of quality frequently adopted in HE (Goodlad, 1995: 8; Green, 1994: 15) is that proposed by Ball (1985: 96): ‘Quality is fitness for purpose.’ However, as Green (1994: 15) points out: ‘The problem with this definition of quality in higher education is that it is difficult to be clear what the purposes of higher education should be.’ Goodlad (1995: 8) identifies an additional weakness of the fitness for purpose definition of quality: ‘it also omits the notion that some universities may be “better” in some way than others in the sense that the purposes they serve may be more comprehensive or more desirable than the purposes of their competitors.’ A similar relativism is favoured by Liston (1999: 4): ‘Quality is related more to the relevance and value of each institution’s mission, purpose, goals and objectives, and the achievement of identified outcomes.’

Crosby (1979: 9) defines quality as ‘conformance to requirements’ and advocates (Crosby, *op.cit.*: 10) ‘the Zero Defects concept ... the thought that everyone should do things right the first time.’ Nightingale and O’Neil (1994: 165), however object:

> Because we believe that the higher education system, including all stakeholders ..., must be dedicated to a conception of quality which is primarily focused on continuous improvement, we reject the idea that quality means ‘getting it right first time’ or ‘zero defects’. Mistakes are OK; they are to be expected. But they must be a source of a new learning experience and of progress.

As a result of eliminating defects, quality for Crosby (1979) becomes ‘free’. Harvey (1997: 133), however, objects that Crosby’s interpretation of the cost of quality does not apply in HE: ‘Higher education is not producing a “consistent product”, be it research output or
graduating students’ (Harvey, op.cit.: 133). Moreover: ‘When shifted from the production of inanimate objects to the realms of education, perfectionist approaches to quality have not only little to say about “standards” but also devalue the transformative process’ (Harvey, op.cit.: 140).

Birnbaum (2000: 215) employs an analogy to encapsulate the dilemma of seeking to transfer quality management systems between operational environments: ‘Thinking that what is good for one kind of organization is good for another is like thinking that what is good for dogs is also good for cats. Universities and businesses are different kinds of organizations.’ This contextual relativism raises the important question: ‘Who is the judge of quality?’ (Deming, 1986: 168). Sallis (1993: 23) points out that fitness for purpose or use in terms of measuring up to specification is a ‘producer definition of quality’, but that quality can also be defined as ‘that which best satisfies and exceeds customer needs and wants. This is sometimes called “quality in perception”’ (Sallis, op.cit.: 24). Such a market driven approach to quality is identified in HE by Williams and Loder (1990: 4), who see a direct link between quality and marketing. Similarly, Green (1994: 7) notes: ‘In competitive markets, quality is seen as a vital tool for those organizations wishing to maintain current market share or secure a competitive advantage.’ For Davies and Scribbins (1985: 5): ‘“quality” is a rather meaningless concept unless it refers to market perceptions’; and Peters (1989: 64) advocates that every firm must ‘[e]nsure that quality is always defined in terms of customer perceptions.’

If educationalists are recognizing the need for more client-centredness (Theodossin and Thomson, 1987: 32), the needs and priorities of customers may nevertheless be extremely difficult to establish (Roberts, 1992: 3). Grönroos (1997: 330) argues that customer-
perceived service quality is two dimensional in that customers are interested in both technical quality, concerning what they receive, and functional quality, concerning how they receive a service. Roberts (1992: 3) further points out that customer needs and aspirations change as they experience a service. Haller (1995: 26), for instance, demonstrates that, 'students' evaluated importance of [quality] criteria, as well as their perceptions, changed during the period of the service process.' Consumer driven approaches to quality are made manifest in the British government's (DfES, 2003: para.4.2) White Paper *The future of higher education*:

To become intelligent customers of an increasingly diverse provision, and to meet their own increasing diverse needs, students need accessible information. We will ensure that the views of students themselves are published in a national survey available for the first time in Autumn 2003, which will explicitly cover teaching quality. We also expect institutions to make progress on their own internal systems for securing student feedback.

However, the realization that such national surveys are complex undertakings requiring careful piloting and taking a number of years to become established means that the planned publication of a national survey in 2003 is cancelled, and that the operation in 2004 will also be only developmental (Baty, 2003).

Barnett (1992: 6) draws attention to a further dimension of quality: 'The debate over quality in higher education should be seen for what it is: a power struggle where the use of terms reflects a jockeying for position in the attempt to impose own definitions of higher education.' Similarly, Alderman (1996: 178) detects a dramatic deepening in the tension at the heart of the relationship between, on the one hand, the measurement and assurance of quality in British HE and, on the other hand, the theory and practice of academic autonomy, stating (Alderman, *op.cit.*: 178): 'it is no exaggeration to assert that [this tension] is at the very epicentre of the debate about the future of higher education in Britain'. As Green
(1994: 3) points out: 'Concern about quality and standards is not new. However, until the mid-1980s, any debate was mainly internal to the higher education system. ... the response of academics when the issue became a matter of public concern was bewilderment and a sense of injustice.' Brennan and Shah (2000: 2) explain:

If 'find good people and let them get on with it' characterizes [the] traditional approaches to quality in higher education, they are frequently complemented today by processes which regulate how people 'get on with it', which require them to continuously demonstrate satisfactory 'performance', which encourage them to improve, which provide new systems of rewards and sanctions, which involve shifts in the locus of authority within higher education institutions and between those institutions and the state and other parts of society. (emphases in original)

Quality systems are themselves articulated through processes such as quality assurance, quality control, quality management, quality audit and quality assessment.

Quality assurance is defined by Ellis (1993: 4) as: 'a process whereby a manufacturer or producer guarantees to a customer or client that the goods or service concerned will meet standards consistently.' Additionally, West-Burnham (1994: 168) stresses the idea of defect prevention: 'The fundamentals of quality assurance are meeting specifications through a system designed to ensure prevention. Assurance rejects the notion that mistakes are inevitable' (emphases in original). Whilst Middlehurst (1992: 28) claims that quality assurance is usually defined as a management responsibility, Goodlad (1995: 9) sees quality assurance as concentrating largely on administrative procedures. Ellis (1993: 5) points out that quality assurance commonly subsumes quality control and quality management.

Quality control, the oldest quality concept (Sallis, 1993: 26), is concerned in HE with 'the stuff of the academic enterprise' (Goodlad, 1995: 9). As Middlehurst (1992: 27) explains: 'In higher education, quality control lies largely in the hands of academic staff; it is their
(professional) responsibility to use methods and to design activities which will lead to the
delivery of quality outcomes. Given that quality control is 'an after-the-event process
concerned with detecting and rejecting defective items' (Sallis, 1993: 26), it is, as Ellis
(1993: 5) argues: 'more difficult to apply to a service since once a service has been
delivered it cannot be retrieved and recycled. But at least a service can be identified as
deficient and steps taken to ensure a better performance next time.'

Quality management is explained by Ellis (1993: 5): 'The total process whereby a particular
organization is managed to achieve and hence to be able to assure quality is quality
management, or, less ambiguously, the management of quality' (emphasis in original).
Barnett (1992: 80), however, argues that, 'an idea of management for quality is both
appropriate and desirable; but ... the management of quality is to be mistrusted. Academic
management is more like that of the leadership and direction exerted by an orchestra's
conductor than by an army's general' (emphases in original). Quality management systems
are for Liston (1999: 159): 'The collective plans, activities and events established to ensure
that a product, process or service will satisfy given needs.'

Quality audit is '[e]valuation to verify the effectiveness of control' (Liston, 1999: 159). As
explained by Cox and Ingleby (1997: 5): 'Quality audit investigates whether an institution
has appropriate quality assurance mechanisms in place and whether they are working
effectively.' Brown and Holmes (2000: 3) situate quality audit firmly within an enhancement
framework: 'Quality audit comprises a range of processes and procedures by which an
institution finds out more about itself and uses this data to make improvements to practice.'

Quality assessment, 'involves the judgement of performance and outcomes against certain

45
criteria or objectives, in order to establish whether the required standard has been achieved, and if failures or shortfalls occur, to ensure that they are corrected’ (Middlehurst, 1992: 28). Quality assessment is carried out internally, as institutions will have, ‘their own systems for quality assessment (such as aspects of staff appraisal or probation arrangements), usually tied into their quality assurance processes as in the case of departmental or programme reviews’ (Middlehurst, op. cit.: 29). In UK HE, quality assessment is also carried out externally by the QAA: ‘The Higher Education Funding Councils have statutory obligations to secure the assessment of the quality of the provision they fund. Information must be provided and quality assurance arrangements undertaken by the Agency in forms acceptable to the councils’ (QAA, 1998b: 2).

In Morley’s (2002: 126) assessment: ‘Quality assurance procedures are influencing organizational cultures, academic identities and pedagogical relations in higher education today.’ Harris (1994: 34) asserts: ‘all universities and colleges are in the business of offering a high quality educational experience to their students: the professionalism of staff would not allow otherwise.’ Nevertheless, ‘the quality machinery, with its many relay points of power’ (Morley, 2002: 127) is very much in evidence. Burleigh (2003: 12) describes British academics as being ‘crushed by assessments of their research and teaching’, which he describes (Burleigh, op. cit.: 12) as ‘these Stalinist measures’. For Barnett (1992: 214): ‘What we are witnessing in the quality debate is none other than a fundamental shift in the relationship of higher education to society.’ It is within that debate that the introduction of subject benchmarking is situated.

A further aspect of the quality debate is summarized by Appleby (1999: 55): ‘Looking back through the history of the quality movement, we can see how the focus has shifted from
quality control and quality assurance to a philosophy of prevention and continuous improvement.' The potential enhancement function of subject benchmarking is considered in a later section.

2.3.ii Quality and standards

Knight (2002b: 107) notes: 'there is a lack of agreement about the meaning of quality; we would expect nothing else of academics.' Similarly, Becher (1997: 158) observes: 'The very concept of standards is riddled with ambiguity.' Brennan (1997: 9) points out: 'As with quality, dispute over terminology concerning standards is also a dispute about values, and the power of one interest group to impose its values on others.'

Barnett (1992: 55) is unequivocal: 'The concept of standards I take to be independent of the concept of quality.' For Bridges (1996: 4): 'Quality and standards are closely related in that standards stem from considerations of quality', whereas Middlehurst (1997b: 47) argues that 'judgements of quality tend to be made in relation to explicit or implicit standards.'

Middlehurst (1999: 189) states that, whereas the terms quality and standards are used interchangeably in many countries, in the UK HE system they have 'different, but overlapping, focuses'. Thus 'quality' refers to 'all those elements that relate to teaching and learning processes, particularly as they impact on students' experiences of higher education and the conditions that support student learning' (Middlehurst, op. cit.: 189), and 'standards' refers to 'expected and actual levels of attainment' (Middlehurst, op. cit.: 189). This separation would appear to be reflected in the QAA's outlook: 'We are concerned principally ... with safeguarding the public interest in sound standards of higher education.
qualifications and encouraging continuous improvement in the management of the quality of higher education' (Williams, 2003: 2).

Nevertheless, as Yorke (2000: 69) points out, both quality and standards 'bear on the concept of "the student experience"'. Holloway and Francis (2002: 239) seek to assess for subject benchmarking 'the extent to which, in the longer term, this particular addition to the quality assessment armoury may improve the quality of the higher education experience for all its stakeholders', thus situating standards-related benchmarking in a quality assessment context. An alternative view of this interrelationship is given by Margham and Jackson (1999: 103), who see quality as referring to 'the level of service provided for students. … Quality can be measured and assured and forms the basis of programme validation, monitoring and review.' Standards, however, 'relate both to the perceived reputation of the subject, in relation to other subjects, and to the reputation of the institution. Together these reputations are a measure of the "worth" of an award' (Margham and Jackson, op. cit.: 103). Whilst for Margham and Jackson (op. cit.: 103), quality and standards 'are often used interchangeably to refer to the value and reputation of an institution's programmes of study', the reality is likely to be that the value and reputation of programmes are in fact generated by the interaction of quality and standards. It is surely in this sense that Seville (1997: 68) observes: 'The "quality" of the degree award is a key aspect of its nature'; or Morley (2003: ix): 'Quality, in the public services, invariably relates to performance, standards and output.'

A logical consequence of this symbiotic relationship between quality and standards is that an improvement of standards can also be understood as an enhancement of quality, even if the nature of improvement in higher education remains an unresolved issue (Houston and
Similarly, as expressed in the QAA’s (2003b: para.27) Handbook for enhancement-led institutional review: Scotland: ‘It is important not to confuse the maintenance of standards of awards with the standards of outcome achieved by the students. Improving the student learning experience will, potentially, improve the standards of outcomes achieved by students and result in an increase in the number of students progressing and achieving awards or achieving higher grades of awards.’

2.3.iii Synthesis

In the benchmarking study, the quality of academic programmes is interpreted holistically in the sense suggested by Seville (1997) and Morley (2003), so that stratagems to improve standards are situated in a quality enhancement context.

2.4 Benchmarking

Although benchmarking is still relatively young (Spendolini, 1992: 3), it has nevertheless given rise to a very considerable body of literature, ‘the sheer volume of which threatens to overwhelm the aspirant benchmarker’ (Lund, 1997:2).

2.4.i Definitions and typologies

Commentators note variously that the idea of benchmarking originated with craftsmen (Appleby, 1999: 54) or, more usually, with surveyors Camp (1989: 12). The notion is now associated with the world of business, where it is ‘a common enough term, used widely’ (Appleby, 1999: 53). Price (1996: 3) records: ‘By common consent [benchmarking] is usually acknowledged as rooted in Japanese manufacturing philosophies of continually seeking inspiration and example from other organisations. Its adoption into western management practice is normally credited to Rank Xerox during the mid 1980s’. More
recently, benchmarking has become one of a number of tools being adopted from industry and business by higher education (Weeks, 2000: 59).

For all that it is widespread in its application, benchmarking remains a problematic concept in so far as 'there is no one notion or definition of benchmarking' (Woolf et al., 1999: 146). Price (1996: 3) suggests: 'There are as many definitions of benchmarking as there are consultants advocating their services.' Indeed, Spendolini (1992: 8) records easily collecting 49 definitions. If, for Schofield (1998a: 4), the word 'benchmarking' is being used in 'widely varying ways', for Fielden (1997: [1]): 'Like many management terms the word "benchmarking" is misunderstood and widely misused.' A survey of UK benchmarking in the early 1990s (Coopers & Lybrand and CBI, 1993: 3), for instance, concluded: '67% of Times 1000 companies now use benchmarking'; but also (Coopers & Lybrand and CBI, op.cit.: 4): 'Many different definitions of benchmarking suggest considerable confusion as to the meaning of the term.'

The 'Xerox definition' (Coopers & Lybrand and CBI, op.cit.: 4) given by Coopers & Lybrand and CBI (op.cit.: 4) is: 'A continuous, systematic process of evaluating companies recognised as industry leaders, to determine business and work processes that represent "Best Practices" and establish rational performance goals.' For Camp (1989: 12): 'Benchmarking is the search for best practices that lead to superior performance.' Spendolini (1992: 15) emphasizes: 'Perhaps the simplest one-phrase response to the definition question would include some reference to "learning from others."' In Alstete's (1996: 20) view: 'fundamentally, benchmarking involves analyzing performance, practices, and processes, within and between organizations and industries, to obtain information for self-improvement' (emphases removed). A similar definition is given by Liston (1999: 98):
Benchmarking is a quality management tool used when comparing one organization with another on some aspect of performance. Aspects of performance include processes, products and services. Searching to find information on these various aspects in which another organization excels, with the objective of finding ways in which to improve current performance is benchmarking.' Appleby (1999: 55) finds 'the key elements of benchmarking to be a continuous, systematic process, involving internal and external measurement of products services and processes, which leads to better practice and improved performance.'

Benchmarking is characterized by McNair and Leibfried (1994: 323) as 'a class on learning how to learn.' However: 'A target benchmarking company expects to give and receive information' (McNair and Leibfried, op.cit.: 44; emphasis in original). As a result, benchmarking is seen as a partnership (Tucker, 1996: 9), sometimes organized as a club activity (Fielden and Carr, 2000: 195). Indeed, Jackson (2001a: 221) points out: ""purists" argue that there is only one practice model (collaborative partnerships'). At the same time, it should not be forgotten that the business origins of benchmarking are also linked to competition: 'Because of the competitive pressures and the mandate to "improve or perish", U.S. businesses were forced to learn quickly how to identify best practices and to implement them in their own companies' (Tucker, 1996: 4). Weeks (2000: 60) notes of a benchmarking project in Australian HE: 'We did find some resistance to the process from some of the possible partners we approached. Some saw us as competitors and did not want to share information with us.' Additionally, benchmarking 'is not as quick and easy as it first seems' (Alstete, 1996: 26), and 'if done correctly is a complex process. It is also an expensive one, in terms of time and money' (Lund, 1997: 4). Fielden and Carr (2000: 196) report of a benchmarking operation involving universities in America, Canada and later
Australia in annual quantitative studies: 'institutions received statistical reports showing how they compared with others. While membership had once been as high as 300, it was steadily falling and there were signs in 1994 that some members felt that the statistical and analytical effort needed was not repaid by the information gained.'

Alstete (1996: 27) records that there are at least four kinds of benchmarking:

- Internal benchmarking
- Competitive benchmarking
- Functional/Industry benchmarking
- Generic benchmarking

In internal benchmarking: 'processes are compared between operating units, divisions, or sister companies' (Alstete, op.cit.: 28). 'The goal of competitive benchmarking is to study the product designs, process capabilities, and/or administrative methods used by an organization's competitors or peers' (Alstete, op.cit.: 29). Spendolini (1992: 20) explains that functional benchmarking 'involves the identification of products, services, and work processes of organizations that may or may not be your organization's direct competitors. The objective ... is to identify best practices in any type of organization that has established a reputation for excellence in the specific area being benchmarked.' For Spendolini (op.cit.: 21), another name for functional benchmarking is generic benchmarking, of which Alstete (1996: 31) states: 'Generic benchmarking, also called "best-in-class", uses the broadest application of data collection from different kinds of organizations. Generic benchmarking compares work processes at one organization to others who have truly innovative and exemplary performance.'

Schofield (1998b: 13) adds to Alstete's (1996) list a fifth type, implicit benchmarking, 'to cater for situations where the initiative for some variant of benchmarking within higher
education results from the market pressures of privately produced data, from central funding, or from co-ordinating agencies within individual systems.' Price (1996: 3), however, warns that the term benchmarking 'is also frequently used by consultants, inside as well as outside organisations, to justify and add "image" to an ordinary process of gathering comparative data.'

Woolf et al. (1999: 146) stress: 'When applied to higher education, benchmarking has been limited to quantifiable comparisons of, for example, financial systems'. As Yorke (2000: 81) observes: 'benchmarking academic aspects of higher education is a much more subtle and complex matter than benchmarking non-academic aspects of performance (premises-related expenditure and so on) or benchmarking industrial processes.' It is, however, possible to envisage qualitative benchmarking, as explained by Price (1996: 6): 'some benchmarking exercises involve no more than the qualitative comparison of a particular function, or service, or technology amongst a group of competitors. Such benchmarking requires as a prerequisite an independent, honest, consultant who must be trusted not to bias the exercise with a personal judgement.' Price (op.cit.: 6) also points out an inherent problem with this approach: 'Unfortunately [such exercises] are also easier to dismiss as subjective if the results do not happen to match the expectations of the managers whose activity is being benchmarked.'

The GSP (BEQC: 1997: para.5.6) Final Report identifies a need for measures enabling programmes and qualifications 'to be mapped and understood, both in relation to each other, and to agreed bench-marks.' There is, however, no explanation of such an approach. Additionally, the status of these 'bench-marks' is clouded by the 'Recommendations Endorsed' (HEQC, op.cit.: para.7.2), which include:
i) promoting and supporting institutional explicitness about standards
x) aligning assessment conventions and benchmarking practice
xiii) ensuring that each institution clarifies its own threshold standards

The Dearing Report (NCIHE, 1997: Recommendation 25) also offers no definition or theoretical underpinning of benchmarking, neither does it specify 'how it would be done' (Jackson, 1998b: 75). Similarly, Armstrong (1999: 24) notes of QAA documentation: 'nowhere is the concept or theory of “benchmarking” defined.' In its brief for benchmarking groups, the QAA (1999a: 4) merely states:

The task of a benchmarking group is to produce broad statements which represent general expectations about standards for the award of honours degrees in a particular subject area. ... [Benchmarking] is about the conceptual framework that gives a discipline its coherence and identity; about the intellectual capability and understanding that should be developed through the study of that discipline to honours degree level; the techniques and skills which are associated with developing understanding in that discipline; and the level of intellectual demand and challenge which is appropriate to honours degree study in that discipline.

At the same time: 'There is no prescribed format for a benchmarking statement' (QAA, *op.cit.*: p.6).

Brown (1999a: 37) refers to a QAA briefing meeting in 1998 at which it was indicated that 'the way the term [benchmarking] is being used by the QAA is more as a point of reference than as used in the industrial/commercial world.' The QAA (2000b: 4) itself describes the first 22 subject benchmark statements as 'reference points - to be used as appropriate, when programmes are designed, approved, reviewed and explained in programme specifications.'

Jackson (1998a: 5), however, writing on behalf of the QAA, states: 'UK higher education is attempting to use the idea of benchmarking as an aid to improvement in a regulatory as well as a developmental sense' (emphasis added). Implicit in this last statement is the view that UK HE is voluntarily and uniformly undertaking benchmarking in the way indicated by
Jackson (op.cit.: 5) when proposing a more appropriate definition of benchmarking as it is being developed in UK HE: ‘a learning process to facilitate the systematic comparison and evaluation of practice, process and performance to aid improvement and regulation’ (emphases removed).

This study investigates whether UK HE is acting in the way identified by Jackson (1998a) or is, perhaps, being made to act in that way, as, indeed, is subsequently suggested by Jackson himself (2000a: 201) – though now no longer writing on behalf of the QAA: ‘The fact that the benchmarking information will be used as a reference point for external peer review (rather than the information being used solely as a prompt for self-reflection) will create a dynamic that logically will drive higher education towards compliance.’ No study of this kind is currently known to exist. A project announced by the QAA (Laugharne, 2001: 9) in 2001 ‘to gain insight into the ways the benchmark statements and other reference points are being used and understood within the higher education sector’ had become by 2003 ‘an evaluation of the use and value of existing benchmark statements’ (Gale, 2003: 9). In 2004, however, the QAA (Gale, 2004: 4) announced the formation of a group merely ‘to evaluate the use of benchmark statements in institutions.’

2.4.ii  Benchmarking and quality enhancement

Appleby (1999: 55) summarizes: ‘Looking back through the history of the quality movement, we can see how the focus has shifted from quality control and quality assurance to a philosophy of prevention and continuous improvement.’ Middlehurst (1997a: 7), for instance, states: ‘In a large and diverse higher education system, identifying and sharing good practice is an important part of quality enhancement.’ This corresponds to the principles underlying the QAA’s ‘new approach’ (QAA, 1998b: 3) and re-affirmed in the
external review process, the first objective of which is: ‘to contribute, in conjunction with
other mechanisms, to the promotion and enhancement of high quality in teaching and
learning’ (QAA, 2002a: para.15). Benchmarking is firmly situated in this when Wright and
Williams (2001: 11), explaining how the QAA system of quality assurance and the standards
infrastructure ‘all fits together’ (Wright and Williams, op.cit.: 11), list the elements of the
infrastructure as being:

- The frameworks for higher education qualifications ...;
- the Code of practice for the assurance of academic quality and standards in
  higher education ...;
- Subject benchmark statements ...;
- programme specifications ...

Jackson (2001a: 219) sees benchmarking as ‘part of the continuum of performance
assessment activities that began in the mid-1980s with Government aspirations for a more
cost-effective, publicly accountable and regulated HE system.’ More precisely, Appleby
(1999: 55) notes: ‘Many would argue that benchmarking is simply a part of the Total
Quality Management (TQM) process of continuous improvement.’ Indeed, Appleby’s
(op.cit.: 57) conclusion is that benchmarking is ‘a tool of TQM, part of the process
methodology of continuous improvement.’ Nevertheless, benchmarking, in particular
subject benchmarking is devoid of certain key attributes of TQM: ‘In TQM staff members
are known as internal customers. ... One of the aims of TQM is so to change the institution
that it operates, without internal conflict and competition, as a genuine team with the single
objective of satisfying customers’ (Sallis, 1993: 32). Even if it is accepted that students are
becoming ‘increasingly discriminating investors in higher education’ (NCIHE, 1997:
para.1.21), the nature of the institutional change involved sets TQM apart from subject
benchmarking. Whereas externally imposed quality reviews can be seen by academics as
invasive and professionally diminishing, ‘TQM requires a change of culture. ... The motivation to do a good job comes from a leadership style and an atmosphere which heightens self-esteem and empowers the individual’ (Sallis, 1993: 37). Fielden (1997: 1) suggests that confusion over benchmarking may be due in large measure to the fact that it appeared at the same time as techniques such as TQM, going on to ask (op. cit.: 1): ‘is TQM part of benchmarking?’ Although Fielden (op. cit.) does not answer the question, there is no reason to suppose an automatic dependency in either direction between benchmarking and TQM.

At the same time, however, benchmarking is seen as one of the ‘tools and techniques for quality improvement’ (Sallis, 1993: 105). It is, after all, one of the features of benchmarking that comparison with best practice is said to reveal ‘performance gaps’ (Alstete, 1996: 23) which ‘must then be used for adapting improvement efforts, and setting operation goals for change’ (Alstete, op. cit.: 23). Yet Appleby (1999: 58) also identifies ‘a gap between the practice and theory of benchmarking’, going on to explain how this can come about: ‘the focus is often on measurement and comparison, rather than on identifying and then exploiting best practices’ (Appleby, op. cit.: 58). Alternatively, as Hammer and Champy (1993: 132) warn: ‘By aspiring only to be as good as the best in the industry, the team sets a cap on its own ambitions. Used this way, benchmarking is just a tool for catching up, not for jumping way ahead.’ The reminder from McNair and Leibsried (1994: 36) is stark: ‘Parity, though, is not the key to long-run competitive success; superiority is.’

Referring to the publication of the first batch of 22 subject benchmarks, Baty (2000: 60) reports that almost all of them include ‘a checklist detailing the skills and intellectual abilities expected.’ However: ‘the benchmarking groups in history and English refused to
provide a checklist. Anthony Fletcher, chair of the history benchmarkers ... said his group was concerned that a 'tick box' was too crude' Baty (op. cit.: 60). It is Brown's (reported in Baty, op. cit.: 60) view that 'the benchmarks raise questions about diversity and autonomy.'

For Brown (2001: 12), quality improvement is dependent on institutional risk taking; but: 'It is very hard to reconcile this vision of an enterprising, risk taking, outward looking sector with the day to day realities of accountability as we currently have it in higher education!' (Brown, op. cit.: 12). Hannan and Silver (2000: 115) conclude that innovation in HE teaching and learning is most likely to be obstructed by, among other things, 'quality assessment procedures or other procedures that inhibit risk-taking.' Whereas the QAA (1998b: 4) states: 'Quality assurance must be rigorous and effective, but it must not place a disproportionate burden on institutions and their staff, or distort the teaching process', Holloway and Francis (2002: 246) observe: 'The potential of the new system as a whole to stifle innovation under a morass of codes of practice, benchmark statements and frameworks is still high.'

In response to Recommendation 25 of the Dearing Report (NCIHE, 1997), the QAA (1999a: 6) stipulates that two main standards of attainment are to be identified in subject benchmarking statements:

The first is the attainment expected of the 'typical' student whose results would fall into the main modal cluster. Benchmarking groups should regard this as the main standard to be identified.

The second is the threshold, the minimum requirement that must be met by anyone graduating with honours in the discipline.

Goodlad (1999: 79) identifies in this second variety the possible danger of "dumbing down" by defining subjects in terms of minimal achievements with the associated danger of teaching to the syllabus.'
It can thus be argued that subject benchmarking may be hindered as a possible vehicle for quality enhancement both by the manner in which it is conceived and by the quality regime within which it has to operate. Moreover, the fictitious report on quality and standards, provided by the QAA (1999d) to indicate ‘what might be expected as a result of the Agency’s new method for quality and standards’ (QAA, op.cit.: 16), albeit before the promulgation of the external review process, gives no indication whatever of how subject benchmarking and threshold standards contribute to quality enhancement. Any misgivings arising from this state of affairs are unlikely to be assuaged when the acting chief executive of the QAA (Williams, 2002: 14) refers to ‘”enhancement” - whatever that word means’.

Brown (2000: 330) observes that quality evaluation regimes generally ‘seek to combine two main elements: accountability ... and enhancement or improvement (systematic learning from quality processes and from best practice). ... at least in the longer run, it is enhancement that is likely to have the greatest payoff in terms of lasting quality improvement.’ Brown (op.cit.: 330) further notes the QAA’s reference to enhancement in its presentation of ‘the new approach’, concluding (Brown, op.cit.: 330): ‘It seems clear ... that the overall thrust of the new framework is towards accountability to third party funders and users.’ Nevertheless (Brown, op.cit.: 330):

At the same time many of the processes involved are potentially valuable forms of enhancement, for example, the benchmarking of academic standards, programme specifications etc. Whether they turn out that way will depend very much on how they are implemented in institutions and how they are regulated by the Agency, including funding and other consequences of adverse judgements.

As HE institutions prepare to undergo their first audits under the QAA’s external review process, Brown’s (2000) observations have an even greater relevance. This study investigates perceptions of subject benchmarking and threshold standards as motors for
quality enhancement and also looks at institutional attitudes to their introduction.

2.4.iii Benchmarking and conformity

Harvey and Knight (1996: 102) argue that 'accountability-led, funding-linked, quality monitoring will, at best, only have a short-term impact on quality and is much more likely to lead to a compliance culture in the long-term.' According to Barnett (2003: 93): 'quality, as a state-backed project, is always likely to produce a compliance culture' (emphasis in original). On benchmarking in UK HE, Jackson (1998a: 5) notes laconically: 'Interest in benchmarking has increased significantly as it has become a key element of the policy framework proposed by the [NCIHE]'. Goodlad (1999: 71) is more direct, referring to the QAA ‘pressuring universities into producing “benchmarks” and “templates”’, and to ‘the threat of withholding funding ... used as an instrument to force conformity’ (Goodlad, op.cit.: 72). Similarly, Yorke (2000: 68) identifies the primary characteristic of the QAA’s notion of benchmarking as ‘an orientation towards control’.

Commenting on the QAA’s ‘new approach’, Baty (2001c: 3) remarks that institutional audits will make universities conform to programme specifications and subject benchmarks. The force of this is illustrated by Hargreaves and Christou (2002: 187): ‘In common with many institutions, Sheffield Hallam University has aligned its own internal quality assurance systems to the QAA requirements.’

Subject benchmarking raises particular issues at the subject level and for individual academics. For Elton (quoted in Utley, 1998: 8): 'Benchmarking is an unbelievable simplification of what we do and lacks any intellectual basis'. Utley (op.cit.: 8) reports that many delegates at a conference on benchmarking and threshold standards at the end of
1998, 'feared that innovation would be stifled by the prescriptive approach, leading to a national curriculum for higher education.'

The GSP (HEQC, 1997: para.5.24) concludes: 'Although threshold standards were widely regarded as desirable, the [HE Quality] Council's detailed investigations through the GSP did not suggest that they were yet generally feasible.' Armstrong (1999: 21), however, notes: 'Dearing seems not to have accepted the serious cautions expressed by the GSP concerning threshold standards'. Additionally, the GSP (HEQC, 1997: para.7.3) specifically rejects as an option: 'the development of national curricula and national assessment arrangements' (cf. Silver and Williams, 1996: 42). Dearing (NCIHE, 1997: para.10.3) states: 'Uniformity of programmes and national curricula, one possible approach to the development of national standards, would deny higher education the vitality, excitement and challenge that comes from institutions consciously pursuing distinctive purposes, with academics having scope to pursue their own scholarship and enthusiasm in their teaching.'

For Baty (1997: 4), the immediate post-Dearing position is clear: 'Dearing has stopped short of recommending a national curriculum for higher education - but only just.' Reacting to the Dearing quality assurance model, QAA Chairman Kenyon (1997: 19) stresses the need for the QAA 'to avoid the merest hint of a national curriculum'. Nevertheless, even though the subsequent QAA consultation on developing the quality assurance and standards framework (QAA, 1998a: Part V, para.3) restates the need to avoid 'the risk of curricular prescription', the QAA (1998b: 11) acknowledges: 'Some concerns about benchmarking have emerged from the consultation. Fears were expressed that they might become tantamount to a national curriculum, or alternatively that they would be so general as to be meaningless.' Goodlad (1999: 78), for instance, identifies in the consultation document
(QAA, 1998a) 'the danger of sliding into a national curriculum for higher education.' Yet, in its account of the responses to its consultation (QAA, 1998a), the QAA (1998b: 19) stresses: 'It is important to emphasise that the Agency has never intended that the process of benchmarking should be concerned with the question of curricular content. On the contrary, its focus has consistently been on defining the general intellectual outcomes appropriate to a particular subject.' Picking up on this, Armstrong (1999: 27) asks: 'if this is the case, why do we need the huge exercise of 42 subject benchmarks?'

In the light of such misgivings, it is perhaps understandable that the QAA's (2002b) draft Handbook for institutional audit should be greeted by Baty (2002: 4) with the words: "'Guidance' ousts "compliance" at QAA.' After all, the draft handbook (QAA, 2002b: para.55) states: 'When considering the institution's management of quality and standards, the audit team draws upon a range of external reference points, including the FHEQ, Subject benchmark statements and the Code of practice ... . In so doing, it is not seeking evidence of compliance ...'. Alderman (2002: 16) summarizes 'the good news' of the document (QAA, 2002b): 'The draft makes it clear that the QAA's mammoth code of practice, the numerous subject-benchmark statements and even the Framework for Higher Education Qualifications no longer have the status of prescriptive regulations. Henceforth they will merely be "reference points" (Alderman, 2002: 16). For Baty (2002: 4): 'The code of practice for quality assurance, incorporating hundreds of new rules, and the national "subject benchmark statements" setting out minimum standards for courses will become non-compulsory elements of the new QAA audit regime.'

Initial reactions of this kind require careful consideration, however. As has already been discussed, the QAA (2000b: 4) stresses as early as April 2000 that benchmarking statements
are 'reference points'. Similarly, Wright and Williams (2001: 11), on behalf of the QAA, explain: 'A set of subject benchmarks could be, and has been taken by some to be, a list of outcomes to be achieved. This is not our intention; we interpret subject benchmarks as reference points.' The QAA's (2001b: 4) own brief guide to quality assurance in UK HE additionally indicates that not only subject benchmark statements, but also the qualifications frameworks, the Code of Practice, and programme specifications are all to be understood as reference points.

Alderman's (2002: 16) interpretation of the QAA's reference points is thus less than apposite. Similarly, Baty (2002: 4) overlooks that the 'Operational description' of the external review process (QAA, 2002a: para.55), published one month before the draft Handbook, expressly rules out the search for 'unthinking compliance' with the QAA's reference points. Baty (2002: 4) also overlooks that the draft Handbook contains no explicit reference to threshold standards (Baty's 'minimum standards': op.cit.: 4), simply specifying (QAA, 2002b: para.16) that QAA audit reports will set out the audit team's judgements on: 'the confidence that can reasonably be placed in the soundness of the institution's present and likely future management of the quality of its programmes and the academic standards of its awards'. Even at the very beginning of the benchmarking process, the QAA (1998a: Part IV, para.4) describes the statements of standards to be produced by the benchmarking groups as 'reference materials'. Nevertheless, the comments of Baty (2002) and Alderman (2002) do serve to illustrate the strength of the scepticism felt by the HE sector. Hargreaves and Christou (2002: 190), for instance, state that, when originally introduced, subject benchmarks 'were clearly part of a compliance quality assurance model.'

Interesting evidence in the compliance debate is also provided by the consultation document
The proposed [quality assurance] method would operate within the framework being established by the QAA of standards benchmarks in each subject area, the qualifications framework, the various sections of the QAA code of practice, and the programme specifications. Between them, those provide common points of reference for institutions as they assure the quality and standards of their programmes and awards. They act also as common points of reference for reviews - not as prescriptive blueprints which must be adopted, but as expectations that will be met in different ways to reflect the differing needs, traditions, cultures and decision-making processes of individual institutions. The elements of that framework will be reviewed from time to time to maintain their currency.

In spite of these reassurances indicated above, the QAA (2001c: para. 15) records in its report on responses to HEFCE 01/45: ‘There was strong encouragement to reinforce the message that the elements of the academic infrastructure - Code of Practice, benchmark statements, qualifications frameworks etc - should be seen as advice and guidance, rather than prescription and expectation.’ Of the two major HE trade unions, the QAA report (op. cit.: para. 58) states: ‘AUT and NATFHE generally supported the stated principles and objectives, the audit model, and the “guidance not prescription” statements about benchmarks, codes of practice etc (but questioned whether this was how it worked in practice and suggested that the operation of the benchmark statements should be reviewed).’

A closer look at the unions’ detailed responses to HEFCE 01/45, however, reveals a somewhat different picture. The AUT (2001: 3) comments:

We are ... disappointed that the mighty structure of qualification frameworks, programme specifications, benchmarks and codes of practice remains in place ...

... We welcome the commitment to reviewing the quality and standards infrastructure on a periodic basis. Our members have raised particular concerns about the impact of benchmark statements on curriculum planning, teaching, learning and assessment. They fear that benchmarks - in conjunction with the development of an HE national framework for qualifications - will limit constructive change in disciplinary developments and restrict the development of teaching by insisting upon the
mechanistic fulfilment of demands set out as skills and attributes. In order to assuage fears that the QAA are ushering in a "national curriculum" by the back door, an initial review of benchmarks should take place in the near future.

Similarly NATFHE (2001: [2]):

Whilst much of what is in the current QAA benchmark statements is uncontroversial, and sufficiently generally phrased not to inhibit innovation and difference, it is vital that this remains the case and that further development of an HE national framework for qualifications and benchmark standards does not result in moves towards a national curriculum. We welcome the commitment to review of the elements of the framework and believe that an initial review of the operation of the benchmark statements should take place before long.

Doubts about the threat of prescription and compliance are thus deep-seated and resistant to what may be interpreted as mere lexical ploys by the QAA to overcome them.

Billingham (1999: 196) observes: 'The QAA intends that benchmarking should focus on "the intellectual attributes associated with successful study of a discipline to degree level" (QAA, 1998[b]). If this can be successfully unpacked, then benchmarking need not interfere unduly with academic freedom to determine the subject knowledge content of the curriculum.' The dangers arising if this is not achieved are identified by Brown (1999a: 47):

We must ... ensure that we avoid the worst excesses of the National Curriculum as applied to schools in the UK, in which a programme to drive up pupil performance by the identification of levels of achievement for each Key Stage has resulted in what many describe as an ossification of the content of the curriculum, loss of morale among teachers and a diminution of their ability to be innovative and creative.

As Brennan (1995: 37) notes of secondary education: 'The introduction of the National Curriculum and league tables has resulted in an obsession with recording, and subsequently, publicizing the product. ... The loss of teacher autonomy is there for all to see as teachers hasten through units of work so that, primarily, a checklist can be completed.' In the particular case of history, Phillips (1991: 22) reports: 'A questionnaire sent to heads of department after the publication of the final report [for National Curriculum history]
confirmed that their major concern was that the National Curriculum would undermine autonomy.' Phillips’ (op. cit.: 22) account shows a further parallel with the potential predicament of HE:

After the publication of the final report, ex-members of the working party and officers of the Historical Association made strenuous efforts to play down its prescriptive nature, emphasizing that the final report was only a ‘guide-line’ for the encouragement of existing ‘good practice’ and that it provided ample opportunity for teacher creativity, flexibility and autonomy.

As far as content is concerned, the parallel may not be exact, but the similarity is certainly striking when Phillips (op. cit.: 23) states: ‘Whatever professional officers of the National Curriculum Council ... may claim, the content requirements of the National Curriculum are formidable.’ It may at all events be unsettling for HE academics to encounter Jackson’s (2001b: 3) explanation that subject benchmark information ‘is an attempt to provide an alternative policy solution to a national curriculum or national assessment.’

Apprehension about a possible national curriculum for HE is almost palpable in some quarters, as is the impact of the post-Dearing quality agenda. Answering their own question about how to initiate new approaches to group-based learning in university history, Nicholson and Ellis (2000: 218) begin:

- Be clear about objectives and outcomes: ...
- Start small: ...
- Work with the grain: point to the increasing emphasis on particular teaching, learning and assessment strategies promoted by the Dearing Report, Quality Assurance Agency, Higher Education Funding Council and others.

As far as subject benchmarking is concerned, an additional problem is that ‘little practical guide is offered as to how the information should be used’ (Jackson, 2001b: 1). Even if compliance with subject benchmarks is not an avowed part of the external review process, the Handbook for institutional audit: England (QAA, 2002c: para.55) nevertheless states
that audit teams be looking for: 'evidence that the institution has considered the purpose of
the reference points, has reflected on its own practices in the relevant areas, and has taken,
or is taking, any necessary steps to ensure that appropriate changes are being introduced'
The obligation for academics to engage with subject benchmarks is thus clear, as is the
obligation to demonstrate such engagement to an external audit team. Far from clear,
however, remain the criteria for assessing 'necessary steps' and 'appropriate changes'.

This study investigates how academics perceive this enforced engagement with
benchmarking, and the effect it has on their professional lives. In doing so, it is
acknowledged that academics from different disciplines may a priori be disposed differently
to the problems. The GSP (HEQC, 1996: para.10.12.4), for instance, notes: 'Chemists ...
pointed to the existence of a core curriculum in chemistry, supported by the Royal Society
of Chemistry.' Dearing (NCIHE, 1997: para.10.4), too, observes: 'For some programmes ...
the setting of some form of core national curriculum may be appropriate, for example in
medicine where doctors need to be trained to consistent standards, and in teacher training
where the existence of the national curriculum in schools demands a degree of consistency.'

2.4.iv The appropriateness of benchmarking

In view of the potential opportunities and weaknesses of benchmarking, some consideration
should be given to the appropriateness of the approach when applied to quality and
standards in HE.

An initial difficulty may well be an affective reaction to a technique originating in an
environment perceived as alien: 'We can borrow from the rhetoric of industrial quality
models, auditing our systems and processes, examining our customer-care features, writing
mission statements, practising kaizen and adding value, but we cannot adopt them wholesale without the loss of much we hold dear’ (Brown (S.), 1997: 185).

A doubt must also be raised by the nature of benchmarking as a predominantly quantitative approach. Gaither et al. (1994: 6) point out that the development of high-performance, user-friendly IT systems was a key factor in the emergence of PIs: ‘What was available was collected and measured. What was measured - or measurable - was given value, and what was given value was reviewed for accountability and funding (Gaither et al., op. cit.: 6). As Peters (1989: 490) observes: ‘“What gets measured gets done” has never been so powerful a truth.’ This can generate a checklist mentality, with the dual danger of, on the one hand, distracting from full engagement in the educational process by provoking mere ‘procedural compliance’ (Bryman et al., 1994), and, on the other hand, engendering ‘Beckmesser’s Fallacy’ (Hart, 1997: 308) through the potential for objections to items not included in pre-determined criteria: ‘We cannot reduce ... judgement to a list’ (Wolf, quoted in Leon, 2000: 2). The QAA’s (2002a: para.24) statement that, in the external review process: ‘Audit teams will not expect to see mechanistic or “checklist” approaches’, reveals how great the temptation is. Scott (2003a: 14) warns that the new performance culture in universities ‘may actually undermine performance, indirectly by diverting energy and ingenuity into game-playing and directly by discouraging risk-taking.’

Benchmarking as advocated by Dearing and operationalized by the QAA may be perceived as a rather static phenomenon, the published statements having an undetermined validity and no known mechanism for revision, even though the QAA is ‘committed to setting up arrangements for reviewing and refreshing the information produced by the early phases of the benchmarking project, but not before 2003’ (Laugharne, 2002: 136). Any static
interpretation accords ill with the normal view of benchmarking as a dynamic process: ‘constant, consistent and comprehensive’ (Camp, 1989: xii). More importantly, it is difficult to reconcile both with the non-static view of standards espoused by the QAA (Randall, 2001: 1) and with the dynamic nature of academic disciplines (Holloway and Francis, 2002: 241). Subject benchmarking may thus be seen as offering a bureaucratic approach to an intellectual challenge: ‘Some academics find [the quality assurance framework] burdensome and over prescriptive, others regard it as “harmless, but not terribly helpful”’ (BRTF, 2002: para. 7.7).

For Morley (2003: 45): ‘A criticism of the benchmarking industry is that it trades in norms, threshold standards and peddles a reductive functionalism.’ It has to be taken on trust, however, that the norms applied in each of the 42 academic disciplines covered are of equivalent standard, which, in turn, raises the spectre of having to benchmark the benchmarks: ‘How far is it necessary to establish some level of comparability between the benchmark statements for all subjects in higher education?’ (Holloway and Francis, 2002: 241).

It has been seen that benchmarking is normally associated with some exemplar of best practice, though Holloway and Francis (op. cit.: 242) note: ‘There is ... limited empirical evidence of wide-ranging uptake of “best practice” benchmarking in higher education.’ The QAA benchmarking model, however, does not put forward a real-world exemplar at all, but instead relies on a theoretical construct. As Morley (2003: 45) notes: ‘benchmarking suggests a type of harmony that does not easily exist in the academy.’ It may be felt that this distancing from reality is taken further with the need, ultimately, for institutions to be compared: ‘there might be felt to be something curious in supposedly autonomous
institutions, each with its own "mission", being subject to comparative evaluations" (Barnett, 2003: 78).

Potentially a much more significant problem with the benchmarking approach is that of control. Fletcher (2001: 15) writes of David Blunkett's intervention in the QAA's policy: 'The present debacle is formally about the burden of review. It is really about whether university teaching and standards should be inspected or purposefully reviewed in a collegial manner conducive to improving standards.' Holloway and Francis (2002: 241), referring to the benchmarking group for General Business and Management, record: 'The ground rules set ... by the QAA were not always clear ... reflecting a mix of political and technical aspects of the overall subject benchmark process.' As for Fletcher's (2001: 15) experience in the History benchmarking group: 'it became evident that, in pursuing its plans for benchmarking, programme specifications and codes of practice, the last thing the QAA wanted was partnership with academics.' Indeed, Barnett (2003: 188) comments: 'the [History benchmarking] group declined fully to play the hand dealt for it by the ... Quality Assurance Agency ... and ran into much flak from the QAA as a result.'

According to Morley (2003: 45): 'The argument used to justify benchmarks is that they provide a structure and set of organizing principles in the face of possible chaotic innovations. In these cases, the audience is the course development and validation team.' This would imply an undermining of the original significance of QAA subject benchmarking, perhaps resulting from the abandonment of universal review, as is suggested by Hargreaves and Christou (2002: 190). Alongside the other research questions already identified, this study seeks to investigate whether the shift from subject review to institutional audit is perceived as altering the status and significance of subject benchmarking.
2.4. Synthesis

The literature reveals a significant discrepancy between the normal conception of benchmarking and that emanating from the Dearing Report: on the one hand, a dynamic process of enhancement and continuous development; on the other hand, what may be little more than a static listing of comparators potentially inducing a stifling conformity. This study seeks to investigate these contrasts and their perceived consequences.

Given that subject benchmarking has been introduced, it is important to understand how it is perceived and interpreted. Is it seen as a means of achieving quality improvements, in the normal sense identified in the literature; and, if not, what is its function, and why was it introduced? Spender (2000: 53) notes: 'In acceding to government pressure for quality assurance and accountability, universities may have improved standards in specific areas, but it has been done at the cost of a substantive and damaging shift in concepts of professionalism and in the culture of HE.' This study investigates whether subject benchmarking is perceived as a new manifestation of this government pressure, and whether it is felt to be damaging the self-image of academics, perhaps by undermining their professionalism and imposing a revised notion of the university based on conformity to regulatory norms.
3.1 Research design

Of central importance to the study is the search for insights into how members of the academic community, particularly academic staff, perceive and interpret subject benchmarking and its consequences. Fundamentally, then, the investigation is qualitative in nature rather than quantitative: 'One facet of the distinction between quantitative and qualitative research is that the former is orientated to the specific concerns of the investigator and the latter to subjects' perspectives' (Bryman, 1988: 142).

In Hammersley's (1995: 55) view: 'the differences in philosophical presupposition that are often thought to exist between quantitative and qualitative research are less clear and sharp than is commonly thought.' Morrison (2002: 14) indicates that 'matters of “quantity” and “quality”' are frequently used to typify the distinction between positivism and interpretivism. As expressed by Arksey and Knight (1999: 10), the central idea of positivism
is that 'there are objective facts “out there” to be discovered by rigorous enquiry, leading to laws or generalizations that describe the world and, ideally, allow good predictions to be made.' By contrast, reality in the interpretivist view is ‘a construct in which people understand reality in different ways. (It may be that some human groups perceive reality similarly, but this does not diminish the potential for reality to be construed differently.)' (Morrison, 2002: 18). Ontologically, in its understanding of ‘the nature of reality’ (Denzin and Lincoln, 2000: 19), the positivist view has heavily influenced quantitative research (Morrison, 2002: 16), in which ‘the emphasis is very much upon the individual as the object of research; the aggregation of individualized data provides overall measures’ (Morrison, op. cit.: 17). Qualitative research employs strategies ‘that take the subject’s perspective’ (Morrison, op. cit.: 19), in line with an interpretivist ontology (Morrison, op. cit.: 19). Epistemologically, ‘which has to do with the question of what is to pass as warrantable, and hence acceptable, knowledge’ (Bryman, 1988: 104), qualitative research is shaped by interpretivism, which emphasizes ‘words rather than numbers’ (Morrison, 2002: 21). As Miles and Huberman (1994: 7) indicate, words ‘can be organized to permit the researcher to contrast, compare, analyze, and bestow patterns on them.’

Notwithstanding these differences, there are ‘areas of overlap’ (Bryman, 1988: 172) between quantitative and qualitative research, and they can be combined (Miles and Huberman, 1994: 40). Seale and Kelly (1998: 156), for instance, note: ‘Counting in qualitative research can help in reassuring the reader that the researcher has not simply trawled through a mass of data and selected anecdotes to report that support his or her particular bias.’ Similarly, in qualitative research, ‘the use of multiple methods, or triangulation, reflects an attempt to secure an in-depth understanding of the phenomenon in
question' (Denzin and Lincoln, 2000: 5). Cohen and Manion (1994: 233) observe: ‘The use of multiple methods ... contrasts with the ubiquitous but generally more vulnerable single-method approach that characterizes so much research in the social sciences.’ As Arksey and Knight (1999: 22) put it: ‘in recent years this approach [triangulation] has been seen as a way to obtain greater completeness.’

3.2 Validity and reliability

The concepts of validity and reliability have their origins in the natural sciences (Kirk and Miller, 1986: 13), and are said by Silverman (1993: 145) to be ‘central ... in any discussion of rigour in scientific research’. Nevertheless, even outside quantitative research, ‘provided the researcher is committed to providing a faithful description of others’ understandings and perceptions, then ideas such as validity and reliability can provide a very useful discipline’ (Easterby-Smith et al., 1991: 40).

Whilst Maxwell (1992) discusses five different types of validity, and Seale (1999: 43) lists eleven varieties of validity, Bush (2002: 66) identifies the main distinction as being between internal validity, which ‘relates to the extent that research findings accurately represent the phenomenon under investigation’ and external validity, which ‘relates to the extent that findings may be generalised to the wider population which the sample represents, or to other similar settings’ (Bush, op. cit.: 67). For Hammersley (1987: 75), most definitions of validity claim that it ‘represents the extent to which an instrument measures the property it is intended to measure.’ Gilbert (1993: 27), for instance, states that measurements are valid if they are ‘accurately measuring the concept’.

Lincoln and Guba (2003: 274) note: ‘Nowhere can the conversation about paradigm
differences be more fertile than in the extended controversy about validity.’ Indeed, if it is reasonable to suppose that ‘researchers want their indicators to be as good as possible’ (Gilbert, 1993: 27), for the qualitative researcher, the epistemological problems posed by this endeavour are considerable. A striking expression of the dilemma is given by Kirk and Miller (1986: 21):

To discuss the validity of a thermometer reading, a physical theory is necessary. The theory must posit not only that mercury expands linearly with temperature, but that water in fact boils at 100°. With such a theory, a thermometer that reads 82° when the water breaks into a boil can be reckoned inaccurate. Yet if the theory asserts that water boils at different temperatures under different ambient pressures, the same measurement may be valid under different circumstances - say, at one-half an atmosphere. In the case of qualitative observations, the issue of validity is not a matter of methodological hair-splitting about the fifth decimal point, but a question of whether the researcher sees what he or she thinks he or she sees.

The fundamental problem is that raised by Russell (1959: 3) when he discusses how a table will appear different in size, shape, colour and texture, depending on who observes it, in what light, at what angle, etc:

it becomes evident that the real table, if there is one, is not the same as what we immediately experience by sight or touch or hearing. The real table, if there is one, is not immediately known to us at all, but must be an inference from what is immediately known. Hence, two very difficult questions at once arise; namely, (1) Is there a real table at all? (2) If so, what sort of object can it be? (emphasis in original)

This contrasts with the view of objectivists, for whom the environment is ‘a comprehensive, objective, intractable set of facts’ (Tierney, 1987: 64). But just as Russell (1959) enumerates individual perceptions of a table without being able to assert with confidence that there is a real table at all, so the comforting assumption: ‘There is a world of empirical reality out there’ (Kirk and Miller, 1986: 11) is undermined by the observation: ‘The way we perceive and understand that world is largely up to us’ (Kirk and Miller, op.cit.: 11). On this view, then, reality, whatever that may be taken to mean, is necessarily mediated by the
individual and as a function of each individual's subjectivity: 'Reality is not something objective or external to the participants' (Tierney, 1987: 64). Nor should the concern be forgotten that 'the natural subjectivity of the researcher will shape the research' (Marshall and Rossman, 1999: 194).

The sources of difficulty, then, are complex and numerous. Cohen and Manion (1994: 281) suggest the use of convergent validity as a possible approach: 'One way of validating interview measures is to compare the interview measure with another measure that has already been shown to be valid.' For Kvale (1996: 236): 'the emphasis on validation is moved from inspections at the end of the production line to quality control throughout the stages of knowledge production.' In particular: 'Validation comes to depend on the quality of craftsmanship during investigation, continually checking, questioning, and theoretically interpreting the findings' (Kvale, op. cit.: 241). Arksey and Knight (1999: 52) suggest eight stratagems for enhancing the validity of interviews, including: 'Interviewing techniques that build rapport, trust and openness and which give informants scope to express the way they see things'; 'A set of questions that fully covers the issues raised by the research question – key aspects are not ignored'; 'A sample that fits the purpose of the research.' These stratagems were adopted in the benchmarking study.

Bush (2002: 66) identifies bias as the 'main potential source of invalidity in interviews'. Even though various ways of reducing bias are proposed in the literature (Cohen and Manion, 1994: 282), Bush (2002: 66) contends that 'bias is likely to be endemic, particularly in semi-structured and unstructured interviews, and is difficult to eliminate.' As Arksey and Knight (1999: 51) say of the threats to validity: 'Many of them we just have to live with, on the principle that flawed information is better than none, and there is
sometimes little that can be done about some threats.

Kirk and Miller (1986) identify three kinds of reliability. For Hammersley (1987: 76), most definitions of reliability refer to 'the scores produced by repeated efforts to measure the same property by means of the same instrument.' More generally, Gilbert (1993: 27) states that measurements are reliable if they are 'consistent from one measurement to the next'. Kirk and Miller (1986: 19) again use the example of the thermometer to illustrate the concept of reliability:

A thermometer that shows the same reading of 82 degrees each time it is plunged into boiling water gives a reliable measurement. A second thermometer might give reading over a series of measurements that vary from around 100 degrees. The second thermometer would be unreliable but relatively valid, whereas the first would be invalid but perfectly reliable.

Such a view, however, relies on the assumption 'that there is a stable reality "out there", which is to be precisely measured and described' (Arksey and Knight, 1999: 52, emphasis added). But what if a more appropriate view is that indicated by Heraclitus (Kirk et al., 1995: 195): 'Upon those that step into the same rivers different and different waters flow ... They scatter and ... gather ... come together and flow away ... approach and depart'? In that case, as explained by Marshall and Rossman (1999: 194):

Positivist notions of reliability assume an unchanging universe where inquiry could, quite logically, be replicated. This assumption of an unchanging social world is in direct contrast to the qualitative/interpretive assumption that the social world is always being constructed and that the concept of replication is itself problematic.

This may be, as Silverman (1993: 146) objects, 'not a comfortable position for social scientists', but as Kirk and Miller (1986: 42) emphasize: 'In the study of sociocultural phenomena, it is often dangerous to assume that configurations of data would be isomorphic across substantial intervals of time. To make such an assumption is to deny
history.' Schuman (1982: 25), for instance, takes a question already used in a study to ask whether Russian reporters should be allowed to report freely from the USA, and American reporters from the Soviet Union, asks it again 30 years later, and finds that 'answers to the question not only varied by context, but this variation in turn interacted with time.' Silverman (1993: 10) refers to the difficulty that respondents may lie to interviewers, whereas Dean and Whyte (1958: 38) stress that 'informants can and do hold conflicting sentiments at one time and they hold varying sentiments according to the situations in which they find themselves.' The benchmarking study had to rely on the information as provided, though all respondents were offered the opportunity to reconsider and revise their statements by checking an interview transcript (cf. 3.5.i and Appendix 1). Of the 34 respondents, 18 asked to see a transcript: no changes of substance were made, though some reassurances were sought regarding anonymity (but cf. 6.3).

Both validity and reliability are contested concepts within qualitative research. In Leininger's (1994: 96) view: 'qualitative researchers should not rely on the use of quantitative criteria such as validity and reliability to explain or justify their findings. Such dependence reflects a lack of knowledge of the different purposes, goals, and philosophical assumptions of the two paradigms.' Creswell (2003: 195), whilst according only a minor role to reliability and generalizability in qualitative inquiry, records that validity 'is seen as a strength of qualitative research, but it is used to suggest determining whether findings are accurate from the standpoint of the researcher, the participant, or the readers of an account.' Bassey (1999: 75) argues that Lincoln and Guba's (1985) concept of trustworthiness as an alternative to reliability and validity 'successfully illuminates the ethic of respect for truth in case study research.' Similarly, Denzin and Lincoln (2000: 21) note of
constructivist interpretative paradigms: ‘Terms such as credibility, transferability, dependability and confirmability replace the usual positivist criteria of internal and external validity, reliability and objectivity.’ Marshall and Rossman (1999: 195) assert: ‘Qualitative research does not claim to be replicable.’

A further consideration is that ‘what interviewees say they believe, prefer, or do, may not correspond with “actuality”’ (Powney and Watts, 1987: 190). Powney and Watts (op. cit.: 193) may be right to conclude: ‘In the end, the best an interviewer can hope for is insight into the respondent’s favourite self-image.’ Consequently, if reliability is: ‘the extent to which a test or procedure produces similar results under constant conditions on all occasions’ (Bell, 1993: 64), then semi-structured interviews can be said to be reliable only in so far as the unavoidable uncertainties will again obtain, albeit in ways not accessible to precise comparison. However, as Seale and Silverman (1997: 379) observe: ‘Authenticity rather than reliability is often the issue in qualitative research. The aim is usually to gather an “authentic” understanding of people’s experiences’. The benchmarking project, too, seeks via interviews authentic reactions to the topics being studied, the reliability of the project being understood as its dependability in doing so.

The dilemma of validity and reliability in qualitative research is summed up by Arksey and Knight (1999: 54): ‘The quest for situational and conditional understanding is quite different from the quest of positivist social science. Consequently, the concepts of reliability and validity cannot be imported from positivist approaches to qualitative ones.’ Paradoxically, however, the qualitative researcher may derive some comfort from the world of experimental science itself, as Black (1937: 426) explains: ‘While the mathematician constructs a theory in terms of “perfect” objects, the experimental scientist observes objects
of which the properties demanded by theory are and can, in the very nature of measurement, be only approximately true.' For Black (op. cit.: 429): 'Vagueness is a feature of scientific as of other discourse.' Furthermore, in the world of atomic physics, Heisenberg (1959: 53) shows that an element of uncertainty is introduced into measurement by the very act of observation.

This is not to say, however, that validity and reliability can simply be ignored in qualitative research. Indeed, Freebody (2003: 168) asserts: 'Qualitative researchers in education need to take seriously the demands for reliability and validity', pointing to 'the enhanced purchase on validity and reliability that comes along with the less grandiose but more empirically defensible aims of understanding interviews as cultural practices about cultural practices' (Freebody, op. cit.: 169; emphasis in original). In the benchmarking project, the problems surrounding validity and reliability are considered to touch upon the study's trustworthiness. As summarized by Arksey and Knight (1999: 55): 'In a nutshell, the qualitative response to the issue of reliability and validity is to require researchers to demonstrate that what they do is fit for their research purpose.' The steps taken to achieve this in the benchmarking study are described below, in the discussions of interviewing and of documents.

3.3 The sample

As Arber (1993: 68) points out, sample design is determined by research goals. Arber (op. cit.: 71) identifies the two broad types of sampling method as probability sampling and purposive or non-probability sampling.

In probability sampling, each element of the population has a known probability of being
included in the sample (Burgess, 1993: 25), and the approach is best ‘if the researcher wishes to describe accurately the characteristics of a sample in order to estimate population parameters’ (Arber, 1993: 71) or ‘for analytic studies which involve testing empirical hypotheses’ (Arber, op.cit.: 71).

By contrast, non probability sampling can be used when, rather than seeking representativeness, ‘the researcher’s aim is to generate theory and a wider understanding of social processes or social actions’ (Arber, op.cit.: 71). Cohen and Manion (1994) identify four basic types of non-probability sampling. ‘Convenience sampling ... involves choosing the nearest individuals to serve as respondents and continuing that process until the required sample size has been obtained’ (Cohen and Manion, op.cit.: 88). In snowball sampling, individuals identified as having the characteristics required ‘are then used as informants to identify others who qualify for inclusion’ (Cohen and Manion, op.cit.: 89), and so on. Quota sampling ‘attempts to obtain representativeness of the various elements of the total population in the proportions in which they occur there’ (Cohen and Manion, op.cit.: 89). Finally: ‘In purposive sampling, researchers handpick the cases to be included in the sample on the basis of their judgement of their typicality. In this way, they build up a sample that is satisfactory to their specific needs’ (Cohen and Manion, op.cit.: 89).

Given the material constraints imposed on the benchmarking study, using probability sampling was not a practical proposition: accordingly, purposive sampling as defined by Cohen and Manion, (1994) was employed. Initial information about each of the four departments to be considered was obtained from the internet, each department publishing a full staff list. It quickly became apparent that it would not be feasible to deal with whole departments, as the numbers of academics concerned were too great. In particular, the
largest department, OldU Chemistry, proved to have approximately forty members of staff, organized in three distinct units. Consultation with the research supervisor produced the decision to interview four members of each unit of OldU’s Chemistry Department, plus the Head of Department. In OldU History and NewU Chemistry, each approximately twenty strong, five academics and the Head of Department were to be interviewed. In NewU History, with only ten members, four academics plus the Head of Department were to be interviewed. For both NewU Chemistry and NewU History, the Head of Department was not a member of the subject group, but each Head of Subject was included in the sample. In some instances, indications of age, and so of experience, were included on the website; in others, assistance was sought from the Head of Department. The identity of the person with a departmental quality role emerged through contact with the departments. This purposive approach of course raises the paradox of sampling: ‘A sample is misleading unless it is representative of the population; but how can we tell it is representative unless we already know what we need to know about the population, and therefore have no need of samples!’ (Rowntree, 1981: 23).

A clear consequence of this is that the study may lack generalizability: but, not only is this a characteristic feature of qualitative research (Marshall and Rossman, 1999: 193), as Firestone (1993: 16) observes: ‘Generalizability is clearly not the strength of qualitative research. ... Generalization from data is always problematic at best.’ However, Bryman (1989: 170) contends that ‘[m]ost qualitative research is in fact a form of case study’, and Yin (1994: 79), listing the six major sources of evidence in case study research as: ‘documentation, archival records, interviews, direct observations, participant observation, and physical artifacts’, points out (op.cit.: 90) that ‘[a]ny of the [six] sources of evidence
can and have been the sole basis for entire studies'. As Nisbett and Watt (1978: 8) point out, case studies may be used to help formulate hypotheses, an approach typical of interpretative qualitative research (Silverman, 1993: 21).

A further consequence of the sampling strategy is the relatively small size of the staff sample; but, as Silverman (1993: 10) points out: 'Qualitative interview studies are often conducted with small samples'. Arber (1993: 73) notes:

> Important sociological work is often based on relatively small samples drawn from one local area. Although these samples may attempt to be representative of a specific category of people, they are not probability samples from which precise inferences can be made about the characteristics of the population from which the sample was drawn. Using a probability sample is often unrealistic for small scale or qualitative research.

Similarly, Cohen and Manion (1994: 88) note that small-scale surveys often use non-probability samples because, in spite of the disadvantage of their non-representativeness, they are easier to set up and cheaper to implement.

It should be noted that the sampling strategy precluded gender sensitivity. Moreover, in the pilot institution, neither of the participating departments had any female members of academic staff, and in the study proper, by far the largest single department had only one female permanent full-time academic.

Given the wish to investigate the impact of benchmarking in higher education, it is necessary to delimit the study with a view to its feasibility (Lewis and Munn, 1997: 15), not only in terms of the time available (Bell, 1993: 21), but also with regard to the single researcher's limited resources (Silverman, 1993: 3). The constraints of the single researcher project made it pragmatically feasible to consider departments of chemistry and history in only two universities, one old and one new.
3.4 Ethical issues

Cohen and Manion (1994: 360) identify an ethical dilemma at the very heart of the research enterprise when they write of: 'the tension that exists between two sets of related values held by society: a belief in the value of free scientific inquiry in pursuit of truth and knowledge; and a belief in the dignity of individuals and their right to those considerations that follow from it.' In a similar vein, Kimmel (1988: 11) draws attention to the concern that: 'the basic scientist's "right to know" ... can conflict with the obligation to do no harm.' Such considerations confront the researcher with the difficult problem of the cost/benefit ratio, described by Cohen and Manion (1994: 347) as: '[a dilemma] which requires researchers to strike a balance between the demands placed on them as professional scientists in pursuit of truth, and their subjects' rights and values potentially threatened by the research.' Punch (1986: 81) points out that there is 'no hard-and-fast way to calculate the costs and benefits of social scientific research', but a sense of perspective can perhaps be achieved by considering what would doubtless be most people's intuitive approval of the advice given by the composer Janáček on the project to make field recordings of Moravian folk-songs between 1909 and 1912 (quoted in Broughton, 1999: 106): 'Do not force the song out of the singer, do not change his environment, do not discourage him by your presence, or by the polished microphone of the phonograph. The recording will be most natural when you listen hidden, unnoticed, to the song.'

Fontana and Frey (1994: 372) observe in connection with interviewing: 'Traditional ethical concerns have revolved around the topics of informed consent (consent received from the subject after he or she has been carefully and truthfully informed about the research), right to privacy (protecting the identity of the subject) and protection from harm (physical,
emotional, or any other kind)' (emphases in original). Bell (1993: 52), stressing the importance of informed consent in negotiating research access, points out: 'People will be doing you a favour if they agree to help, and they will need to know exactly what they will be asked to do, how much time they will be expected to give and what use will be made of the information they provide.' For Arksey and Knight (1999: 129), the purpose of informed consent is: 'to safeguard participants' privacy and welfare, and to give them a choice about whether or not to take part in a study.' They (Arksey and Knight, op.cit.: 129) recommend that informants taking part in qualitative interviews should be made aware of:

- The purpose and nature of the study, including the research methods and timing.
- The anticipated benefits, risks or costs (for those taking part in the study, and also the wider society).
- Contact details of the researcher, and the research base.
- The names of any funders or sponsors.
- The sorts of questions being asked, and how long the interview should take.
- Their right not to answer specific questions, or to change their mind and withdraw from the research altogether.
- The degree of anonymity and confidentiality: what information will be disclosed, for what purposes and to whom; the use of quotations.
- Arrangements to safeguard confidentiality.
- Plans regarding dissemination .... Graduate students undertaking research ought to let informants know that the resulting dissertation is a public document which can be obtained from libraries. If the eventual aim is to produce a book version of the thesis, then this should also be stated.

The approach adopted in the benchmarking project was Bell's (1993: 52) simple injunction: 'Be honest', particularly by seeking informed consent, including an explanation of the research as part of a formal academic qualification (cf. also 3.5.ii). When explaining informed consent to participants, it was additionally pointed out that no value was attached to their responses, in the sense that they might be seen as correct or incorrect, approved or disapproved. Given this, respondents were requested to say what they thought, rather than what they thought the researcher might like to hear. Roth (1961: 283; quoted in Burgess,
1985: 144) observes ‘All research is secret in some ways and to some degree - we never tell the subjects “everything”; nevertheless, some exposition of the assumptions underlying the research was given to the respondents after the interview, if time permitted.

Qualitative research is inevitably influenced by value systems in the sense explained by Usher and Scott (1996: 179), for whom education research ‘cannot be free of the value commitments of those who conduct it: social research is always valued research.’ Kelman (1972: 991) indicates a particular danger of this state of affairs: ‘The investigator usually defines and takes charge of the situation on his own terms and in line with his own values and norms, and the subject has only limited opportunity to question the procedures.’ Research interviewing, then, is based on what Mishler (1986: 54) calls an ‘asymmetry in power’, with the balance in favour of ‘interviewer dominance’ (Mishler, op. cit.: 54). Indeed, for Kelman (1972: 989): ‘The ethical problems surrounding social research ... can be conceptualized in terms of the power relationships between the subjects of social research, on the one hand, and the social scientist, as well as the sponsor and user of social research, on the other hand.’ In view of this, it is incumbent upon the researcher not to lose sight of the ethical frameworks and value systems of the subjects, and also not to be indifferent to the subjects’ predicament both during and after the research project, either in terms of known or expected effects on the subjects (Homan, 1991: 160), or in terms of unintended effects of the research (Ball, 1997).

A further difficulty stems from the notion that ‘[e]ducational research ... is always political’ (Usher and Scott, 1996: 179), in that each participant can be seen as a micropolitical case requiring extensive explication in order to be interpreted with confidence. A parallel may be seen in Black’s (1937: 427) observation: ‘The line traced by a draughtsman, no matter how
accurate, is seen beneath the microscope as a kind of corrugated trench, far removed from
the ideal line of pure geometry.' This same general problem underlies the caution of
Easterby-Smith et al. (1991: 64): 'the researcher needs to be aware of conflicts that may be
far deeper and more complex than will be evident to a relative newcomer in the
organisation.' A dilemma potentially facing the researcher is well summed up by Ball (1991:
181) in relation to personal questioning: 'How much does the researcher need to know, as
against how much should they be allowed to know?'

As Robson (1993: 30) reminds: 'ethical dilemmas lurk in any research involving people'
(emphasis in original). When the research is carried out by means of interviews, Fontana and
Frey (1994: 372) point out: 'Because the objects of inquiry in interviewing are human
beings, extreme care must be taken to avoid any harm to them.' The care thus called for
expresses itself in ethical behaviour, defined by Cohen and Manion (1994: 359) in the
following terms: 'Whatever the specific nature of their work, social researchers must take
into account the effects of the research on participants, and act in such a way as to preserve
their dignity as human beings.' So important is the need for ethical behaviour that, for Miles
and Huberman (1994: 288): 'Any qualitative researcher who is not asleep ponders moral
and ethical questions'. Kimmel (1988: 34), however, stresses that 'the distinction between
ethical and unethical behavior is not dichotomous', that there is, rather, a continuum ranging
from the clearly ethical to the clearly unethical. Moreover, as Smith and Berard (1982: 210)
cautions: 'We do not assume that a certain practice is ethical merely because subjects do not
object to it.' Indeed, in Busher's (2002: 86) view: 'Fully ethical research is impossible to
achieve'.

87
3.5 Research instruments

3.5.1 Interviews

As Burgess (1993: 2) points out: 'There is no “best” method of conducting social research. Instead, researchers need to consider the kind of research question they wish to pose and the most appropriate methods of research and techniques of data collection.' In this study, the aim is to obtain individuals' views through the use of a technique encouraging the respondents to comment on the issues under consideration, and as Kvale (1996: 105) stresses: 'interviews are particularly suited for studying people’s understanding of the meanings in their lived worlds, describing their experiences and self-understanding, and clarifying and elaborating their own perspective on their lived world.' For Arksey and Knight (1999: 33), the 'main rival' of the interview is the self-administered questionnaire; however, 'such questionnaires are markedly inferior [to interviews] as a way of exploring the world of informants. Social interaction is important here, as is the scope to explore informants’ answers with them' (Arksey and Knight, op.cit.: 17). Discussing the relative merits of interviews and questionnaires, Cohen and Manion (1994: 283), point out that questionnaires tend to be more reliable, as they encourage greater honesty through anonymity. In an interview, however, the interviewer can clarify questions and the purpose of the interview, can offer reassurances, and 'an interview can be conducted at an appropriate speed whereas questionnaires are often filled in hurriedly' (Cohen and Manion, op.cit.: 283). Bryman (1989: 147) indicates a general tendency in the choice of interview format: 'By and large, quantitative research tends to use interviews at the structured end of the continuum; unstructured interviews tend to be used in the context of qualitative research'. Broadly following this categorization, the benchmarking study uses semi-structured interviews based on a standardised interview schedule offering both a consistent
framework for information gathering and a structure making it possible to use interview
time effectively and efficiently: 'A semi-structured interview schedule tends to be the one
most favoured by educational researchers as it allows respondents to express themselves at
length, but offers enough shape to prevent aimless rambling' (Wragg, 2002: 149).

The construction of the interview schedule for the benchmarking study was informed by an
initial survey of the literature, which produced approximately 60 thematic headings (cf.
Miles and Huberman, 1994: 58). On the assumption that busy academics and administrators
would be unlikely to be willing to agree to participate in an interview predicted to last
longer than 30 minutes, an assumption proven by the fieldwork to be correct, the original
list of headings was reduced, first to approximately 35, and then to around ten focusing on
the most essential issues manageable, allowing the construction of the schedule. The
interview schedule used with academic staff is given in Appendix 1.

For Arksey and Knight (1999: 52), using interview schedules containing 'questions drawn
from the literature and from pilot work with respondents' enhances the validity of interview
research. Mishler (1986: 64), however, stresses: 'serious questions must be raised about the
validity of the usual assumption of a "community of meaning" between researchers and
respondents and among respondents.' The problem is that 'data recording is necessarily
selective and always involves some interpretation, however minimal' (Hammersley and

In order to seek 'to gain entry into [the participants'] worlds' (Marshall and Rossman,
1999: 194), the interview schedules used for the various respondent groups contained open
questions, closed questions and scale items (Robson, 1993: 233), though even after closed
questions and scale items, respondents were invited through careful probing to comment on their answer, thus giving a fuller picture (Cohen and Manion, 1994: 277). This does, of course, also raise certain difficulties. Silverman (1993: 114), for instance, notes: ‘Interviews share with any account an involvement in moral realities’; whilst Finch (1986: 209) records: ‘The idea that all social science knowledge is ultimately political and not neutral is a well-established position in sociology.’ Given this, it is perhaps not surprising that the researcher might be viewed at times as a ‘a tool of management’ (Bryman, 1989: 4), ‘an undercover agent’ (Hammersley and Atkinson, 1995: 81), or as some sort of spy (Horn, 1996: 551).

In accordance with recommended practice (Fielding, 1993; Foddy, 1993; Fowler, 1993; Robson, 1993; Wengraf, 2001), the interviews were piloted, using a sufficiently wide range of respondents to enable the research instruments to be adequately tested (Arber, 1993: 72). The pilot study was conducted in the departments of chemistry and history at a new university (not NewU), generating interviews with two Heads of Department and four members of academic staff. Neither department had any inexperienced staff, though the two staff members from history were relatively less experienced than the two chemists. Cohen and Manion (1994: 374) see the establishing of good relations as the key to resolving ethical and methodological questions in education research. The pilot phase was an excellent opportunity to refine the skills required to achieve this: the acceptability of the questions and manner of questioning was confirmed by the piloting. A further benefit of the piloting was the acquisition of a sense of the pace needed to complete the schedule in the time available. In spite of the very large investment of time needed (Arksey and Knight, 1999: 105; Burgess, 1984: 121), all the pilot interviews were transcribed in full. Approximately one third of the respondents opted to check transcripts, but no respondent asked for any
change to be made, an outcome helping to build up the cumulative confidence that other transcripts which were not checked were also accurate records of what was said. Whilst the pilot produced no ‘nasty surprises’ (Oppenheim, 1992: 64), one respondent (Pilot/Acad: C1) preferred not to answer one question (Question 8: ‘Can I pass on that?’); but even then, the rest of the schedule was completed.

3.5.ii Data gathering

Potential respondents were sent a letter explaining the project briefly and giving a full assurance of adherence to the principles of informed consent, particularly anonymity. A subsequent telephone call to fix an appointment for the interview also provided an opportunity to give reassurances that no knowledge of subject benchmarking was necessary for a valuable contribution to the project. Explanations of the motivation behind the research topic were delayed until after each interview had taken place. On the one hand, this avoided the danger of suggesting a perspective to respondents and thus potentially contaminating data, and on the other hand, it favoured the collection of unprepared, spontaneous data.

However, the interview is ‘an unnatural situation’ (Measor, 1985: 67) and, as Robson (1993: 243) observes, it is ‘a complex social interaction’. On the one hand, it is inconceivable, as Brenner (1978: 138) points out, that: ‘the interviewer could ever not influence the situation of action in the interview by means of his own performance’ (emphasis in original); and on the other hand, the interviewer, by entering into a reciprocal relationship, is also subject to external influence: ‘In an interview, like any other instance of social interaction, both individuals influence each other’ (Powney and Watts, 1987: 35). Nevertheless, as Kvale (1996: 20) indicates: ‘In professional interviews there is usually an
asymmetry of power ... In contrast to the reciprocal interchanges of everyday and philosophical conversations, there tends to be a one-sided questioning of the subject by the professional. Hammersley and Atkinson (1995: 152) remind that the research interviewer 'has a research agenda and must retain some control over the proceedings.'

In the benchmarking study, scrupulous adherence to the principles of informed consent sought to mitigate against '[t]he pattern of interviewer dominance and respondent acquiescence' (Mishler, 1986: 54). Of particular importance here were the assurances that respondents could withdraw from the project at any point, and that they could seek clarification of any aspect of their involvement.

In order to record the 'spoken texts' (Brown and Yule, 1983: 9) produced during the interviews, note-taking and tape-recording were available: video-taping was neither feasible nor appropriate (Arksey and Knight, 1999: 106). Note-taking has the advantage of being cheap (Arksey and Knight, op. cit.: 105), but has the disadvantage that 'it may disrupt the effectiveness of the communication between interviewer and respondent' (Borg and Gall, 1979: 314). Moreover, note-taking: 'can be slow, is open to charges of selective recording and requires practice and skill' (Arksey and Knight, 1999: 105). Tape-recording, by contrast: 'reduces the tendency of the interviewer to make an unconscious selection of data favoring her biases' (Borg and Gall, 1979: 315). Additionally: 'The interviewer can concentrate on what is said. There is a permanent record that captures the whole of the conversation verbatim, as well as tone of voice, emphases, pauses and the like' (Arksey and Knight, 1999: 105). Indeed, as Burgess (1984: 120) points out: 'without a tape-recorder much important data would be lost.' Accordingly, the preferred method of data recording during the fieldwork was on audiotape, with note-taking as a stand-by.
Each interview schedule in the benchmarking project ended with a question as to whether or not the respondent wished to check a transcript of the interview. As Poland (1995: 292) observes, however: 'the very notion of accuracy of transcription is problematic given the intersubjective nature of human communication, and transcription as an interpretive activity.' Nevertheless, such checking was considered to be not only a way of enabling those respondents who so wished to check and control data being held about them, but also a source of 'respondent validation' (Silverman, 1993: 159). Comments made on tape after the completion of the schedule were counted as data (Platt, 1981a: 79) and included in the transcript: similarly, comments made after the tape recorder had been switched off were included in summary form at the end of the transcript.

3.5.iii Documents

For the benchmarking project, access was sought to departmental and institutional documentation. This was partly in order to seek triangulation (Marshall and Rossman, 1999: 194) of issues discussed with interviewees, and partly in order to seek supplementary data and insights. As Finley (1987: 15) observes: 'Records and documents, record-keeping and archives are a function of the society which produces and preserves them'.

University documents are, in Bell's (1993: 68) terminology: 'Inadvertent sources, which are used by the researcher for some purpose other than that for which they were originally intended' (emphasis removed). Moreover, they fall into Scott's 'restricted' category: "'restricted' documents are accessible on an ad hoc basis under specified conditions to those outsiders who are able to secure the permission of insiders; they are therefore, normally closed to outside access, though their authors or custodians may be willing to grant access on application" (Scott, 1990: 14). But, as Bell (1993: 69) warns: 'it cannot be
discounted that inadvertent documents were intended to deceive someone other than the researcher, or that what first appear to be inadvertent sources (some government records, for example) are actually attempts to justify actions to future generations’. It is important, therefore, to remember that documentary sources are also socially produced (Macdonald and Tipton, 1993: 188), and should be approached bearing in mind three questions (Jones, 1994: 6):

- Who wrote or created the document, and for what purpose?
- When was it written or produced?
- For what audience was it created?

This in turn complicates the application of Scott’s (1990: 19) quality control criteria for assessing documentary sources: ‘authenticity, credibility, representativeness and meaning’, especially as any access is completely subject to institutional control. The search for Finley’s (1987: 18) ‘stigmata that automatically distinguish fiction from fact’ may therefore be fraught with difficulty, but Bell (1993: 69) makes a useful distinction in the types of evidence available from documents: “’Witting’ evidence is the information which the original author of the document wanted to impart. ‘Unwitting’ evidence is everything else that can be learned from the document ... All documents provide ‘unwitting’ evidence, but it is the task of the researcher to try to assess its precise significance’ (emphases removed).

As Platt (1981b: 45) observes: ‘To treat all accounts as merely “accounts” solves some problems but ... does so at the cost of ignoring the reason for being interested in the matter in the first place.’

Whilst it was considered highly improbable that any of the documents released for the benchmarking study would lead to problems of ‘guilty knowledge and dirty hands’ (Fetterman, 1998: 143), the possibility nevertheless remained that they might in some way
call into doubt deeply held beliefs regarding, for instance, the institution's officially promulgated educational philosophy or the nature of strategic decisions. As is amply demonstrated in studies such as those by Ball (1991), Burgess (1985) or Punch (1986), a researcher may have severe problems with the presentation and publication of results not conforming with the institution's own views. Similarly, making public certain information about institutions could damage the position of those institutions in the educational marketplace, either by raising doubts about such things as quality and standards, or by undermining the institution's market position in some other way. Institutions need public faces, not only in order to attract students, but also in order to maintain good relations with external agencies, the media, funding bodies, local and national government, and so on. This public face will be largely created and interpreted by actors within the institution, and any finding indicating a conflict between the appearance and the reality of an institution's activities needs to be treated carefully. As Platt (1981b: 49) concludes: 'there are important senses in which documentary research has problems which are not significantly different from those of research using other data sources.' Once more, the rôle of informed consent in the benchmarking study is crucial to an ethically sound investigation.

3.6 Analysis of data

Flick (2002: 176) stresses: 'The interpretation of data is at the core of qualitative research'. In the case of semi-structured interviews: 'Interpretation is not fixed to a specific method. But coding procedures ... seem to be most appropriate' (Flick, op.cit.: 79). Miles and Huberman (1994: 56) define codes as 'tags or labels for assigning units of meaning to the descriptive or inferential information compiled during a study.'

For Miles and Huberman (op.cit.: 56): 'Coding is analysis.' That analysis may be continued
through the process of thematic analysis, which is ‘not a separate method ... but something to be used to assist the researcher in the search for insight’ (Boyatzis, 1998: vi). A theme is ‘a pattern found in the information’ (Boyatzis, op.cit.: vii) and, as Boyatzis (op.cit.: vii) explains: ‘A theme may be identified at the manifest level (directly observable in the information) or at the latent level (underlying the phenomenon). The themes may be initially generated inductively from the raw information or generated deductively from theory and prior research.’ In the benchmarking study, the brief and limited access to respondents inevitably means that any insights into factors shaping attitudes must remain heavily conjectural.

For Powney and Watts (1987: 161), the task of the analyst is: ‘to work through the data and to re-present it in a form that can be appreciated by the intended audience.’ Ryan and Bernard (2000: 780) emphasize that ‘no particular tradition, whether humanistic or positivistic, has a monopoly on text analysis.’ Neuendorf (2002), for instance, advocates content analysis, defined as: ‘the systematic, objective, quantitative analysis of message characteristics’ (Neuendorf, op.cit.: 1). It is this quantitative foundation which rules out content analysis for the benchmarking study: ‘by most definitions, [content analysis] fits the positivism paradigm of social research’ (Neuendorf, op.cit.: 11). Brown and Yule (1983: 1) point out that the analysis of discourse is ‘necessarily, the analysis of language in use.’ For the researcher using semi-structured interviews, however, there is immediately a problem relating to naturally occurring talk, defined by Potter (1997: 148) as: ‘spoken language produced entirely independently of the actions of the researcher.’ In the benchmarking study, the context established by the researcher together with the potential linguistic channelling resulting from the pre-determined interview schedule combined with prompting
and probing make it impossible to assume the production of the respondent’s naturally occurring language, especially within such tight time constraints. Therefore, discourse analysis is also ruled out for the study. As Potter (op. cit.: 150) notes: ‘Discourse analysts are increasingly turning away from interviews to focus on materials less affected by the formulations and assumptions of the researcher’

An initial problem of analysis is the sheer bulk of the data produced by a qualitative study (Arksey and Knight, 1999: 161; Miles and Huberman, 1994: 56). Powney and Watts (1987: 158) state that analysis is ‘an act of constructive interpretation’ and is to be understood as ‘a process of gains and losses’ (Powney and Watts, op. cit.: 143). The process of analysing qualitative data: ‘usually begins at the design stage ... and continues, albeit informally, while the interviews are being done’ (Arksey and Knight, 1999: 161). Hammersley and Atkinson (1995: 204) add that, in ethnography, the analysis of data ‘continues through to the process of writing reports, articles, and books.’ A similar lack of ‘a distinct stage’ (Hammersley and Atkinson, op. cit.: 205) for data analysis is assumed in the benchmarking study.

After an initial indexing of the data, analytic categories were developed by studying the texts (Arksey and Knight, 1999: 163). Given the relative brevity of the sections of text resulting from fairly short interviews, the coding was descriptive and interpretative, rather than pattern (Miles and Huberman, 1994: 57). Techniques such as linguistic repertoires (Wooffitt, 1993: 292) or paralinguistics (Wengraf, 2001: 216) were not employed, for similar pragmatic reasons. Meaning was generated from the data bearing in mind Miles and Huberman's (1994: 245) 13 tactics. The results of this analysis are presented discursively and, as appropriate, in tabular form (Marshall and Rossman, 1999: 197).
Some of the questions included in the interview schedules also involved simple Likert scales (Bell, 1993: 139). As has already been indicated, these questions were used partly as an attempted precursor to reflection on the issues. However, they also helped to address a problem identified by Silverman (1993: 162): 'The critical reader is forced to ponder whether the researcher has selected only those fragments of data which support his argument. Where deviant cases are cited and explained ..., the reader feels more confident about the analysis.' As Silverman (op. cit.: 163) explains:

> simple counting techniques can offer a means to survey the whole corpus of data ordinarily lost in intensive, qualitative research. Instead of taking the researcher’s word for it, the reader has a chance to gain a sense of the flavour of the data as a whole. In turn, researchers are able to test and to revise their generalisations, removing nagging doubts about the accuracy of their impressions about the data.

In line with this rationale, the outcomes of the counting techniques are presented essentially in tabular form, whilst also informing the discursive analysis.

It was intended that such straightforward numerical techniques would also be applied to the documents made available in the benchmarking study. On the one hand, it is significant to know how many departmental and institutional documents relate to subject benchmarking, and whether documents are discussion papers, minutes, directives, guidelines or something else; on the other hand, it is important to note the frequency, nature and extent of references to subject benchmarking.

Situational and contextual considerations are also important, including: 'How are documents written? How are they read? ... For what purposes? On what occasions? With what outcomes?' (Hammersley and Atkinson, 1995: 173). Here, as elsewhere: 'The key factors are to work with the data and to keep in mind the goals of the project' (Arksey and Knight, 1999: 169).
3.7 The study

This study adopts a qualitative methodology, seeking to understand respondents' perceptions of subject benchmarking and its impact. Based on a small population (n = 34) selected by purposive sampling, the study makes no claim to generalizability, but does aim to achieve trustworthiness.

The principal method of data collection is the semi-structured interview, producing transcripts offered to the respondents for checking, verification and amendment. Additional data and insights are sought from documentary sources, an approach also affording triangulation of results. All data gathering is firmly based on informed consent.

Descriptive and interpretative coding of the data collected permits the generation of analytic categories. Whilst the results of the analysis are presented largely discursively, additional texture is provided by using simple counting techniques generating results presented in tabular form.

The findings derived from the data collection and analysis are presented in the next chapter.
Chapter 4

Findings

In the fieldwork for the benchmarking study, data were obtained from one old university (OldU) and one new university (NewU). Within each university, data were gathered from the departments of chemistry and of history, and from the centres with an institutional responsibility for quality. This produced 34 semi-structured interviews: 30 with academic members of staff, and four with institutional quality administrators. Where possible, departmental and institutional documents relating to subject benchmarking were consulted.

4.1 The respondents

Individuals were identified as possible respondents either from internet sources or by personal recommendation within a department or administrative unit. At OldU, only three approaches for participation were declined, all very quickly and by e-mail: at NewU, just one person was unwilling to participate, the reluctance,
expressed in the initial telephone conversation, being based on a lack of familiarity with subject benchmarking. Even though it was explained that an interview would nevertheless produce interesting and valuable results for the study, the reluctance persisted and so the approach was abandoned.

Of the 30 academics interviewed, ten claimed no familiarity with subject benchmarking (seven from OldU, three from NewU); though, of those ten, three (all from OldU) indicated that they knew that subject benchmarking had been introduced. For biographical data, cf. 4.1.ii and Table 4.1.

The semi-structured interviews were conducted using one of three standard schedules. Appendix I gives the schedule used with academics (varied only marginally with Heads of Department), and includes prompts: the version used with members of institutional quality services was also very similar, but assumed, for example, that the respondent would have experience of working with the old QAA system of Subject Review.

Time constraints meant that some interviews were a little hurried, particularly for the later questions, so that it was not always possible to cover all issues in a uniform way. Attempts to obtain clarification of details and/or additional comments subsequent to interview also met with levels of success varying from no response at all at one end of the scale to short texts at the other. Very occasionally, interviews were interrupted by telephone calls or knocks at the door, which did cause some loss of data (eg. NC5, whose train of thought was irretrievably broken at one point). Similarly, there were occasional instances of responses not clearly matching the
stimulus question, or being shaped by the respondent's preoccupations rather than the interviewer's objectives. If gentle probing achieved no shift, no further attempt was made to impose the researcher's agenda on the interview.

4.1. i Identifier codes

All of the respondents agreed to the use of three part identifiers for reference and quotation purposes: OldU and NewU have identifiers beginning O and N respectively, followed by either C for chemistry or H for history, and then a number, to which no hierarchical significance is attached (eg. OC1). Departmental heads are identified by H instead of a number (eg. OHH). Quality administrators (Q) are identified by E for enhancement or S for standards (eg. NQS).

At NewU, neither departmental head was a member of the academic discipline under consideration: the subject heads of chemistry and history are NC4 and NHI respectively.

4.1. ii Academic departments

As indicated in Chapter 3, resource and time constraints made it impossible to interview whole departments or to introduce gender as a variable for analysis. Nevertheless, in an attempt to identify differences in perceptions of subject benchmarking perhaps attributable to age and experience, and thus to differing generation-related expectations, the fieldwork sought the views of both experienced and inexperienced members of staff. In all four departments, experienced staff greatly outnumbered inexperienced staff: NewU History had only one inexperienced member. Of the 30 academics interviewed, 19 had ten years'
experience or more of teaching in UK HE (an arithmetic mean of 24 years); one had eight, and one seven years' experience; nine had five years' experience or less (an arithmetic mean of 2.8 years). In some instances, respondents indicated that some of their teaching experience was part-time and/or undertaken as a postgraduate student. Other than noting the information, no attempt was made to factor this part-time experience into the data explicitly.

Amongst the 21 academics with teaching experience amounting to seven years or more, 16 had been in their current post for ten years or more, one for seven years, one for three years and three for two years. Amongst the remaining nine academics, one had been in post for five years, two for three years, two for two years, three for one year, and one for approximately six months. Table 4.1 shows the distribution of academic staff by level of experience: five years' experience or less is classified as 'inexperienced'.

<table>
<thead>
<tr>
<th>Table 4.1 Experience of Academic Staff</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th></th>
<th>OldU Identifier</th>
<th>Total</th>
<th>NewU Identifier</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Chemistry</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experienced</td>
<td>OC1, OC3, OC5,</td>
<td>9</td>
<td>NC1, NC3, NC4,</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>OC6, OC7, OC9,</td>
<td></td>
<td>NCH</td>
<td></td>
</tr>
<tr>
<td></td>
<td>OC11, OC12,</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>OC13</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>OC2, OC4, OC8,</td>
<td>4</td>
<td>NC2, NC5</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>OC10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inexperienced</td>
<td></td>
<td>4</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>All</td>
<td>13</td>
<td></td>
<td></td>
<td>6</td>
</tr>
<tr>
<td><strong>History</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experienced</td>
<td>OH2, OH4, OH5,</td>
<td>4</td>
<td>NH1, NH2, NH3,</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>OHH</td>
<td></td>
<td>NHH</td>
<td></td>
</tr>
<tr>
<td></td>
<td>OH1, OH3</td>
<td>2</td>
<td>NH4</td>
<td>1</td>
</tr>
<tr>
<td>Inexperienced</td>
<td></td>
<td>6</td>
<td></td>
<td>5</td>
</tr>
<tr>
<td>All</td>
<td>13</td>
<td></td>
<td></td>
<td>6</td>
</tr>
</tbody>
</table>

Experienced = more than five years teaching in UK HE
One anomaly occurred in the constitution of the sample. When granting permission to undertake research in the department, the Head of OldU Chemistry (personal communication) wrote: ‘I would, however, emphasize that I have never been involved in defining and benchmarking chemistry, strange as this might seem to you.’ At the same time, however, the Head of Department also supplied the names of two colleagues ‘who are veritable “experts” in this area and I feel you should contact them directly with my approval’ (personal communication). One of these colleagues proved to be co-ordinating the production of Programme Specifications within the department, and the other to be the person with departmental responsibility for quality issues. The Head of Department also cited a third colleague, ‘who is now retired, is also someone whom you should speak with’ (personal communication). Each of these three individuals agreed to be interviewed, the retired academic being not only an eminent chemist, but also, as emerged during the interview, a former Head of Department at OldU. This last piece of information became particularly significant when, in spite of repeated efforts, it proved impossible to acquire fieldwork data of any kind from the existing Head of Department. Whilst the retired academic had been initially approached partly in order to avoid a potential discourtesy to the actual Head of Department, and had been interviewed thanks to consent being granted immediately, it was not at first clear what status should be accorded to the resulting data. However, in the absence of data from the current Head of Department, the retired academic was included as a full member of the sample, able to give insights not only into the national academic and professional contexts, but also into the departmental ethos of
OldU Chemistry.

In the remaining three departments, the Head of Department agreed to be interviewed, as did the person with local responsibility for quality. Without exception, the person with the departmental quality remit was an experienced member of staff.

4.1.iii Administrative units

Both OldU and NewU proved to have central quality services configured in two sections, covering professional standards and enhancement. The Head of Standards at OldU had thirteen years’ HE experience, eleven of them in quality, had been at OldU for nine years, but Head of Unit for only one month. It was the OldU Head of Standards who recommended a member of the Enhancement Unit, a colleague with four years’ HE experience, all gained in the current post. The Head of Standards at NewU had 31 years’ experience as an academic in HE, all at NewU and 20 of them including direct involvement in quality issues, but had been Head of Unit for only one month. Similarly, the individual at NewU with overall responsibility for quality enhancement had initially been an academic, with 19 years’ experience overall in HE, 11 of them at NewU, four years involving a principal focus on quality issues, two of those being in the centre guiding quality enhancement, and one as Head of Centre.

4.2 Documents

Neither OldU nor NewU was able, at the time of the fieldwork, to provide any substantial documentation relating to subject benchmarking.
4.2.1 Departmental documents

*OldU Chemistry* The academic with a co-ordinating role for quality issues (OC10) undertook to provide such departmental documentation as was available relating to the introduction of subject benchmarking, though, some seven months after the original interview, nothing was forthcoming. At that point, OC10 observed in a telephone conversation: 'We've not actually formally written down that sort of thing, I don't think.' He referred, however, to 'an internal q.a. [quality assurance] visit' carried out approximately two years earlier, at a time when he had been absent himself, and offered to check with the departmental administrator about documentation produced for that event. One year after the original interview, still nothing had been made available: 'We're having trouble locating the documents' (OC10: telephone conversation). OC3, who proved to be Chair of the department's Teaching Committee, stated that benchmarking had had 'a huge impact' (emphasis in original) on that committee, subsequently explaining (e-mail) that this had been because of the need to produce Programme Specifications. However, when asked about benchmarking-related documents emerging from that committee, OC3 (private communication) reported: 'nothing documented'.

Ultimately, and in spite of prolonged negotiations, no documentation whatever was made available by OldU Chemistry.

*OldU History* Just before the point was reached in the fieldwork interview at which access was to be sought to documents recording the History Department’s handling of subject benchmarking, OH5, the academic responsible for quality issues,
observed: 'I thought you were going to ask me about the way it was introduced here, and what we actually did mechanically, in the Department: and I would have had to have said [laughs], 'I've no idea, 'cause I wasn't here.' OH5 added, however: 'but I have a sense that this department took it pretty seriously, that there were a lot of discussions, that they did give careful thought to: 'Do we fit in?' ... But, I think the conclusion was: 'Well, yes, we probably do.' In other words, as OH5 remarked after the interview, it was not that things had not been done in the department, just that proper records had not been kept. Nevertheless, OH5 did give permission for departmental documentation to be consulted, suggesting that a better time might be after a forthcoming QAA review. According to OH4 (e-mail), however, the Chair of the departmental Teaching Committee: 'I don't think there was a time when we officially discussed the [History Benchmarking Statement], so I cannot think of any records to consult.'

Three and a half months after the original interview, OH5 was again approached about departmental documentation, and replied: 'I've been asking and rooting around the Department, and literally nothing seems to have survived of formal meetings on Benchmarking. ... Even [OH2], the Benchmarking guru, can find nothing (though he remembers it well enough)' (e-mail). Three days later, OH5 (e-mail) further noted: 'I've been collecting some stuff for Subject Review and, again, it's surprising how little there is directly on benchmarking.'

Several months later, and now almost a year after the original interview, OH5, contacted in the hope of gaining access to the internal documentation prepared for
the QAA visit, suggested approaching directly OH2, but OH2 (e-mail) replied: 'I can say that there is no departmental documentation re subject benchmarking.' Subsequently, however, OH2 (private communication) commented that, in preparing the Self Evaluation Document for the QAA visit, OldU History 'tried hard to conform as closely as possible to the H[istory] B[enchmark] S[tatement]. This tells you how influential the HBS was in determining our thinking.' As OH4 put it: 'we used the benchmarking as a Bible [laughs].'

OH4 provided the Programme Specification for the BA (Hons) History and a sample Module Specification: the former merely mentions the QAA subject group into which the programme falls, whilst the latter makes no mention at all of benchmarking.

*NewU Chemistry* NCH was unable to say what material existed in the department arising from the advent of subject benchmarking, as he had not been Head of Department at the time, though he did undertake to give access to a document setting out generic features of a set of related degree programmes, 'which would show you where things like learning outcomes and so on relate to benchmarks.' In the event a photocopy of just one page of the document was provided; though it does indicate that programme modifications introduced in the document resulted from several factors, including 'the interpretation of Chemistry Benchmarking statements'.

The Head of Chemistry (NC4) was 'aware of statements [from the QAA benchmarking document] which were then faithfully reproduced' in NewU
Chemistry’s own documentation.

A later request to NC4 for the Programme Specification for the BSc (Hons) Chemistry produced some additional pages from the programme document just mentioned, from which it emerged that the chemistry benchmark statement (QAA, 2000c: 3) bullet points under ‘Subject knowledge’ and ‘Abilities and skills’ had been adopted almost verbatim.

A grid for a skills mapping exercise, developed and provided by NC1, the academic with responsibility for quality issues, also adopts almost the precise wording of the benchmark statement for chemistry: ‘I mean: the first thing, we were very keen to show that we were complying with the benchmarks’ (NC1). NC3 provided a copy of a similar grid as an example of the extra work occasioned by benchmarking. Samples were also provided of a Student Module Guide and a Module Specification Proforma: neither mentions subject benchmarking.

Requests to the chemistry subject group for documentation concerned with benchmarking went unanswered: it is therefore unclear whether any existed. The supposition that there may actually be no such documentation is partly borne out by the statement from NC1 that an exercise mapping NewU courses against the chemistry benchmarking statement was not supported by anything in writing: ‘[NC4] and I just sat down and did it.’

NewU History When interviewed, NHH was ‘almost certain’ that there was no departmental documentation on benchmarking, but was ‘almost certainly’ willing
to grant access to documents 'if I can think of any.' No such documents were ever forthcoming. The Head of History (NH1) undertook to check minutes from subject group meetings, but discovered that 'there's almost nothing there. What we've got is really quite minuscule [being in the nature of] [name of academic] discussed benchmarking' (NH1: telephone conversation). An approach to NH2, the academic with responsibility for quality issues, produced the reply: 'the problem is that this paper trail is through a range of different committee minutes. I don't recall a single document which deals with this issue' (NH2: e-mail), and no documents were forthcoming. A copy was provided of the Programme Specification for the BA (Hons) History, but it contained no reference to benchmarking.

4.2.ii Institutional documents

OldU OQE provided certain standard documentation mentioned in the course of the interview. The documents, almost exclusively sections of OldU’s institutional quality manual for academic staff, include a brief section on subject benchmarks, with the statement: 'An essential component of the course review process leading to the preparation of programme specifications is to ensure that the course(s) reflect the relevant QAA subject benchmark statements.'

OQS observed that benchmarking had led to changes in OldU’s approach to quality, 'largely in terms of the sort of documentation we expect our departments to create'. Asked if there was central guidance on what departments were expected to do, OQS replied: 'At a general level. Again, in terms of the form of documentation we expect Schools to produce, and the areas they're expected to address; but no more specific
than that.' Moreover: 'The only direction [given to Schools] is that they should address [the benchmark statement]. It doesn't mean they should meet it, match it. If they choose, if there are bits they just don’t like and choose to ignore, that's for them' (OQS). All institutional guidance, OQS explained, was contained in OldU's Quality Manual: OQS provided copies of the sections of the Quality Manual setting out the format for Programme Specifications and Module Specifications. Programme Specifications are required only to state the QAA benchmarking group to which the programme belongs; the Module Specification Template Guidance makes just one reference to benchmarking, in the section on learning outcomes: 'You should refer to the relevant subject Benchmark statement(s), any professional or statutory requirements and the University Learning & Teaching Strategy as a stimulus to reflection.'

In a follow-up e-mail, OQS was asked about documents from the Departments of Chemistry and History showing how they had taken their benchmark statements into account, possibly the sort of thing to be used in evidence to the QAA, replying: 'I don’t think there is anything to be honest' (OQS: e-mail).

NewUNCI explained that, when benchmark statements appeared, 'our institution, as part of its internal review [of] quality management procedures, became very keen that people took account of benchmarking.' No documentary evidence of this was forthcoming. Asked if there was any central guidance on the exploitation and application of benchmarks, NQE replied: 'There’s guidance in the sense that there is advice to do with writing learning outcomes, and benchmarks play a part in that:
so we do a development session with staff, taking them through, showing them the links.' Unfortunately, NQE did not respond to requests for access to associated documentation.

NQS confirmed after the interview that there was no institutional documentation on benchmarking, adding simply: 'but there will be.'

4.3 The interviews

After a rapid check of the respondent's length of experience in higher education and in the particular institution, the interview proper began. Occasionally, data occurring out of sequence with the interview schedule is re-located within the account of the findings. Each item from the interview schedule (Appendix 1) is presented in turn.

Given the small size of the sample, the findings are not amenable to analysis by the application of statistical techniques.

Experience of quality assurance and enhancement

The purpose of the first question in the interview schedule was to establish a context for the interview and to invite the respondents (n = 34) to begin the process of reflection. Typical probes asked the respondent to evaluate his or her experiences of quality assurance and enhancement and/or QAA Subject Review. Given the need to address without undue delay issues surrounding subject benchmarking, little time was invested in this item and no attempt was made to canvass attitudes to particular themes.
Generally speaking, the academics interviewed were at least used to the concepts of quality assurance and enhancement, even if their current involvement tended to be limited to student feedback forms and peer observation of teaching.

Some of the academics interviewed did not, or could not, give an evaluation of Subject Review, though almost all of those who expressed a view spoke of it as having both positive and negative aspects. A typical expression of this duality was that given by OC12:

'I think that the principle, initially the principle, I don't have a problem with; that some sort of overall quality assurance is a good thing. It certainly developed in such a way that it became seriously onerous, in terms of the amount of work that had to be done, and particularly the amount of paperwork that had to be generated.'

NH3, who was broadly in favour of quality assurance and enhancement, and who liked the idea of peer review felt that 'the administrative burden that [Subject Review] placed on people ... far outweighed any benefits.' Indeed, for NC3, Subject Review had been 'basically an enormous amount of resource for not very much'.

Approximately one quarter of the academics interviewed situated quality assurance and enhancement in an accountability context. Some did this explicitly: 'Well, clearly there is in the air, now, the notion of accountability for taxpayers' money, and therefore: there's a lot of money being spent, then one of our key activities will be monitored and vetted. I find it a little hard to disagree with that' (OH4). Some respondents indicated an accountability relationship implicitly: 'I think we owe our students a responsibility, which is to ensure that the programmes we deliver are of
good quality’ (NH2). By contrast, OC6, fully supporting the principle of public accountability commented on Subject Review: ‘I mean – I see the point, and if I believed it made the teaching better, I’d be enthusiastic.’ Moreover, a significant shortcoming of the system was identified by NHH, drawing on personal experience of a highly successful Subject Review:

it just seemed to take up an awful lot of our time, and at the end was more a case of ‘phew!’ rather than inspiring us to go on and make the changes which were suggested by the Review. It wasn’t supportive ... it wasn’t developmental; it felt like it was, you know, a, sort of, summative assessment, rather than being part of an ongoing formative process.

OC1 identified a troubling development over the last ten years, involving ‘almost an explosion’ in quality assessment requirements: ‘I think there has been a development of a situation where there’s a lack of trust. And I think this is political in origin. But, basically, there seems to have been a loss of trust, and ... it is not accepted that we are able to control ourselves, in a sense.’ Only one other academic (OH4) raised the question of trust at this stage, but interpreting the shift from Subject to Institutional Review as a positive development: ‘I suppose there’s an element of progress, isn’t there; that the QAA ... is now saying: “We’ve probably got you all up, to a certain degree, up to speed, and therefore we can move to a lighter touch, and so on, and just trust the universities.”’

Academic identity was also raised by just one other academic. OC11, whilst acknowledging that Subject Review stimulated reflection on the quality of courses and also had the beneficial effect of improving paper trails, commented that having to have those paper trails ‘does ... subtract from ... just being considered
professionals who can get on and do the job, in some ways.’

Representatives of institutional quality systems were asked at this stage to give a general assessment of the QAA Subject Review system. At OldU, OQE felt that ‘it was not a bad system, in that it’s a job which needs to be done.’ Asked why the job needed doing, OQE replied:

Because – you’re going to get a slightly biased view, here – because academics do not manage, and students need to come into institutions which have managed systems, and there’s a lot of work to be done in managing systems which focus on student achievement. And that’s where I think the QAA has got it right, they do talk in terms of student achievement.

For OQS, the strength of the system was that ‘it did actually look at teaching: they actually looked at delivery.’ However, ‘an obvious problem’ for OQS was the amount of documentation required and the time spent producing it: ‘it’s hard to really, I think, justify the amount of value added of doing that’.

For NQE, Subject Review had been helpful in identifying areas needing development and also good practice. However: ‘I suppose the negative sides are the pressure that it puts on staff ...; and also the emphasis was more on systems and processes than actually on the development side’ (NQE). According to NQS, Subject Review was ‘a very hit-and-miss system of academic review.’ On the positive side: ‘It got academics to concentrate, for example, on the student learning experience, on curriculum design, on methods of assessment – these type of things, which I think were entirely beneficial’ (NQS). On the other hand: ‘Where I found ... it fell down is that it became unduly bureaucratic, in becoming an actual paper trail exercise’ (NQS).
At this stage, then, a picture emerges, across both institutions and subjects, of acceptance of the principles of external quality assurance, combined with objections to the excessive administrative burden involved.

*Experience of working with criteria set by an external agency, such as a professional body*

This item was addressed only to academic staff (n = 30) and served a twofold purpose. On the one hand, it formed a mechanism for checking whether chemists and historians did actually differ in this regard. On the other hand, it encouraged respondents to reflect on possible direct involvement of external agencies in their teaching programmes, and thus helped to move them from the generalities of quality assurance and enhancement to the specifics of subject benchmarking.

As expected, both chemistry departments ran courses accredited by the Royal Society of Chemistry (RSC). Only two chemists (OC2, OC4) claimed to have no involvement with the RSC, though it was clear that, in some cases (eg. OC9) the direct effect on the individual was perceived as being minimal. In spite of the considerable effort required for RSC accreditation, chemists generally found it to be ‘a worthy aim to actually have a valid, highly regarded, professional qualification of CChem’ (OC12): indeed, for NC1, RSC accreditation was ‘one of those things that’s expected.’

In addition to the RSC, just one chemist (NC1) mentioned experience of the CNAA.

With only two exceptions, the historians interviewed identified no experience of external agencies: ‘We’re pretty unregulated’ (OH1). NH1 had some experience of
the CNAA and OHH, whilst having no engagement with an external agency for history, had done internal reviews in areas such as law, with a strong professional association requirement.

Reactions to the introduction of benchmarking

This item marked the start of the discussion specifically on benchmarking. Respondents (n = 34) were asked to rate the introduction of benchmarking on a simple Likert scale from ‘very good’ to ‘very bad’ (plus ‘other’): the breakdown of the responses is given in Table 4.2.

<table>
<thead>
<tr>
<th>Table 4.2 How do you feel about the introduction of benchmarking?</th>
</tr>
</thead>
<tbody>
<tr>
<td>It is:-</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>OldU</td>
</tr>
<tr>
<td>Chemistry</td>
</tr>
<tr>
<td>History</td>
</tr>
<tr>
<td>Quality</td>
</tr>
<tr>
<td>Total</td>
</tr>
<tr>
<td>NewU</td>
</tr>
<tr>
<td>Chemistry</td>
</tr>
<tr>
<td>History</td>
</tr>
<tr>
<td>Quality</td>
</tr>
<tr>
<td>Total</td>
</tr>
<tr>
<td>All</td>
</tr>
<tr>
<td>Chemistry</td>
</tr>
<tr>
<td>History</td>
</tr>
<tr>
<td>Quality</td>
</tr>
<tr>
<td>Total</td>
</tr>
</tbody>
</table>

It is striking that no respondent rated the introduction of subject benchmarking as overall ‘very bad’ or even ‘bad’, though one individual did remark: ‘Well, I can’t say it’s bad’ (OC3), and a further individual, having stated: ‘It’s a difficult one’
(NH3), observed: 'I don’t think it’s an entirely bad thing [laughing], and I think that’s as far as I’m going to go on that' (NH3). At the other extreme, only one respondent gave the rating ‘very good’, seeing the system as a mechanism for facilitating the recognition of degrees: ‘so, the benchmarking was basically saying: Right, now, anybody who claims to be a Bachelor in chemistry must understand this’ (OC6).

Overall, 18/34 respondents rated the introduction of benchmarking as ‘good’, and 11/34 as ‘neither good nor bad’.

At OldU, 13/21 were in favour of ‘good’, with 6/21 declaring ‘neither good nor bad’. In particular, 9/13 OldU chemists found the introduction ‘good’, and only 2/13 ‘neither good nor bad’. The one remaining OldU chemist felt unable to answer the question. OldU’s historians were equally split between ‘good’ (3/6) and ‘neither good nor bad’ (3/6), as were OldU’s quality administrators.

At NewU, a rather different distribution appears, with 5/13 finding the introduction of benchmarking ‘good’, and 5/13 ‘neither good nor bad’. Strikingly, only 1/6 chemist rated the introduction ‘good’, as opposed to 3/6 declaring ‘neither good nor bad’. 2/6 chemists interviewed were classified as ‘other’. ‘Lack of knowledge’ made it impossible for NC2 to make a definite evaluation: nevertheless: ‘it’s got to be an improvement … if there’s been nothing there before’ (NC2). NC3 provided a split evaluation: ‘I’ve got things for and against, I think.’ NH4 took a similar view: ‘I guess it’s a mixed feeling … it’s not “neither good nor bad”, it’s a bit of ‘good’ and ‘bad’ [laughs]’. Otherwise, NewU historians were equally divided between
'good' (2/5) and 'neither good nor bad' (2/5). Both quality officers at NewU rated the introduction of benchmarking as 'good'. If the equivocal responses from NC3 and NH4 are factored in divided between 'good' and 'bad', the NewU totals become: 'good', 6/13; 'bad', 1/13; and the overall totals: 'good', 19/34; 'bad', 1/34.

All four inexperienced chemists at OldU rated the item 'good', whereas at NewU, 1/2 inexperienced chemist was neutral and 1/2 was classed as 'other'. Both inexperienced historians at OldU were neutral, whereas the one inexperienced historian at NewU was split between 'good' and 'bad' and initially classed as 'other'.

Typical of the view taken by chemists was the reaction of OC10: 'I think that everyone would agree in chemistry [that everybody] needs to know the basics, attain a particular level, and it would be nice if that level is standardized across the country.' As NC1 put it: 'I'd rate it at the "good" level, because it's about trying to define what you think a graduate in the subject is.'

Amongst the historians, NH2 also thought that benchmarking 'should ensure that all of the courses are meeting certain standards.' A contrasting view was given by OH1: 'although ... I can see the advantage of ... some outside observation and some standardization, I think that the realities of placing lines which we have to meet, or boundaries over which we must hurdle, is, they're often quite false.' A particular bone of contention for OH1 was transferable skills: 'it's just a nonsense, you know, that we have to attach what we do in our courses to these skills which are integral, and ever have been, in history degrees'. Whereas OH5 and NH1

119
commented that the benchmarking process had given rise to beneficial reflection on the nature of the subject, OH2 objected: 'it's not exactly helped departments to reflect; it's merely provided a template that you follow.' Indeed, OHH rated benchmarking as 'good' on the grounds that it was useful to have a reference point in Subject Reviews.

Misgivings about subject benchmarking were typified by OH5, who, whilst recognising the usefulness of inter-departmental comparability in terms of what a history degree is 'meant to look like' (OH5), commented: 'My whole fear with anything, when you write down rules, is it can also become restrictive, in their turn.' Such unease about a possible 'lack of flexibility' (OH5) was expressed in a number of ways. NCH, for instance, took as a given that there was 'a core of chemistry', but also wished to preserve particular staff specialisms, which might be endangered 'if the benchmarking became too prescriptive all the way through'. Similarly, NHH stated of his reaction to reading the benchmarking statement: 'yes, this affirms that we're doing some things which have some kind of recognition, so that's a useful idea'; adding, however: 'But, then again, I thought, there are distinctive things about the degree that I teach on that I wouldn't want to be regarded as illegitimate.' For OHH it was important that the benchmarks should be guidance and not prescription. At OldU, both quality officers found the benchmarking statements unhelpful in their formats, as they were constructed differently, using different criteria. By contrast, NQE found benchmarks good, 'because it's giving us a set of markers, if you like, that we can say: if we're meeting these, we can put ourselves in comparison with other institutions.' This
depended, however, on there being ‘flexibility in their use’ (NQE): ‘I think they would only be considered bad if they are taken as a straitjacket’ (NQE). For NC4, too, an evaluation of benchmarking ‘probably depends on how people apply it.’ The major concern here was that benchmarking was ‘getting close to a national curriculum for higher education’, which was interpreted as a bad thing. In sharp contrast to this, NQS stated: ‘I don’t think there are any dangers in subject benchmarking.’ Benchmarks were good because, ‘I think that you need consistency across higher education in what’s being taught; not how it’s being taught, but what’s being taught’ (NQS). Such an approach, NQS suggested, corresponded to there being ‘accepted paradigms within subjects’. Asked about the description of this being ‘almost like in a national curriculum’ (NQS), NQS replied: ‘I would say that it’s a good thing.’

One academic answered in terms of academic identity, finding it hard to reconcile the attempt to maintain standards with ‘what I was saying before about trust’ (OC1).

Reactions to the performance criteria and statement of threshold performance (chemistry) or the statement of the threshold standard (history) established for single honours degrees in chemistry and history.

The same Likert scale was used here as for the previous item: the breakdown of the responses (n = 34) is given in Table 4.3.

No respondent rated this item as ‘very good’, though NC5 did describe it as ‘good to very good’. Whilst no rating of ‘very bad’ was recorded, 5/34 respondents, all chemists (four of them from OldU), did consider this item to be ‘bad’. NCH
objected: ‘we found it difficult to interpret at what level some of these benchmarks should apply.’ Similarly, OC6 observed: ‘The attainment thing I find confusing. It isn’t clear to me whether what it’s trying to do is define First Class, II, III, Third Class degrees’. For the others, the problem lay with the formulation of the document, descriptions being used such as ‘wishy-washy’ (OC4) or ‘vague’ (OC5). More fundamental was OC9’s criticism: ‘I’m very worried by the general idea that you can sort of apply any kind of quantitative scale to something that isn’t really quantifiable, which seems to me to be what this is all about. So I would say I don’t like that idea.’

Table 4.3 How do you react to the performance criteria and/or threshold standard?

<table>
<thead>
<tr>
<th>It is:-</th>
<th>Very good</th>
<th>Good</th>
<th>Neither good nor bad</th>
<th>Bad</th>
<th>Very bad</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Nr</td>
<td>%</td>
<td>Nr</td>
<td>%</td>
<td>Nr</td>
<td>%</td>
</tr>
<tr>
<td>OldU</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chemistry</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>46.15</td>
<td>2</td>
<td>15.38</td>
</tr>
<tr>
<td>History</td>
<td>0</td>
<td>0</td>
<td>46.67</td>
<td>2</td>
<td>33.33</td>
<td>0</td>
</tr>
<tr>
<td>Quality</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>50</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>47.62</td>
<td>5</td>
<td>23.8</td>
</tr>
<tr>
<td>NewU</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chemistry</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>66.67</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>History</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>40</td>
<td>2</td>
<td>40</td>
</tr>
<tr>
<td>Quality</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>50</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>46.15</td>
<td>3</td>
<td>23.08</td>
</tr>
<tr>
<td>All</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chemistry</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>52.63</td>
<td>2</td>
<td>10.53</td>
</tr>
<tr>
<td>History</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>54.55</td>
<td>4</td>
<td>36.36</td>
</tr>
<tr>
<td>Quality</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>50</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>0</td>
<td>0</td>
<td>16</td>
<td>47.06</td>
<td>8</td>
<td>23.53</td>
</tr>
</tbody>
</table>

OldU: n = 21  NewU: n = 13  All: n = 34
OldU: 13C + 6H + 2Q  NewU: 6C + 5H + 2Q  All: 19C + 11H + 4Q

For this question, 5/34 respondents are classified as ‘other’. OC1 and NHH felt unable to answer the question, and NQE observed: ‘Threshold is a difficult one,
because we don’t work with thresholds, and I think that’s caused a problem.’ NC1, in spite of gentle probing, was very difficult to pin down, but did offer the view that a threshold level was ‘a simplistic measure’. OQS said of having threshold standards: ‘I think, really, just as an idea it’s good’, but was ‘disappointed’ with the resulting documents overall: ‘So, I would say bad’ (OQS).

Overall, 16/34 respondents rated this item ‘good’ and 8/34 ‘neither good nor bad’.

At both OldU (10/21:5/21) and NewU (6/13:3/13), twice as many gave a positive assessment as did a neutral one, though comments were muted, being of the order ‘fine’ (OC3), ‘a worthy effort’ (OC12), ‘sensible’ (OH1), ‘a fair enough statement’ (NC2), ‘a useable set of criteria’ (NH2).

In chemistry, the proportion of responses in the category ‘good’ was rather lower at OldU (6/13) than at NewU (4/6), and the proportion in the category ‘bad’ rather higher (4/13 and 1/6 respectively).

In history, the distribution between ‘good’ and neutral was very similar in the two institutions, the small numbers involved exaggerating the percentage difference. It is noticeable, however, that no historian rated this item ‘bad’.

7/10 inexperienced respondents rated the item ‘good’, 2/10 ‘neither good nor bad’, and 1/10 ‘bad’.

At both OldU and NewU 1/2 quality administrators was classified as neutral and 1/2 as ‘other’.
Benchmarking and threshold standards as evaluated by quality administrators

Quality administrators (n = 4) were additionally invited to give a general assessment of benchmarking and threshold standards to date.

At OldU, OQE found the items to be useful and constructive and felt 'Broadly positive' about the system so far. OQS suggested that benchmarking and threshold standards were probably 'a reasonably helpful tool' in curricular review, and assumed that the discussions prior to the production of the benchmark statements had helped people to think about their disciplines and what they were trying to achieve in them. On the negative side: 'Well, an obvious thought is that – again, purely just looking at it from one institution: not much evidence it's actually led to curriculum change' (OQS).

At NewU, NQE stated: 'They are starting to work, and I think staff are starting to engage with them.' NQS saw it as 'a positive achievement' that in certain subjects 'they basically have been not too prescriptive. ... they're almost generic. And therefore within that, it's not too difficult to actually say: "Well, we are doing this."' However, NQS suspected that one of the disadvantages of benchmark statements was that 'they tend to be drawn up by people who you might say are within the inner circle of a particular discipline, and that reflects perhaps a particular normative set of values at that point of time.' Overall, however, benchmarking and threshold standards 'seem to become an accepted part of the higher education framework' (NQS).
Benchmarking and change

Academics were asked if benchmarking had led them to change their approach to their courses. For quality officers, the reference was to ‘changes to the university’s approach to quality’. These two sets of results (n = 34) are combined to produce one set of aggregated results linking subject benchmarking and change.

Overwhelmingly, this question prompted a negative response, with 24/34 respondents, including 6 of the 10 inexperienced staff, answering ‘no’. In this group is included NC3, who replied firmly ‘no’ but did also add: ‘What it has changed is the course materials, in terms of what we issue in the way of a syllabus and a module proforma, and so forth; because we now include these benchmarking things within that.’ At the same time, however, NC3’s general comment was typical of this category: ‘We were doing it before, and we’re still doing it now that they’ve articulated these various things.’ OC6’s interpretation was uninhibited: ‘all that the benchmarking contains, to somebody like me, to be blunt, is blindingly obvious.’

For OQE, the answer was clearly ‘no’; OldU had clearly done things in response to the overall changes which included benchmark statements: ‘But, in itself, there are things that are more significant in terms of changing systems.’ NQE replied ‘no’, but added that NewU was very aware of quality issues: ‘enhancement, management, assurance, the quality and standards agenda; and benchmarks play a part in that agenda. So we had the agenda in place already. I don’t think benchmarks themselves have done anything to shift that; they’re just part of one total operation.’

A small number (5/34), all but one (NH2) inexperienced members of staff, fell into
the category 'other', essentially because they felt unable to answer the question: NH2 did not provide an unequivocal answer.

The category 'yes' also provided 5 responses. OHi4, who originally answered 'no', recalled at the end of the interview that the benchmarking document had given a lead on how to 'revamp' grade descriptors. NC1, adopting a perspective opposite to that already recorded for NC3, answered 'yes' because benchmarking 'affected how we presented our modules and our courses externally', but also made it clear that there had been no other changes: 'Because I think, on the whole, we discovered that we did comply' (NC1). For NH1, the only real changes had occurred in the area of assessment: 'to make sure that the assessment accurately reflects the content of the modules, particularly the learning outcomes et cetera that we’d been asked to indicate.'

OQS explained that, whilst there had been fairly significant changes at OldU as a result of the various QAA developments, the existence of benchmark statements had only been one aspect of that: 'In themselves, they’ve not led to major changes. And, as I say, they’re changes to our quality systems, not the quality of provision'. However, the only example given of such a change was the requirement to produce Programme Specifications, a point also made by OC3.

Initially, NQS commented: 'It is one of the actual criteria along with the QAA Code of Practice and the HEFCE (sic) Academic Framework, where we need to align our quality particular procedures with those of outside agencies, of which benchmarking is one. So it has.' However, NQS went on to explain that
benchmarks 'were there, they were basically in everything this university’s been doing; but they need now to articulate it very explicitly; exactly how they are meeting these benchmarks.'

No significant differences could be discerned between institutions or subjects. Moreover, any change mentioned was slight.

**Academics' perceptions of the impact of benchmarking on their professional lives**

This item, addressed only to academic staff (n = 30), sought to assess the perceived direct impact of benchmarking on academics, but also to prepare the ground for questions 11, 12 and 13 in the interview schedule.

For 15/30 respondents, including 5/9 inexperienced academics, there had been no impact. An unequivocal reaction was that of OC6: 'I can’t imagine how ... . I would guess that the benchmarking document has had zero effect on this department.' NC4 also replied 'no', but thought there was a potential for impact, because 'I am all the time worried about a national curriculum for higher education.' Included here is NC3, who replied: 'Not really, apart from the extra work in producing the documentation'.

Paperwork and administration formed the major reasons for answering 'yes' to this question (11/30): OC3, for instance, referred to 'this deluge of stuff'. NH1 observed that, as a result of rewriting module specifications, benchmarking 'became part of my professional life'; and OH2 stated: 'it's helped us, or it's made us, focus the learning objectives and outcomes more clearly around the things that are in the
Benchmarking Statement. For OC4, the only inexperienced academic to reply 'yes', there was more documentation to go through: 'So you do find yourself trying to jump through these little hoops that are set, just to get the documentation in order, which, in a way, detracts from the teaching and detracts from other aspects of my job, as well. So, there are some difficulties there.' NH2 took a very different line, referring to a major responsibility held for quality, but also observing: 'I think, personally: yes, it's had quite an impact on how I see myself and how I perform as an academic in a history department.' OC1's view was rather negative: 'we have to do a lot more in terms of administration to justify your courses. So, really, what's happening is there's a lot of paperwork to be originated, most of which I wonder if, how many people read it.'

4/30 respondents fell into the category 'other'. OH3 and NC2 both suspected an increase in paperwork. OHH gave an unfocussed answer, but commented: 'I think we are conscious of the benchmarking if we're designing new courses.' NH4, completely unfamiliar with subject benchmarking, offered a different perspective: 'My feeling is that, by creating a kind of, a standard approach, that it can only benefit institutions like ours, which maybe don't have a traditional reputation, being a new university'; this enhanced reputation then 'reflects on me, doesn't it?' (NH4).

As with the previous item, no significant differences could be discerned between institutions or subjects.

Benchmarking and improvements in quality

The purpose of this item was to seek insights into the meaning attached to
benchmarking by members of the HE community (n = 34). Whilst it is true that enhancement is not a stated aim of subject benchmarking, it is, as the literature review has shown, a key element of most definitions of benchmarking. Tight time constraints precluded a discussion of possible interpretations of the concept of quality: rather, an opportunity was created for respondents either to question the idea of quality, or, and possibly also, to indicate in their answers what they considered quality to be. Only 5/34 respondents asked for any kind of clarification of the term ‘quality’, the interviewer then referring typically to the quality of Single Honours degrees. Additionally, one respondent observed: ‘I’m very unhappy with the word “quality”, anyway, I think’ (OC9).

Approximately half of the respondents indicated some interpretation of quality, most of them in terms of teaching: ‘we’re talking teaching’ (OH4). Occasionally, something else was also mentioned, such as ‘content or teaching methods’ (OC9) or ‘the delivery of historical knowledge’ (OH1). For OC8: ‘Every institution is under pressure to show that they’re excellent at everything’. In one instance, a respondent answered in terms of standards: ‘I would say no. I’d say the best it could hope for is to maintain levels’ (OC2). Answering the question as to what would lead to quality improvements, two respondents appeared to use ‘quality’ and ‘standards’ interchangeably. For OH4: ‘the challenge is to ... have larger numbers, yet not ... deterioration of standards.’ In OC10’s view: ‘the way to increase standards is not to teach to the lowest common denominator’.

For 15/34 respondents, benchmarking was not leading to quality improvements. In
OC9’s view: ‘the things that actually get altered are very much exterior gloss rather than genuine content or teaching methods.’ As NC1 put it: ‘with a lot of people, benchmarks are something that they just tick off.’ OC13 commented: ‘I think it will lead to people gearing their teaching and things to fit in with [the benchmarking statement]. I don’t think it will improve things, no.’ For NH3: ‘I think potentially imposing benchmarks could restrict diversity.’

A further 14/34 respondents fell into the category ‘other’, generally because they did not know. NCH thought that it was still too early to say, but NH2 observed: ‘How would we know if benchmarking had improved the quality of what we were doing? It would be very difficult to isolate it out from all the other factors.’ NQE thought: ‘it may inform development. I probably wouldn’t use the word “improvement”.’ For NQS, the greater conformity in curriculum construction brought about by benchmarking ‘will lead to a fuller and more meaningful curriculum for students; that therefore, by default, leads to greater quality’, whereas ‘if you’re looking at, let us say, methods of teaching and learning, then it may lead to less experimentation, and as a consequence of that, that may reduce the actual quality’ (NQS).

Only 5/34 respondents offered more positive answers, though three of these were very muted: ‘I guess, in a way, it may do’ (NH4); ‘I would hope that it is’ (NC5); ‘It may be doing’ (OH5). OH5 further commented that benchmarking’s impact in this regard ‘may be difficult for us to sort out: I don’t think it will be on a great scale’; adding: ‘I think it probably is keeping up this pressure to be self-aware’. The
remaining two respondents, both from NewU History, were the only people actually to answer 'yes'. NH2 highlighted 'the transparency of information between academics and students' and a rethinking of modules, programmes and assessment strategies, whilst nevertheless recognising, 'The alternative argument ... that it's a paper exercise, and that you're ticking boxes in order to meet the demands of an outside agency, and then the internal agencies that respond to that. I think people have responded in different ways.' For NH1:

Indirectly, yes; because I think benchmarking is part of a whole process of what you might call a regulation culture: and whereas that can be rather stifling and bureaucratic and rigid, it does encourage people to reflect on what they do, and I think that sort of reflection is healthy. They will grumble about it and will often fill the forms in just wanting to get it off their back, but I think underneath they do reflect on what they do. And I think that's helpful.

If the muted responses of OH5, NC5 and NH4 are classed as 'other', the totals for OldU are: 'yes', 0/21; 'no', 9/21; 'other', 12/21; and for NewU: 'yes', 2/13; 'no', 6/13; 'other', 5/13; inexperienced respondents: 'no' 5/10; 'other' 5/10.

**Benchmarking and the dissemination of best practice**

As the dissemination of best practice is one of the stated aims of the QAA external review process (QAA, 2002c: para.61), an integral part of which is subject benchmarking, this item was addressed to all 34 respondents: the breakdown of the responses is given in Table 4.4.

As can be seen, the overall responses fell into three almost equally sized groups for 'yes', 'no' and 'other'. The range of responses went from: 'It has to help' (NC4), through: 'Don't think it can discourage it: let's put it that way' (OC1), to: 'I don't
see how it relates to that at all' (OC3). OH4 observed astutely: "Well, obviously, not in itself, because it just sits there, doesn’t it: so, somebody has to act on it.' Benchmarking was ‘a starting point, something you could refer back to’ (OH4). As NH2 put it: ‘I think, on its own, it may not have got beyond an initial discussion and a filing in a relevant place.’ Moreover, for OC4: ‘I’m not sure how many lecturers are going to actually read the data that comes from the benchmarking. There’s a bit of an inhibition, I guess, to actually read all this information when you’ve got lots of other things to do as well.’ A rather different view was taken by OC9: ‘I’m a little unconvinced that there is such a thing as best practice.’

| Table 4.4 Do you think that benchmarking will encourage the dissemination of best practice? |
|---------------------------------|---|---|---|
|                                | Yes | No | Other |
|                                | Nr  | %  | Nr  | %  | Nr  | %  |
| **OldU**                       |     |    |     |    |     |    |
| Chemistry                      | 4   | 30.77 | 4 | 30.77 | 5 | 38.46 |
| History                        | 2   | 33.33 | 2 | 33.33 | 2 | 33.33 |
| Quality                        | 1   | 50 | 1 | 50 | 0 | 0 |
| **Total**                      | 7   | 33.33 | 7 | 33.33 | 7 | 33.33 |
| **NewU**                       |     |    |     |    |     |    |
| Chemistry                      | 3   | 50 | 1 | 16.67 | 2 | 33.33 |
| History                        | 2   | 40 | 2 | 40 | 1 | 20 |
| Quality                        | 0   | 0 | 1 | 50 | 1 | 50 |
| **Total**                      | 5   | 38.46 | 4 | 30.77 | 4 | 30.77 |
| **All**                        |     |    |     |    |     |    |
| Chemistry                      | 7   | 36.84 | 5 | 26.32 | 7 | 36.84 |
| History                        | 4   | 36.36 | 4 | 36.36 | 3 | 27.27 |
| Quality                        | 1   | 25 | 2 | 50 | 1 | 25 |
| **Total**                      | 12  | 35.29 | 11 | 32.35 | 11 | 32.35 |

OldU: n = 21 NewU: n = 13 All: n = 34
OldU: 13C + 6H + 2Q NewU: 6C + 5H + 2Q All: 19C + 11H + 4Q

Those answering ‘yes’ did not do so, generally, with any strong endorsement of this aspect of benchmarking. Those who made explicit what they meant by ‘best practice’ mostly referred to teaching. Some thought that the dissemination would be
achieved by the published benchmarking statements themselves: ‘Well, I think, when you look at the sections in the benchmark statement which has got, for example: “alternative methods of assessment”, it might lead you to think of assessing work in more ways than one currently does’ (OC7). In NH2’s view: ‘I think the [subject benchmarking] document itself is part of the impetus. ... I think as part of a wider process of review, and as part of a response at university, faculty and department level, it’s contributed to developments.’

By contrast, among those answering ‘no’ to this question, OH2 stated: ‘I don’t think it was ever intended to do that job.’ OC6 stated: ‘I don’t see how, because I think it’s so lacking in detail.’ In OC12’s view: ‘I think there have been, over the last few years, significantly improved mechanisms for the dissemination of good practice, but not necessarily tied to the benchmarking exercise.’ For the dissemination of best practice, OC12 referred to the RSC and to the LTSN, the latter being cited by several respondents, as were conferences. Only one academic mentioned Subject Review in this context, but commented: ‘Well, this is one problem, I think, about best practice and the dissemination of it through the reports on departments. ... I don’t think that, as academics, we sit and read [them] – even in our own discipline, even in departments round about here’ (OH5).

A very forthright view was given by NC1: ‘Well, I mean, to be brutally honest, Steve: benchmarks are about government watchdogging and not about academic standards. Aren’t they? Let’s be brutally frank! This is a political game, not an academic game.’ This political aspect, one of the main centres of interest of the
study, was mentioned by no other respondent at this stage: standards were
mentioned by just two others, both categorised as 'other'.

In the category 'other', respondents tended to be unsure of the contribution of
benchmarking or very guarded, offering comments such as 'I would hope so' (OH1)
or 'it could help, like the Subject Review, in making people aware ... of various
standards and requirements' (OHH). As NQS put it: 'I don’t necessarily think that
it's going to have too much of an impact.' An interesting element of doubt was
introduced by OC10: 'It'll certainly try and homogenize the system. Whether that is
best practice, I think remains to be seen.' OHH also suggested that benchmarking
might help with 'the general problem ... of feeling comparability of degrees across
the country.' OC10 cautioned that, whilst homogenization would be a good thing if
it meant bringing everyone up to the standards of the country’s top chemistry
departments, 'If it means that, for the sake of having a general standard, that
standard is lowered to the average, then that is a bad thing.' OC8, struck by the
question: 'That’s a good one', commented: 'if we want to be rated as a good
institution ... and if this [Subject Benchmarking Statement] is the mechanism by
which we’re being assessed – then we’re going to make sure we fulfil all of these
criteria. Now, whether that makes it better or worse, I don’t know.'

The overall distribution of responses was generally even across institutions, subjects
and length of experience. Only NewU Chemistry showed a slightly different
pattern, with 1/6 'no' and 3/6 'yes'. The 'yes' responses were muted, however: 'I
would hope so; I would think it would' (NC2); 'To some extent, I suppose' (NC3);
and even NC4's 'It has to help' being explained as: 'QAA monitors, showing there's a proper link to External Examiners.'

The perceived necessity of subject benchmarking

This item began the process of investigating perceptions of the reasons for introducing subject benchmarking: the breakdown of the responses (n = 34) is given in Table 4.5.

| Table 4.5 Do you think that it was necessary to introduce subject benchmarking? |
|--------------------------------|----------------|----------------|----------------|
|                                 | Yes            | No             | Other          |
|                                 | Nr | %    | Nr | %    | Nr | %    |
| **OldU**                       |    |      |    |      |    |      |
| Chemistry                      | 5  | 38.46| 4  | 30.77| 4  | 30.77|
| History                        | 2  | 33.33| 4  | 66.67| 0  | 0    |
| Quality                        | 1  | 50   | 1  | 50   | 0  | 0    |
| **Total**                      | 8  | 38.1 | 9  | 42.86| 4  | 19.05|
| **NewU**                       |    |      |    |      |    |      |
| Chemistry                      | 2  | 33.33| 3  | 50   | 1  | 16.67|
| History                        | 2  | 40   | 2  | 40   | 1  | 20   |
| Quality                        | 0  | 0    | 0  | 0    | 2  | 100  |
| **Total**                      | 4  | 30.77| 5  | 38.46| 4  | 30.77|
| **All**                        |    |      |    |      |    |      |
| Chemistry                      | 7  | 36.84| 7  | 36.84| 5  | 26.32|
| History                        | 4  | 36.36| 6  | 54.55| 1  | 9.09 |
| Quality                        | 1  | 25   | 1  | 25   | 2  | 50   |
| **Total**                      | 12 | 35.29| 14 | 41.18| 8  | 23.53|

OldU: n = 21 NewU: n = 13 All: n = 34
OldU: 13C + 6H + 2Q NewU: 6C + 5H + 2Q All: 19C + 11H + 4Q

Overall, 12/34 answered 'yes' and 14/34 'no', with 8/34 responses categorized as 'other'.

OC3 found this an 'interesting question' and thought that benchmarking probably was necessary because of 'the great variety of institutions'. Similarly, NH4 saw the advantage of achieving 'evenness between institutions'; or, as OC5 put it: 'So that
the – I think it’s a minority, but I think the weak and poor minority can’t get away with it.’ Several respondents took the sort of view expressed by OH3: ‘If this process means that there is a minimum standard and that there is a set of aims that we’re going about to achieve, then that’s no bad thing, as far as I’m concerned.’

For OH4: ‘if it’s deemed necessary to have a system of Subject Review, at departmental level, I think you had to have the benchmarking.’ By contrast, OHH, who replied ‘no’, explained: ‘I think the Subject Review was doing most of the things quite adequately, as it was.’

OC13 gave a very different reason for replying ‘yes’: ‘I think, in the climate we exist in now, yes; because everything has benchmarking of some sort.’ OC13 further explained that benchmarking ‘just came out of a general need – not just in education, but in all areas – to try and justify funding and suchlike.’

OQE was the only administrator to respond ‘yes’, doing so because benchmarks ‘provide a starting point for working with departments’. However, pressed slightly as to whether benchmarking was necessary in order to put right things that were defective in the system, OQE replied: ‘in terms of things being wrong, no.’

In the ‘no’ category, OC7 ‘hadn’t perceived any problem that [benchmarking] would be an answer to’, and OC9 observed: ‘Was there some perception that things were going wrong? Because, if so, that wasn’t clear to me.’ NC4 thought that benchmarking was ‘not entirely necessary, but could be useful.’ Similarly, NC1 responded ‘no’ ‘because of the impact of the professional body’, but nevertheless
qualified the response because 'the need for subject benchmarks was about showing to the wider community what was a degree in chemistry.'

OQS also responded 'no', pointing out: 'really, the position afterwards doesn't look very different from the position beforehand. It's difficult to imagine why one would have felt it was necessary, given that it doesn't seem to have had much impact.' A similar retrospective assessment was given by OH5 and NC3.

In the category 'other', NHH thought that benchmarking was not necessary, though possibly desirable in terms of guidance when setting up a new course. For OC2, 'it was probably needed, but I wouldn't say necessary. ... it may help to maintain standards.' OC6 found the term 'necessary' 'over-strong', but thought that benchmarking had been 'an interesting exercise'. Two respondents (OC8, OC10) were unable to answer. In the opinion of NQE, the introduction of benchmarking was 'inevitable'. Two respondents (NC2, NQS) gave mixed responses. For NC2:

I guess I've got two opinions. One: the higher education institutions seemed to work well for god knows how many years before it was introduced. Now, you could take the view that, you know, why fix something if it's not broken? On the other side of the coin is that, obviously, the profile of higher education has changed over the years. ... we are educating a mass, rather than a minority.

OC4, who replied 'yes', made a related point, but referred to the declining standard of students at intake, an issue also mentioned elsewhere by other respondents. The response of NQS was also two-fold: 'In terms of what universities were wishing to do, it probably wasn't necessary, in that sense. ... I think it makes a lot of sense in terms of ensuring that, within single disciplines, that they actually are up to (sic) the ball.'
Differences between OldU and NewU were slight. Whilst overall responses in chemistry were fairly evenly distributed across the three categories, a small majority (6/11) of historians responded 'no'. Amongst the inexperienced staff, 5/10 answered 'yes', though without strong conviction; 2/10 gave mixed answers; 2/10 did not know; 1/10 answered 'no'.

**Perceptions of why benchmarking was introduced**

This item made it possible to give some control to the respondents (n = 34), even though the time available did not allow for detailed probing or discussion of the meanings attached to responses. The most frequently given explanations for the introduction of benchmarking are shown in Table 4.6: some respondents suggested more than one category, and the distinctions drawn were at times less distinct than is suggested by a tabulation.

<table>
<thead>
<tr>
<th>Table 4.6 Perceived reasons for the introduction of subject benchmarking</th>
</tr>
</thead>
<tbody>
<tr>
<td>Political reasons</td>
</tr>
<tr>
<td>Considerations of standards and/or quality</td>
</tr>
<tr>
<td>Reflects a generalized phenomenon</td>
</tr>
<tr>
<td>Considerations of accountability</td>
</tr>
<tr>
<td>Quantification and/or measurement</td>
</tr>
<tr>
<td>Trust and/or professionalism</td>
</tr>
</tbody>
</table>

Figures indicate the number of times an item was mentioned

14/34 respondents suggested a political explanation for the introduction of subject benchmarking, sometimes directly: 'This is quite political' (OC3); in one instance, hesitantly: ‘That’s very difficult. I’m not wholly sure. I’m not sure if it’s political.’ (OC4). For OC6, there was ‘a fear ... that universities were not doing their jobs properly’, that fear emanating from ‘[w]hat used to be called the Department of Education. I mean, our political masters’ (OC6). In OH5’s view: ‘I suspected that it
was driven by the Treasury wanting academics to justify their existence, again.’ NCI suggested: ‘I think there was a perception somewhere, in the Civil Service, or wherever, that universities weren’t doing their job.’ Otherwise, the political source of benchmarking, if specified, was identified as ‘the government’:

I assume it’s governmental in essence ... if, as a government, you are looking to influence the direction of higher education and what it achieves, part of that process has to be to be able to have some definition of what universities will be delivering, and that benchmark statements is a step in that direction.

(OQS)

A note of scepticism was introduced by OH2, though in answer to the preceding item: ‘I think it was done for political reasons, which could be explained – but probably won’t be – by the QAA.’

12/34 respondents mentioned standards or quality. 8/34 respondents, including NCI in an earlier answer, alluded to standards. Some did so explicitly: ‘That [the general public] can feel, if they’re choosing an institution, that broadly similar and comparable standards are applying in each institution’ (NHI). Others addressed standards indirectly: ‘we’ve got to make sure that people claiming to teach Bachelor’s degrees in all subjects are actually doing the bottom line’ (OC6). 5/34 respondents responded in terms of quality, though the distinction between standards and quality was far from clear, as when OC8 reflected: ‘I don’t necessarily think that [benchmarking] leads to higher perceived quality at the end.’ The extreme case was NQE, who is consequently counted twice: ‘Definitely for the quality agenda. We’re looking at standards, here.’

11/34 respondents answered that benchmarking was an example of a generalised
phenomenon: 'I think it just comes from the way that everything is now being run in this country. I don't think it's anything to do with higher education' (OC9). For OC7: 'Well, it's all part of this general quality assurance mind set we have now, isn't it? In that you have to demonstrate that you are producing some course which is meeting its objectives, and so on; whereas it was all much more sort of laisserz-faire before.' For OH1, benchmarking was part of the 'creeping bureaucratisation' of the HE system; and NH1 observed: 'we live ... in a culture of regulation'.

10/34 respondents gave answers referring to accountability in some form, generally university accountability: 'Now, whether benchmarking fits into that whole process, where that greater accountability and greater transparency is applied to this [benchmarking]; well - I can say that's probably down to a personal opinion - but, it certainly looks like it' (OH3). Respondents typically referred to public funding: 'I think it's government obsession with accountability, and that value for money has to be seen to be delivered: and I can understand that' (NH3). Two of the respondents in this category referred to governmental accountability, either because of the government's responsibility for teaching (OC5) or because of the major expenditure on HE (OH1). OHH possibly straddled these two groups, seeing benchmarking as part of a public accountability requirement to demonstrate efficiency gains in the face of a declining unit of resource.

9/34 respondents answered in terms of quantification or measurement, OC11 replying simply: 'Because we're in a culture that, everything must be measured', the measurement being done by the government. For OC2: 'I think there's a general
drive to try and quantify degree programmes, and see what people are actually learning'. This was seen by OC2 as 'probably good'; but NH3, who declared: 'I have no problem with measurement per se', introduced a note of caution, speaking of 'the situation with things like on-going measurement that is going to get more and more relentless and more and more intrusive, which is a worry.' NQE’s 'slightly cynical' view was that benchmarking was 'one way of starting to compare', and two respondents referred specifically to league tables.

Only 5/34 respondents raised issues of trust or professionalism at this stage; though, in effect, all five alluded to both issues. OC2 replied: 'My glib answer is because the government doesn’t trust academics.' For NC3, quantification reflected: 'a central need in professional bodies and the government and everywhere else ... They're not satisfied with people saying that “we’re doing it”: they want to know exactly how and why they’re doing it, and when they’re doing it.' OC3 summarised: 'Highly professional people are cheesed off because they are not trusted. Job satisfaction and motivation are going down.'

3/34 respondents gave atypical answers. For OHH: 'I think my honest answer would be: yet another stick to hit these lazy academics, who are all skiving.' OQE answered: 'Why was it done? Because, if you’re going to look at dissemination of best practice, you have to have a marker of what best practice is.' OH4 reflected on external review mechanisms, but did not give a direct answer to this item. 4/34 respondents offered no view.
Benchmarking and changes in perceptions of the position of academics

This item was addressed only to the academics in the sample (n = 30) and produced an overwhelming response of 'no' (22/30), generally with little or no further comment.

2/30 respondents were categorised as 'other'. For OC9: 'If I found I was in a situation where my teaching was being completely prescribed, in some way, then I would not wish to do it any more.' NC4 replied: 'Far more time ought to be invested in scholarship, not in regular quality assurance and the taking up of time.'

Only 6/30 respondents answered 'yes': three from OldU Chemistry, all relatively inexperienced; and three from NewU History, one of them inexperienced. In OC2's view: 'I think we're less free to do what we want'; and for OC8: 'I feel a lot more pressure now to make sure that everything I do is in line with all my colleagues in this institution and also my peers at other institutions.' Whereas neither OC2 nor OC8 could say whether the position they described was good or bad, NH4, who replied: 'I suppose it makes me feel less autonomous', indicated that benchmarking implied a negative impact, in terms of some loss of control. OC10 replied in terms of bureaucratic demands regarding teaching: 'I think most people came into the job because they want to do research and they want to enthuse students, and the more that you're put under the spotlight and the more admin that you get, the less time you actually have to do that.' Of the two experienced academics to reply 'yes', NH1 stated: 'Not on its own. It's part of a number of developments in recent years that have encouraged me to reflect on what being an academic means.' For NH2: 'I
guess that the benchmarking documents has been a significant contributory factor in
that, yes'; though NH2 then proceeded to talk about other contributory factors, such
as the expansion of HE and 'just simply the maturing process that one goes
through'.

Overall, no obvious differences emerged between institutions. The generally muted
'yes' responses rule out any meaningful contrast of chemistry and history, though it
is interesting that four of the six 'yes' responses came from relatively inexperienced
academics.

The four administrators in the sample were asked if they thought that benchmarking
was leading to changes in the nature of the academic profession. No clear pattern
emerged. Only NQE was categorised as 'no': 'I wouldn't feel that it alone had that
power of impact.' OQS, the one respondent in this group classified as 'other',
replied: 'It has that potential. It's the first, sort of, toe in the water, I suppose, of
saying to academics: “We want you to teach this”, rather than simply saying:
“Here’s a general area; go away and teach what you want to in that area.”'

OQE and NQS both answered 'yes'. OQE saw benchmarking as 'part of a number
of things which are leading to changes in the academic profession', these including
'a move towards conception of programmes as a whole, rather than individual
lecturers coming in, doing their modules in isolation', academics' responsibility for
student support, and the call for institutional transparency. For NQS, it had become
more difficult in some disciplines to become a radical academic, 'because that
therefore takes you outside that mainstream benchmarking.'
Benchmarking and changes in perceptions of the position of universities in today's society

This item was addressed to all members of the sample (n = 34) and, as with the previous question, produced a clearly predominating response of 'no' (23/34), again generally with little or no further comment. OCI observed: 'Well, the position of universities in today's society has changed enormously from when I was a student. ... I think the major changes have not been due to benchmarking, but have been due to, again, a political desire.' This was the only explicit reference here to a politically motivated intervention into the development of universities, though OC2 commented: 'I think it's, you know, outside influence having more control over the university. I think the university used to be more autonomous'. Probing revealed that this outside influence was coming from, 'I guess government, ultimately' (OC2). However, as an inexperienced member of staff, OC2 was unable to say that subject benchmarking had provoked a rethinking of the position of universities in society: 'it's difficult for me to comment. ... I'm sure people had far more freedom to do what they wanted than I do now, coming into the profession.' NQS found benchmarking 'really insignificant'; and in terms of a possible shift in the perception of what the relation is between universities and society, OQS stated: 'I think the overall impact of the benchmark statements have (sic) been too small to have led to any such [shift].' For OQE: 'Benchmark statements, of themselves, haven't. But, the whole debate that they're part of has.'

6/34 respondents answered 'yes', though the line dividing 'yes' from 'no' was often fine. As OHH put it: 'one obvious answer would be "yes"'; but I think the more
realistic answer of somebody of my age ... is to say that, really, it's just another element of a rethink, you know, over the years.' For NH4: 'I guess, again, it takes away some of their independence, I would say. Makes it more like a school than an institution of higher education.' A similar point was made by OC9, who was the only respondent to reply explicitly in terms of professionalism: 'Yes, because I think it does cease to treat us as professionals, if you like. But I think, as I said before, it's not just this [benchmarking], it's many other things that have already been starting to do that, both outside the university and inside the university.' One of the other themes underlying the research project, that of conformity, was also raised by just one respondent: 'I suppose it does make me feel that there's less autonomy. ... I have a feeling that universities have to be rather uniform, now' (OC7). OC10 answered very emphatically: 'Oh, yes!', but immediately explained this reaction in terms of the vast increase in student intake over the previous ten years: 'We've ceased, for either good or bad, targeting and training the best, and we're now a mass education system. So we've certainly changed our goal on that.' Unfortunately, OC10 did not elucidate the direct link between this and the introduction of subject benchmarking. For OC11, students were 'measured in terms of being a product.'

5/34 responses were categorised 'other'. OC4 also picked up on the shift from an elite to a mass HE system, but did so equivocally: 'I'm not sure if it was benchmarking that's done that. I think it's more the expansion of the university system'. NC4 resisted efforts to elicit a clear response, but commented on 'a downgrading of the role all the time', which was seen as reflecting 'a lack of trust'.

145
In line with this, academic and scholarly freedom were being eroded (NC4). OC8 acknowledged that his thinking on this item had changed, but then commented: 'we still ... have all the bad things about working at universities, and the good things are being eroded. Whether that's down to benchmarking, I don't know'. This conflict remained unresolved at the end of the interview. In spite of probing, NH3's answer remained unclear: NHH felt unable to answer the question.

Here, too, responses were uniform across the institutions, and the nature of the comments precludes meaningful subject comparisons. It is noticeable, however, that 6/10 of the inexperienced respondents answered 'no' here, only 2/10 answered 'yes', and 2/10 did not know.

**Additional comments concerning benchmarking or threshold standards**

At the end of each interview, an attempt was made to take some of the control away from the interviewer and to give the respondents (n = 34) the opportunity to express themselves freely.

Not infrequently, respondents returned, directly or indirectly, to a point made earlier; others raised issues not germane to the study. OHH, however, referred to the challenges posed to British universities by globalisation and the need for resource efficiency in order to be able to compete in a world marketplace, commenting: 'I think benchmarking is probably one of the tools which, you know, has a positive side, that helps in this. So, I'd be a bit more positive than some of my comments.'

A very different point was made by OH2: 'The only thing that interests me about
subject benchmarking is, really, how long it will last.’

Having invited comments in this way, respondents were then asked if they thought that the change in QAA methodology would make any difference to the status of subject benchmarking. 10/34 respondents provided no data answering this question clearly, and a further 11/34 did not know. 2/34 respondents thought that the advent of Institutional Review would produce no change in the status of subject benchmarking.

Of the 11/34 respondents who thought that the new QAA methodology would result in a change in the status of subject benchmarking, five thought that benchmarking would become more important, though one of those five was rather muted: ‘It may do. In some ways, one hopes that the benchmarking statements become more central’ (NH2). OH4 saw the change as coming from the institutional need to know that degrees were ‘up to scratch. ... Possibly, that will make [benchmarking] more important, actually, won’t it; because, instead of having peers, who know it, you’ve got a more general, managerial approach, which doesn’t know about that and is going to want to rely on something.’ NQE thought that the new QAA methodology would ‘give [benchmarks] more significance. Give them a higher profile in terms of what they’re doing, within the institution. ... because they are coming as a package, now, rather than individually ... they are being seen as a whole package alongside other quality and standards initiatives. They are being foregrounded more.’

A further five respondents thought that benchmarking would become less important. NC1, for instance, took a view diametrically opposed to that just quoted.
from NQE:

[benchmarking is] now one of a raft of things ... that you're supposed to have taken into account when designing a course. ... So, it's not just the one thing; it's part of a cluster of things. And therefore their impact as the major influence on curriculum, I think has been diluted, as a consequence of that (NC1: emphasis in original).

NC3 was a little more hesitant: 'I think it's perhaps less important. You know, as long as you can show that you are meeting the standards in some way, and satisfy [the QAA]'. By contrast, OQS immediately commented on the changes in QAA methodology: 'Well, of course, they greatly reduce [the status of subject benchmarks].'

NC5, who had to ask for an explanation of the new QAA methodology, concluded: 'I think it would make quite a significant difference, actually.' Probing revealed that NC5 considered the new methodology to be 'much more open to abuse' (NC5), understanding it to mean that institutions could choose to put up their best departments for inspection: 'benchmarking, in that particular respect, may actually fail; because the weaker departments, which is what you're really aiming for, are going to slip through the net' (NC5). This anomalous response remained unresolved through lack of time.

4.4 Synthesis

The fieldwork shows that quality systems, whilst seen as necessary and beneficial, were also considered to have considerable negative aspects, a dichotomy identified in both institutions, irrespective of subject and age group. External controls were viewed as an acceptable consequence of an accountability relationship. Also evident
was a commitment to standards and quality teaching, with indications of concern to preserve local autonomy.

Close knowledge of subject benchmarking was not widespread, with exactly one third of the academics asked claiming no familiarity with it. Additionally, there was an almost complete lack in both institutions, at subject, departmental or institutional level, of documentation relating to subject benchmarking.

Overall, a small majority (19/34) saw the introduction of subject benchmarking as good or very good, but only 16/34 rated the performance criteria and/or threshold standard as good (0/34 as very good). However, the respondents reported overwhelmingly that the advent of subject benchmarking had led to no changes of substance, a huge majority (29/34) being unable to identify any specific change attributable to benchmarking. Almost nobody (2/34) thought that subject benchmarking was leading to improvements in quality, and only approximately one third (12/34) of respondents thought that benchmarking would encourage the dissemination of best practice. Whilst only approximately one third (12/34) of respondents thought it had been necessary to introduce subject benchmarking, the need to conform to the benchmarking system was attested in both institutions.

No uniform view emerged as to why subject benchmarking had been introduced, though the explanations offered most frequently were political motivation and public accountability, including a concern for quality and standards. The drive for measurement and quantification was often seen as reflecting a generalized phenomenon.
The impact of subject benchmarking on academics’ self-image and on their perception of the position of universities in today’s society was slight. A frequently identified concern, however, was that benchmarks might become inflexible and restrictive, a small number of individuals talking about this in terms of the danger of a national curriculum for HE.

Less than one third (10/34) of respondents thought that the shift to Institutional Review would affect the status of subject benchmarking, though opinions were equally divided between those who expected the importance of benchmarking to increase and those who expected it to decrease.
This section considers possible meanings to be attached to the findings, and relates them to issues arising from the literature review. The discussion is structured in such a way as to correspond to the research questions.

5.1 The introduction of subject benchmarking

5.1.1 Subject benchmarking
As has been seen, the interviews revealed a broad awareness and acceptance of quality assurance principles, not uncommonly expressed in general terms, such as ‘overall quality assurance’ (OC12) or ‘programmes ... of good quality’ (NH2). Underlying this was often a commitment to public accountability and so an acknowledgement of the need to maintain high standards, though there were frequent instances of academics objecting to the excessive bureaucratic burden of quality assurance. Morley (2003: 55) also reports ‘evidence that academics and
managers tended to support the concept of accountability but found the cumbersome bureaucratic requirements unacceptable. However, in the initial phases of the interviews, there was almost no sign of the kind of outlook adopted by Trow (1994: 15): 'bureaucratic institutions and their mechanisms are the alternative to a relationship of trust between the state and universities.' Nor did the idea of quality assurance seem to undermine perceptions of professionalism, also described by Trow (op. cit.: 27): 'almost everything in a university depends on the inner motivations of teachers – their sense of pride, their intellectual involvement with their subjects, their professional commitments to the role of teacher, their love of students, or of learning'. Those respondents who assessed QAA Subject Review in terms of development or improvement all expressed their disappointment that the potential to promote enhancement remained unrealised.

The detail underlying the figures in Table 4.2 is now considered.

It is striking that in OldU Chemistry, 9/13 respondents rated the introduction of benchmarking as good and 1/13 as very good, giving a total positive response of 10/13, including all four inexperienced chemists. Typically, the view taken was that 'there are certain aspects that are important' (OC11). In OC5's view, 'it's very important to normalize the whole system.'

Both experienced and inexperienced chemists at OldU spoke favourably of benchmarking in terms of standards. OC1 thought that benchmarking could also be seen as 'a good thing from the point of view that in essence one is trying to maintain standards across the board.' Knight (2002a: 9) reports academic staff as
saying 'that they now work with a more diverse ... and/or less capable ... student community on curricula fragmented by modularization and the proliferation of electives.' These dilemmas are reflected by OC2, who rated the introduction of benchmarking between good and very good, 'to maintain standards ... I think as an employer, employs a graduate, should expect a certain amount of ability or otherwise the whole thing becomes meaningless. I mean, there's been, with expansion in higher education – I think [a] more standard approach – I can't see any harm in it.' Dearing (NCIHE, 1997: para.10.7) also comments that 'the expansion of student numbers has put the existing quality assurance arrangements under strain'; adding (NCIHE, op.cit.: para.10.8): 'We believe there is a need to develop quality assurance practices which allow for diversity throughout the system, yet ensure that diversity is not an excuse for low standards or unacceptable quality.'

Yet the expressions of approval for the introduction of subject benchmarking need to be put into context. OC1's positive comment, for instance, is off-set by the conviction that, 'basically, there seems to have been a loss of trust, and ... it is not accepted that we are able to control ourselves', thus actually leaving OC1 neutral about benchmarking. On the other hand, OC2 had not only not seen the chemistry benchmark statement before, but was also unfamiliar with the concept: 'Can you give me a breakdown, exactly what you mean by subject benchmarking.' In other words, underlying OC2's approval of the idea of benchmarking would seem to be the tacit assumption that it had no real impact on OldU Chemistry, so that approval would in essence be approval of benchmarking in its application elsewhere.
Support for this interpretation is to be found in the responses of OC8, who rated the introduction of benchmarking as good, but then explained at the end of the interview: ‘we haven’t had a lot of information given to us about this.’ This comment would seem to rule out, at least within the department, the influence of ‘discursive capture’, defined by Becher and Trowler (2001: 165) as, ‘the notion that individuals’ view of the world is limited by dominant discourses’. Moreover, among the ten OldU chemists rating the introduction of benchmarking as good, none reported it as prompting changes to courses, and seven stated that benchmarking had had no impact on their professional lives, whilst three referred to bureaucratic consequences of the system.

The view that benchmarking was good when applied elsewhere emerged very clearly when OC3 assessed the introduction of benchmarking in two ways: whilst, at the national level, it was ‘probably good’, ‘At [OldU], we easily met all the levels, so it is of relatively little value.’ However: ‘So many places can call themselves universities that it may be of some value to define minimum standards’ (OC3). A similar line was taken by OC6: ‘Any decent university course ... would already be doing everything that’s contained in [the QAA benchmark statement].’ These comments are reminiscent of Dearing (NCIHE, 1997: para.3.65), who also refers to the ‘more means worse’ criticism of UK HE. As Gordon (1995: 26) observes: ‘All stakeholders have an interest in comprehending what is meant by an award and knowing that the standards of such awards are consistently high.’ The British government’s (DfES, 2003: para.1.9) view of HE expansion is unequivocal: ‘It is not the case that “more means worse”.’ Specifically on the use of the
university title, the government (DfES, op.cit.: para.4.34) states: ‘we will examine and modernise the criteria for degree-awarding powers to reflect the increasing diversity of higher education. ... But there will be no relaxation of the high standards that have to be reached before taught degree awarding powers are granted.’ The fieldwork suggests that there is scepticism in the HE sector as to the veracity of the government’s claims.

In OC6’s view: ‘I would guess that the benchmarking document has had zero effect on this department. But this one, the RSC one, has had a lot of effect.’ Indeed, the commitment demonstrated by OldU Chemistry to working with the RSC may contribute to the high ‘good’ rating accorded to the introduction of subject benchmarking. OC1 explained that the RSC ‘needs to have some standards that it wants to maintain’. Even though gaining RSC accreditation is ‘quite a substantial hurdle’ (OC3), ‘[The RSC is] the professional body for chemistry, and I think it’s important that there’s some sort of recognised accreditation that relates to the content of the material that you deliver’ (OC13). The idea of obtaining approval from an external body is well established in OldU Chemistry and successfully pursued. Additionally, as OC7 put it: ‘the [RSC] accreditation exercise I think was much more demanding [than QAA benchmarking]’, which may also have had the effect of making benchmarking easier to accept. As the GSP (HEQC, 1996: Appendix, para.5.1.1) observes: ‘RSC accreditation effectively agrees a core curriculum for the preparation of career chemists’.

The profile of responses in NewU Chemistry is very different, with only 1/6
respondent rating the introduction of benchmarking good, 3/6 respondents rating it neither good nor bad, and 2/6 giving no definite answer. At first sight, it may be tempting to interpret the ambivalence towards benchmarking as arising from a need for self-assertion of a kind perhaps underlying a comment such as that from NC5, who acknowledged having no familiarity with subject benchmarking: ‘I would hope that if I see a benchmarking system, I’m already working to that standard.’ Similarly, NC2 found RSC guidelines helpful, ‘because it tells me that the questions that I am asking of students in an exam are the types of questions that they’re going to be asked at a traditional university.’ This relative caution may be a legacy of the institutional history of a new university, as described by Trow (1994: 29): ‘traditional values of autonomy were on the whole less firmly embedded in the polytechnics, which had always had less autonomy, and were, so to speak, used to substantial influence from local authorities, local industry, and later on from the [CNAA].’ As Pratt (1997: 235) observes of the CNAA: ‘its very existence confirmed the status hierarchy in higher education between the universities and the rest.’ In this connection, it will be recalled that NH4 drew attention to the potential benefit of subject benchmarking for an institution such as NewU which may not have a national reputation, ‘being a new university’.

It may also be, however, that the reactions in NewU Chemistry stem from a greater direct involvement with the benchmarking process. NC3, for instance, observed: ‘It’s perhaps good for students, in that they can see the precise skills that we’re trying to teach them. But I think it’s been bad for staff in terms of the extra work that it’s involved’. For NC4: ‘It probably depends on how people apply it.’
Furthermore, NC4 stated: ‘You need some arbiter of standards. But, as to whether benchmarking achieves that, I have severe reservations’, a view in direct contradiction to the QAA’s (1999a) claim that subject benchmarks should make it possible to identify broadly comparable standards of attainment.

Only 1/6 NewU chemist rated the introduction of benchmarking as good, as it promoted reflection on what constituted a graduate in a particular subject: ‘particularly in areas where there are no professional bodies and therefore no definition of what a course is’ (NC1). By contrast: ‘those of us who have to have certified courses, it’s a slightly different kettle of fish, because you already had, to some extent, some idea of what somebody thought ought to be in a degree course’ (NC1). This outlook is rather similar to that identified in OldU Chemistry, in that it supports the introduction of QAA subject benchmarking even though that is seen as having relatively little direct relevance in the respondent’s own department. As with OldU Chemistry, involvement with the RSC seems to have played a significant part in the ‘good’ rating. However, an interesting contrast is provided by NC3, who commented on the introduction of benchmarking: ‘I think it’s a bit of a waste of time, because we are so heavily prescribed anyway, in terms of what we can teach, because of the professional requirement.’

Although neither inexperienced chemist at NewU rated the introduction of benchmarking as good, neither of them rejected it. NC2 observed that benchmarking was better than a situation in which there was nothing of this kind, and NC5 hoped that he would already be at the standard of a benchmark: ‘we
would all like to think that [the benchmark] doesn’t apply’.

It is important not to over-interpret the differences recorded between OldU Chemistry and NewU Chemistry. After all, of the ten OldU chemists who rated the introduction of benchmarking positively, only four also considered it to be necessary. Amongst these four was just one of the inexperienced academics; not included was the one respondent to give the introduction of benchmarking a rating of ‘very good’. The one NewU chemist to rate the introduction of benchmarking as ‘good’ rated it as not necessary. Becher and Trowler (2001: 44) posit: ‘we may appropriately conceive of disciplines as having recognizable identities and particular cultural attributes.’ The fieldwork suggests sub-sets of complexity underlying such identities. As Gordon (2002: 215) observes: ‘Each individual in a course team will hold personal opinions and values and respond to particular drivers and motivators. There can be many common items amongst the team and yet there can still be significant detailed differences in priorities and values.’ Similar variances are likely to have been at play in OldU Chemistry and NewU Chemistry. Lack of time precluded an investigation of individual understandings and value systems, which may have made it possible to disaggregate departmental responses and achieve a more sophisticated analysis of perceived discrepancies between the two chemistry departments.

In the case of history, opinions at both OldU and NewU were exactly evenly divided between good and neither good nor bad. Given the small sample size, it is not possible to draw any reliable conclusions from this result, though it is clear that
support for the introduction of benchmarking is far from strong in this group. Interestingly, however, the strongest expressions of dissatisfaction came not from experienced academics, but from the three inexperienced respondents, who rated the introduction of benchmarking as neutral (OH1, OH3) or as part good and part bad (NH4): good in terms of general criteria for standards, but bad in terms of it being imposed 'from above' (NH4).

Overall, positive evaluations of the introduction of benchmarking were: chemistry, 11/19; history, 5/11 (cf. Table 4.2). The differential reactions between the two chemistry departments make it unlikely that this proportional difference between the two subject areas can be adequately explained by the lack of a professional body for history.

In its consultation document setting out initial proposals for subject benchmarking, the QAA (1998a: Part IV, para.4) states: 'The benchmark information will form an important part of the overall framework for quality and standards that will be developed over the next few years.' The purpose of the statements of standards produced is to form reference materials for the 'task of verifying and comparing standards of student attainment' (QAA, op.cit.: Part IV, para.4). On the outcome of this consultation, the QAA (1998c: 18) reports: 'HEIs were divided: about 25% expressed support for the principle of benchmarking, many proposed positive suggestions of various kinds; but most were neutral or expressed reservations. A minority were opposed in principle, or highly critical.' In the fieldwork, a somewhat different pattern of responses emerged: as Table 4.2 shows, 19/34
respondents reacted positively to the introduction of benchmarking, 15/34 nevertheless remained neutral or equivocal, but there were no negative responses. The highest level of approval came from quality administrators (3/4), though it is striking that NQS welcomed the introduction of subject benchmarks as they were akin to a national curriculum, the very interpretation the QAA (1997: para. 50) was at pains to disavow from the outset. Generally, then, the fieldwork partially reverses the rather negative trend identified by the QAA in 1998 regarding institutional reactions to subject benchmarking.

On the question of possible differences between types of HE institution and amongst different academic discipline communities in terms of their reactions to the introduction of subject benchmarking, both Dearing (NCIHE, 1997) and the QAA are silent. Whilst there are studies considering how subject benchmarks can be used (Jackson, 2000b), how they have been used by a department (Harrison, 2001; Houghton, 2002) or an institution (Hargreaves and Christou, 2002), or what their implications might be in UK HE (Holloway and Francis, 2002), the studies take as their starting point the implementation of the benchmarking system and do not consider reactions to its introduction. The absence of any similar study makes it impossible to assess the findings of the fieldwork for their typicality. Newton (2002: 59), however, emphasizes the merits of 'close-up study and insider research into “views from below”' and the great need for 'insights into issues around the implementation of quality policy ... and how key actors receive and respond to policy and change in higher education organisations.'
One insight to suggest itself here concerns the attitudes of inexperienced members of staff. Whilst the inexperienced respondents in both chemistry and history were generally accepting of, if not necessarily in favour of, subject benchmarking, expressions of disapproval came only from historians. On the one hand, this suggests that a new generation of academics may, indeed, be more ready to accept external scrutiny as a normal part of their activities; on the other hand, it may highlight a fundamental cultural difference between the two disciplines. Indeed, whilst the benchmark statement for chemistry (QAA, 2000c: 3) notes the need for the ability 'to recognise and implement good measurement science and practice', the benchmark statement for history (QAA, 2000d: 3) emphasizes the historian's 'intellectual independence'. Similarly, whereas OC8 found the introduction of benchmarking a good idea because, 'as a scientist, it's good to have something to measure something by', OHI seems to encapsulate the dilemma of the tension between acceptance and independence by commenting on 'the realities of placing lines ... which we have to meet, or boundaries over which we must hurdle ... they're quite often false', and reflecting: 'as a new member of staff, I'm probably more likely to be accepting of these things than the older members of staff; and I am, probably; but ...

5.1.ii **Threshold standards**

Doubts about the usability of benchmark statements had a noticeable effect on the evaluation of the statement of threshold performance in chemistry (QAA, 2000c: 5) or threshold standard in history (QAA, 2000d: 12). Although 10/13 of OldU chemists rated the introduction of benchmarking as good or very good (cf. Table
only 6/13 rated the statement of the threshold standard as good (cf. Table 4.3). The principal objection raised was the lack of objective clarity. By contrast, whereas only 1/6 NewU chemist rated the introduction of benchmarking as good, 4/6 of them declared the threshold statement to be good, seeing it essentially as a mechanism for comparing departments. It would seem, then, that OldU chemists, accustomed to the rigours of RSC validation, were critical of the lack of rigour of the QAA threshold statement; whereas NewU chemists, cautious about QAA benchmarking, were nevertheless reassured by the confirmation to be achieved via nationally applied comparability.

In OldU History, 3/6 rated the introduction of benchmarking as good, and 4/6 rated the threshold statement as good. Since this increase amounts to just one additional respondent, no reliable conclusion can be drawn. However, given the lack of clear support for benchmarking in this group, it is tempting to see the result as an expression of making the best of a bad, or at least not particularly good, job. In NewU History, opinions about both the introduction of benchmarking and the threshold statement were equally split between good and neither good nor bad, thus suggesting that support for the threshold statement was not particularly strong.

Asked generally about the statements of threshold standards established for single honours degrees, no quality administrator gave an overall rating higher than neutral, continuing the pattern of a lack of clear support for the threshold statements.

In its initial evaluation of the benchmarking process, the QAA (1999b: para.6) notes among BMG members 'broad satisfaction with the information produced'. This
finding was not replicated by the fieldwork. On the contrary, less than half of the respondents (16/34) reacted positively to the threshold standards; there was no rating of very good; and 18/34 reacted at best neutrally, with 5/34 respondents giving a 'bad' rating (cf. Table 4.3). Overall, the fieldwork seems to identify the view that statements of threshold standards are a reasonable idea not well executed.

Amongst academic staff, ratings of 'bad' for threshold statements came only from chemists, which may be another manifestation of a cultural difference between chemistry and history worthy of further investigation.

Overall, 19/34 respondents classed the introduction of benchmarking as good or very good, whilst 11/34 remained neutral (cf. Table 4.2). When it came, however, to the statements of threshold standards, only 16/34 respondents classed the items as good, 8/34 remained neutral, and 5/34 classed the items as bad (cf. Table 4.3). The fieldwork does not, then, replicate the QAA's (1999b: para.9) initial finding that the academic community's view of the benchmarking process is 'broadly positive'. Neither, however, does the fieldwork detect much strong opposition to subject benchmarking, of the kind posited by some commentators (Tapper and Salter, 1998; Utley, 1998).

5.2 Subject benchmarking and improvement

From the literature review, it emerged very clearly that the standard understanding of benchmarking is composed of three elements: it is based on analysis of best practice, sometimes with a reciprocal element; it is dynamic, a continuous process and not a one-off; and, above all, it aims at improvement.
5.2.1 Subject benchmarking and best practice

Middlehurst (1997a) identifies an important link between, on the one hand, the identification and dissemination of good practice and, on the other hand, the enhancement of quality. According to the QAA (2002c: para.10), the first objective of institutional audit is: ‘to contribute, in conjunction with other mechanisms, to the promotion and enhancement of high quality in teaching and learning’. In the fieldwork, however, only 12/34 respondents (cf. Table 4.4) thought that benchmarking would encourage the dissemination of best practice, whereas 22/34 either thought that benchmarking would not help in this regard, or did not know. Given the centrality of subject benchmarking to the institutional audit process (QAA, op.cit: para.12), it is clear that, in the sample studies, QAA subject benchmarking was not perceived as fulfilling a purpose normally taken to be integral to any benchmarking system. Particularly forthright was the comment of NCH: ‘If [subject benchmarking] doesn’t [encourage the dissemination of best practice], it will probably have been a waste of time, I would have thought.’

No consistent picture emerged as to what would encourage the dissemination of best practice, though several respondents mentioned the Learning and Teaching Support Network. In particular, NHH observed: ‘you get more worried if you’re being told what ought to be taught’, finding it much more sensible to concentrate on ‘best practice in terms of teaching methods’.

5.2.2 Subject benchmarking and dynamism

From the ‘Xerox definition’ (Coopers & Lybrand and CBI, 1993: 4) onwards,
benchmarking has been seen as a continuous process requiring a dynamic approach to a constantly evolving environment. The fundamentality of this underlying principle led to objections to the QAA’s benchmarking proposals: ‘It was ... pointed out that in normal usage “benchmarking” was a term used to denote a means for continuous improvement of quality and standards; and there was a danger that the proposals ... would lead to stasis’ (QAA, 1998c: 18). In its own ‘response and proposals’ (QAA, op.cit.: 19) in answer to such specific issues, this misgiving is not addressed. The QAA’s original intention had been to exploit benchmarking in Subject Review, ultimately producing an overview report for each subject: ‘In the light of this report the subject community will be invited to consider any adjustment to the subject benchmark statement as may be necessary’ (QAA, 2000f: para.52). With the abandonment of Subject Review, QAA benchmark statements hold sway for an unspecified period and are open to the criticism of being static and inflexible. As OH5 put it: ‘My whole fear with anything, when you write down rules, is that it can also become restrictive’. Not only can it rightly be wondered if such an approach is appropriate to any academic discipline (Holloway and Francis, 2002: 241), there is also a real potential for it to ossify into one of the QAA’s ‘regulatory bureaucratic mechanisms’ (McNay, 2004: 12). If this proves to be the case, then the answer to Brown’s (1999b: 278) question quoted in the introduction to this study is that the efforts to benchmark students’ achievements are focused on measurement rather than on the potential benefits of improvement. In this way, benchmarking becomes a form of performance indicator rather than making a constructive contribution to the enhancement of ‘the quality of
programmes and the standard of awards at the point of delivery’ (QAA, 2002c: para.3).

The static nature of QAA benchmarking also impedes the identification of performance gaps which enable the operationalisation of change and improvement (Alstete, 1996). It may be argued that the publication of institutional audit reports overcomes this impediment: ‘The concluding section of the report may also highlight features of good practice in the management of quality and standards at institutional level and within the disciplines selected for audit trail’ (QAA, 2002c: para.60; emphasis added). The uncertainty of the coverage provided by the reports makes them, however, an erratic source of information. Neither does the QAA (2003c: para.48) overcome the static nature of the system by providing interim insights, stating merely:

In general ..., the [audit] reports suggest that the eight institutions have responded appropriately to the introduction of the FEHQ and subject benchmark statements. ... All of the institutions are found to have given consideration to the purpose of these external reference points and undertaken some initial work to address them.

The static nature of QAA subject benchmarking is thus a further aspect not corresponding to the normal conception of benchmarking which emerged from the literature.

5.2.iii Subject benchmarking and quality enhancement

What clearly emerged from the literature review was that the standard understanding of the raison d’être for benchmarking is to achieve improvement.

The QAA’s (2002c: para.1) commitment to improvement is clearly asserted: ‘The
mission of the Agency is to promote public confidence that the quality of provision and standards of awards in higher education are being safeguarded and enhanced.' It is, then, a striking finding from the fieldwork that 29/34 respondents stated that the introduction of benchmarking had not led them to change their approach to their courses or to the institutional approach to quality, or at least that they were aware of no such changes. This may be because the benchmark statements are perceived to be so obvious as to amount to 'motherhood statements' (OC6), or, as OH5 put it:

I also suspect that, in History, the benchmarking document was drawn up in a catch-all way, in which it probably described what was going on anyway, and that's why it wasn't a fantastically good thing to do, because it's basically said what we were doing already was pretty much what we should have been doing.

The explanation may also be found, however, in the record of the subject communities themselves. The HEFCE (1995a: para.2) Subject Overview Report – Chemistry, for instance, records:

The results of the quality assessment exercise ... show that no unsatisfactory provision has been found in chemistry and that about 18 per cent of the providers of the subject were identified as excellent. Although these results leave no room for complacency, the broad 'satisfactory' category contains a range of institutions which provide a sound education in chemistry, suggesting that chemistry students in England are generally exposed to a beneficial learning experience.

In particular: ‘As judged by external examiners, the standards achieved by chemistry students in English universities and colleges are at least satisfactory and sometimes excellent’ (HEFCE, op.cit.: para.36).

The corresponding overview report for History (HEFCE, 1995b: para.1) affirms that 'history students can confidently expect to receive at least a sound education and often a very fine one', and emphasizes (HEFCE, op.cit.: para.7): 'There were
no unsatisfactory providers in history.' Here, too, no problems were detected regarding standards; on the contrary: 'Overall, the level of student achievement in history is impressive. ... The standard of students' work was at least appropriate to the level of the award, and this was supported by external examiners' comments. The standard of marking was also fair and appropriate' (HEFCE, op.cit.: para.28).

In other words, in neither of the subjects here investigated was external assessment revealing any deficiencies in the quality of provision or the standard of awards.

In the fieldwork, some respondents referred to the challenges of a shift from an elite to a mass system of higher education or of an increasing modularisation of undergraduate programmes. It has also been seen that Brown (2000) draws attention to benchmarking as a potentially valuable form of enhancement, an objective stressed by the QAA (2003d: 59) in its explanation of academic review at the subject level: 'Student achievement of intended learning outcomes is given greater prominence, as are assessment and the management and enhancement of quality and standards.' Nevertheless, when asked if benchmarking was leading to improvements in quality, 32/34 respondents either replied in the negative or could give no clear answer. As NC3 put it: 'I doubt it, because, you know; as I say, we were teaching the same material before, and we shall continue [laughing] to teach the same material afterwards. It's just the way you talk about it that's different.'

The clear finding of the fieldwork, then, is that QAA subject benchmarking is not perceived to be achieving the goal normally considered to be essential to any benchmarking system, in that it is not leading to improvement in quality.
Many respondents had no concrete suggestion to offer as to what would improve the quality of single subject honours degrees. Among those who did make suggestions, however, the recurrent theme was resources, expressed preponderantly as more money, better facilities or, particularly, more favourable staff-student ratios. Knight (2002a: 9) notes laconically of the predicament of academic staff: ‘Resource complaints are endemic.’

Overall, subject benchmarking is not seen as corresponding to any of the three principal characteristics of benchmarking identified from the literature.

5.3 Consequences of subject benchmarking

Only 12/34 respondents in the fieldwork considered the introduction of subject benchmarking to be necessary; for 14/34 it was not necessary, 22/34 giving an assessment that was either negative or ambivalent (cf. Table 4.5).

Dearing’s (NICHE, 1997: para.10.55) premise, leading to the recommendation to introduce subject benchmarking, is that there is ‘concern about the standard required for the award of higher education degrees.’ It has been seen that, when asked why benchmarking was introduced, just 8/34 of the respondents mentioned standards in some way; but no respondent identified standards as a significant problem recognised by the sector. By contrast, OC3 observed: ‘I really do think that in the government philosophy it was necessary to have these kinds of probes and checks to ensure, as they saw it, high standards.’

Fry (1995: 75) sees the quality movement as part of the ‘Rise of the Evaluative
State’. Although the term ‘evaluative state’ occurred at no point in the fieldwork, it may not have been unreasonable to have expected the frequent references to a political explanation for the introduction of subject benchmarking, combined with considerations of accountability requirements, increasing quantification and decreasing trust in professionals to have raised issues such as that described by Finch (1997: 147): ‘A key question which troubles academics themselves is whether the traditional authority of academics over their own subject areas is being compromised by systems of externally driven inspection, particularly of teaching quality and standards.’

5.3.1 Benchmarking and academic identity

In line with Henkel’s (2000a) findings, respondents generally did not object to quality assurance mechanisms, though many did object to the administrative burden involved. Elsewhere (Henkel, 1997), Henkel notes that academics generally accept public accountability, a finding also replicated here. At the same time, it has also been seen that the fieldwork revealed signs of an underlying mood similar to that identified by Power (1997: 135), that public sector audit corresponds to an increase in ‘institutionalized distrust in the capacity of teachers, social workers, and university lecturers to self-regulate the quality of their services’. However, the fieldwork produced little support for Doherty’s (1997: 247) contention: ‘Morale is low.’

According to the QAA (1998c: 19): ‘benchmark information constitutes a key element of the “new compact for higher education” envisaged by the Dearing
report.' Newton (2000: 162) contends that the QAA's quality framework 'shows signs of being far more invasive and forensic in its impact on the work practices and work situation of ground-level academics than anything which has preceded it' (cf. Newton, 2002: 60). In the fieldwork, however, 24/30 of the academics questioned either declared that benchmarking had not led to changes in the way they thought about their own position as academics, or at least gave no clear answer. The impact of subject benchmarking on academic identity was thus negligible as far as the sample group was concerned. At the same time, it was striking that some of the strongest statements against benchmarking and its underlying principles came from academics relatively new to the profession. Whereas OC2 was unsure as to whether the fact that academics were now less free to do what they wanted was a good thing or a bad thing, OH1 (cf. 5.1.i) seemed to acknowledge the intuitive assumption that young, inexperienced members of staff will have grown up more accustomed to a 'regulation culture' (NH1) and so be more relaxed about developments such as subject benchmarking. In fact, OH1 seems to display an outlook similar to that already referred to from Henkel's (1997: 142) work, that academics accept the modalities of a mass HE system, but hanker after 'the values and modes of working that belonged to an elite system.' Thus, OH1 protested: 'these benchmarking levels ... the skills we expect students to develop in the course of a degree: I don't need to specify those, because I know my students will have to do those things in the course of any degree.'

5.3.ii Benchmarking and universities

As has been seen, the fieldwork discovered essentially no effect of benchmarking
on the way respondents thought about the position of universities in today's society, and so provided no strong support for a view such as that expressed by Inglis (2004: 16): 'This loss of authority is acutely marked in the universities.' Here, too, though, there were some signs of allegiance among young academics to a more traditional conception of a university, as when OH1 even expressed hesitancy about the use of the word 'quality', considering it to be part of a business approach to education, which 'I think it is actually in direct conflict, really, with liberal, historical understanding of the higher education system and what we’re here to do. I don’t think I’m here to provide a service to customers. I think that my role as an educator is far more complicated than that.'

Commentators (Trow, 1996; Spender, 2000) argue that government-imposed quality assurance and accountability requirements have had serious deleterious effects in terms of concepts of professionalism and HE culture, leading to 'a kind of mass degradation ceremony, involving the transformation of academic staff ... into employees, mere organizational personnel' (Trow, 1996: 318). The sample size in the present study is clearly far too small to permit any attempt at generalisation from the comments on academic identity or institutional status, but the perhaps counter-intuitive reactions from young academics suggest an intriguing field for future research.

Participants in the fieldwork saw subject benchmarking as having very little impact on academic identity, and almost none on their views regarding the position of universities in today's society.
5.4 Benchmarking and conformity

5.4.1 Cultures of conformity

The fieldwork makes it clear that, in both institutions investigated, steps had been taken to ensure conformity with the principles of subject benchmarking, though it proved almost impossible to obtain documentary exemplification of the processes and actions involved. OH4, however, commented: 'But, what we’re talking about with Subject Review, QAA and all the rest of it is trying to satisfy bureaucracy.' As Gosling and D’Andrea (2001: 10) observe: 'the bureaucratic demands of quality assurance are creating a compliance culture that dampens creativity, rewards conformity and slows down the responsiveness of the system to a rapidly changing environment'. OH4 added after the remark just quoted: 'And then it becomes obsessed with certain rules. You know, if every student doesn’t achieve the learning outcomes you will lose points et cetera; and you’re sort of: “Oo, God! How do I satisfy that? We can’t control this.” [laughs]'. As Brennan (2001: 1) indicates: 'Staff in higher education institutions, especially in the UK, are focused on how to get a good “result” from whatever quality assurance process they face.' Harker (1995: 31) notes: 'While quality assurance is ostensibly about the maintenance of excellence, it is fundamentally about accountability and control.' In Cox’s (1999: 26) view, the UK emphasis on accountability rather than improvement ‘has encouraged a compliance view of quality.’

At OldU, History was seen to have engaged with subject benchmarking rather more closely than had Chemistry. This may be explained partially by OldU Chemistry’s familiarity with the RSC’s accreditation processes, which were seen to be more
rigorous and demanding than QAA benchmarking. The difference may also be explained at least partially by the fact that OldU History was in the process of preparing for a QAA visit. Indeed, OHH commented on the usefulness of benchmarking and threshold standards for inspection purposes, and OH4 felt that he was affected by 'the Subject Review, rather than the benchmarking itself; but they are all rather linked, aren't they? Otherwise, nobody — well: potentially, nobody would have to pay attention to the benchmarking.' Smith (H.) (1999: 206) warns that the imposition of benchmarking will have 'subsequent effects on motivation and compliance'. Enarson (1975: 172) makes the general comment: 'techniques and tools tend not to be the neutral servants we describe them to be. Techniques and tools are used, always, by persons operating in time, place, circumstance, culture, and power relationships' (emphasis in original). Lemaitre (2002: 34) observes: 'Definitions of quality are never neutral, or innocent. They are about balances of power, within higher education and between higher education and other social actors.' As Brown (2004: 69) observes: 'the effect, if not the aim, of the Dearing Committee (and subsequently ... the QAA) was to create a framework through which institutions' compliance with collectively agreed benchmarks could be monitored and regulated externally.'

At the institutional level, OQS was careful to emphasize that changes resulting from benchmarking were 'changes to our quality systems, not the quality of provision'. Nevertheless, OQS did wonder how much people were 'looking at benchmark statements, taking the language from them, reciting the language in their own documents, without actually changing or doing anything. I do wonder to what
extent that's genuine engagement with benchmark statements, rather than just, you
know, going through the motions'. As Holloway and Francis (2002: 246) note:
'Ultimately the benchmark statements will be applied as part of an assessment
process with political origins and potentially serious political, educational and
financial repercussions on individual institutions, which may foster dysfunctional
behaviour rather than encourage a greater willingness to reflect on and improve
practice.'

The contrast in the reactions of the two chemistry departments visited (cf. 4.2.i)
may be explained by the differing institutional traditions, in particular NewU's
former involvement with the CNAA. Lund and Jackson (2000: 17) comment: 'The
Council established what now would be called benchmark "process" standards for
programme validation and approval and other institutional regulatory mechanisms,
with which polytechnics and colleges whose programmes were validated by the
CNAA had to comply.' Nevertheless, representatives of both departments spoke of
the considerable expenditure of effort involved in generating the documentation
required as a result of the introduction of subject benchmarking. Moreover, this
effort was expended even though, for OldU: 'I can see that we will satisfy all of [the
benchmark statements] as our [degrees] currently stand' (OC3: e-mail); and at
NewU the only resultant change reported was presentational, as it was discovered
that NewU Chemistry 'did comply' (NC1). As Brennan and Shah (2000: 15)
observe: 'Quality assessment ... affects the distribution of power within higher
education.'
In NewU History, NH1 referred to the recent rewriting of module specifications, during which, ‘the benchmarking document became very useful, because you rooted around for the desirable phrases, to some extent.’ NH1 thought that benchmarking did stimulate reflection: ‘But it was only limited, in that I think people felt they were undertaking an exercise, really; that it was happening, they contributed to it, they had a discussion, that was quite useful. Then, I think, it stopped. You know, from the point of view of the everyday historian working on the ground.’ Jackson et al. (2000: 201) warn that there is in subject benchmarking ‘very real tension between promoting professional debate about the learning embodied in standards and the political imperative to show that there is indeed a level of consistency in outcome standards that drives systems towards conformity.’

There is a hint of this in NH1’s reflection on the implementation of accountability in new universities: ‘because they’re very anxious to prove that they’re rigorous and as good as anybody else – [it] has meant an overload of paperwork. But also a tendency to produce structures that are rather rigid and stifling.’

The fieldwork revealed no significant differences between the two history departments. Here, too, OldU History’s impending QAA visit is likely to have coloured attitudes. The comment just quoted from NH1 is, however, reminiscent of the general observation offered by the History at the Universities Defence Group (HUDG, 1998: para.6): ‘Our serious doubts about this Dearing Committee recommendation [21, introducing Programme Specifications] are based upon comments received from some of the new universities where historians have found that managerial pressures to comply with standardised procedures are both more...
intense and more distracting than elsewhere.

At the institutional level, the almost total lack of information from NewU makes it difficult to draw reliable conclusions. It does seem, however, that whereas OldU was able to adopt a very relaxed stance on the action required as a result of subject benchmarking, NewU was more concerned about being seen to comply with the system (cf. 4.2.ii). Once again, the explanation for this difference is likely to be found in NewU's institutional background as a former polytechnic, combined with a familiarity with rigorous and demanding CNAA processes.

As has been indicated, no single view emerged as to how respondents interpreted the shift from Subject Review to Institutional Audit in terms of the status of benchmarking. However, an interesting perspective, also suggesting an area for further research, was offered by OQS: 'I think, probably, the relative lack of interest in benchmark statements is to a fair extent tied to the abandonment of the original QAA'.

5.4.ii Benchmarking and national curriculum

The threat to diversity identified by Tapper and Salter (1998) scarcely emerged in the fieldwork, and even then somewhat hesitantly (NH3, NQS): 'I feared, at the beginning, that it might also reduce innovation; but I feel less strong on that than I did' (OH2). A common theme was concern that benchmarking should not be prescriptive (NCH, NHH, OHH), or, as OC6, a former Head of OldU Chemistry, put it: 'nobody really wants to produce a syllabus, on the grounds that universities themselves ought to decide which areas of chemistry are particularly important.'
From the outset, the QAA (1997: para. 50) indicated its wish to avoid the risk of benchmarking creating a national curriculum. Nevertheless, Goodlad (1999: 78) highlights 'the danger of sliding into a national curriculum for higher education ... [which] would spell death to university work.' By contrast, OC12 indicated that subject benchmarking was not problematic in scientific subjects such as chemistry, for which there already existed a core curriculum nationally. Haines (1999: 144), a member of the chemistry benchmarking group, observes that the chemistry document is 'a statement of a very small core of essential knowledge and nowhere near the national curriculum for higher education that many had feared might be imposed on them.' As Holloway and Francis (2002: 247) note: 'Probably the strongest fear held in the academic community – that national curricula would be imposed on universities through the subject benchmark process – has thus far not materialised.'

At the same time, the risk identified by commentators such as Tapper and Salter (1998) and Brown (1999a), that benchmarking could herald a national curriculum for HE, is still present: 'What implicitly happens is that perhaps things carry on because certain people are leading in a certain direction. But the majority of colleagues aren't even aware that you are moving towards a situation where almost there is a standard curriculum' (NQS). NC4 was constantly worried about a national curriculum for HE: 'You need much more flexibility than that, to reflect the departmental ethos. There ought to be a core syllabus, but not rigid prescription.' OC8 recalled colleagues threatening to resign if a national syllabus was introduced for university teaching, and NH3 closely paralleling Phillips' (1991) account,
observed: 'That’s my only concern about it: that it’s the first step on a dreadful route to national curriculum.'

In the fieldwork, only one respondent took a positive view of a potential national curriculum for HE: ‘there are accepted paradigms within subjects, and consequently there is nothing radical in expecting that people would keep within those particular paradigms’ (NQS). NQS did concede, however, that such an approach made it more difficult to become a radical in some disciplines, as that ‘takes you outside that mainstream benchmarking.’

In spite of the fact that the fieldwork identified manifest actions in both departments at each institution to ensure conformity to the requirements of the subject benchmarking system, the approach at OldU was clearly more relaxed than at NewU, and, as such, somewhat reminiscent of the concept of procedural compliance, defined by Bryman et al. (1994: 178) as:

> a response to an organisational innovation in which the technical requirements of the innovation ... are broadly adhered to, but where there are substantial reservations about its efficacy and only partial commitment to it, so that there is a tendency for the procedures associated with the innovation to be adhered to with a less than total commitment to its aims.

The differential between the two institutions may be explained in part by the greater self-confidence of the longer-established OldU and in part by NewU's experience of CNAA control.

Explicit references to a national curriculum were rare, and almost entirely negative. More common, however, was the desire to avoid the rigidity implied by a national curriculum.
5.5 *Synthesis*

The fieldwork demonstrates that, in the two institutions investigated, subject benchmarking has had little effect beyond the bureaucratic. In particular, it is not perceived as having contributed to the enhancement of quality or standards, terms used essentially interchangeably by the respondents; nor is it seen to promote the dissemination of best practice.

Some support was expressed for subject benchmarking in terms of identifying the essence of a discipline (cf. Hargreaves and Christou, 2002:188), thus suggesting sympathy with the notion of conformance as 'a concern for consistency and reliability' (Gaster and Squires, 2003: 263), though whilst voicing the concern that benchmarking should not become 'merely ... conformance and the negation of difference' (Morley, 2003: 45). Specifically, a national curriculum for HE was viewed as 'against the spirit of academe' (OC9).

However, the subject benchmark statements were seen as simply listing current practice; but, as McNair and Leibfried (1994: 330) admonish: 'If you want to maintain the status quo, then, don't benchmark.' Indeed, QAA subject benchmarking does not display the characteristics normally associated with benchmarking. Given the perceived failure of subject benchmarking to bring about improvement, or even change, it becomes a manifestation of 'coercive accountability' (Shore and Wright, 2000) associated with the rise of audit culture in HE. Engagement with QAA subject benchmarking thus becomes a kind of ritualistic conformity motivated not so much by the desire to realise intrinsic...
potential as by the need to satisfy an externally imposed constraint.

According to the QAA (1998c: 19), benchmark information ‘will have no value unless it commands the support of the academic community and other interested parties, is of practical value, and is fully understood by all those involved.’ The fieldwork suggests that this is some way from being achieved.
Chapter 6

Conclusion

6.1 The research questions

This study investigates reactions to the Dearing-inspired (NCIHE, 1997) quality agenda by considering two broad domains. One domain is concerned with how university staff members perceive and react to subject benchmarking. Of particular interest is the extent to which subject benchmarking, as an integral part of the academic infrastructure, is seen as being a means of facilitating improvement, in line with the typical view shown in the literature, or as having some other function within an externally generated quality assurance framework. The other domain involves considerations of power, control and compliance, both at the level of the individual and at the level of the institution. This second domain further includes the possibility of academics in particular, as opposed to university administrators, perceiving the externally imposed system of subject benchmarking as undermining
their own professionalism, and hence as provoking modifications in individuals’ own self-image and understanding of the relationship between universities and society, this perhaps also resulting from the perceived threat of the potential imposition of a national curriculum for higher education.

Even though a single-researcher project could not operate with a sample size claiming to represent the HE sector as a whole, the study does nevertheless seek to juxtapose views from an old university and a new university, including views from quality administrators; to compare reactions from academic chemists and historians, as representatives of science-based and humanities-based disciplines; and to investigate possible differences of perception between experienced and less experienced members of staff.

Within the conceptual framework established by these parameters, the study seeks to test the QAA’s (1999b: para.9) initial finding that members of the academic community take a ‘broadly positive’ view of the benchmarking process.

*The introduction of subject benchmarking*

At the level of the academic discipline, the study detects certain differences between the departments of chemistry and history in the institutions visited. Regarding the introduction of subject benchmarking, there was overall more acceptance of the underlying principle amongst the chemists than amongst the historians. Positive reactions to threshold statements were recorded at approximately equal levels in the two subject areas, though here a significant number of chemists reacted negatively, whereas no historian did.
The differences identified in responses to the benchmarking system are relatively small and so need to be treated with caution. Familiarity with curricular stipulations made by a professional validating body such as the RSC may go some way to explaining the higher level of acceptance of benchmarking amongst chemists, though differences revealed between OldU Chemistry and NewU Chemistry show that the presence or absence of professional body requirements is not, of itself, a reliable indicator of likely responses to subject benchmarking. It was not possible, for instance, to investigate the history and nature of the relationships between the two chemistry departments and the RSC in order to gauge possible sources of differences actually extraneous to the subject benchmarking process or even to the simple familiarity with professional body requirements.

It emerges from the study that institutional history is a further variable to be considered as an influence on the way in which actors respond to the requirements of subject benchmarking. The fieldwork suggests that a background in the polytechnic sector, and in particular experience of the CNAA as a regulator and arbiter of academic quality and standards, is also an important factor shaping attitudes to external control, even at the subject level. This contrasts with the tradition in old universities: ‘an autonomous institution should be free to establish and maintain its own standards of competence without reference to any central authority’ (CHE, 1963: para.713).

Similarly significant in assessing the impact of subject benchmarking and threshold standards is the intrinsic nature of academic disciplines themselves and the
implications this has for cultural differences between them. Trow (1996: 321) may be correct to suggest that education is a process merely pretending to have a measurable outcome, so that all measures of educational outcomes are in fact spurious. Nevertheless, the two subjects considered in this study can be differentiated in terms of their fundamental relationships with their objects of enquiry: whereas scientists like to have something to measure (OC8), historians need to leave much more room for judgment (Hart, 1997). On this view, it is unsurprising that chemists will be more sympathetic than historians to a system offering a form of outcome measurement.

The study suggests, then, that reactions to subject benchmarking do not result from a simple and uniform decision-making process, but are shaped by complex interactions amongst a range of professional, contextual and intellectual variables. Overall, however, amongst academic staff, just 16/30 approved the introduction of subject benchmarking (cf. Table 4.2), and 16/30 gave approval for threshold standards (cf. Table 4.3). If quality administrators are included, the approval ratings are 19/34 and 16/34 respectively. Therefore, the study does not replicate the QAA’s initial finding that the academic community’s view of subject benchmarking is broadly positive.

In the two universities investigated, there seemed to be little difference at the institutional level in the way that subject benchmarking was accommodated within quality management structures: both universities indicated the need for conformity, though OldU seemed able to adopt a more relaxed approach, leaving with academic
departments responsibility for the extent of the compliance. Here, too, institutional history, especially the familiarity with external control associated with NewU’s polytechnic background, may well be a factor contributing to any differential in institutional stance on subject benchmarking. As far as the academic departments are concerned, only NewU Chemistry exhibited manifest conformity with the benchmark statement, though the almost total lack of documentary evidence elsewhere makes it difficult to draw reliable conclusions. As has been seen, however, OldU History was at pains to conform to the subject benchmark statement when drawing up its Self Evaluation Document in preparation for a QAA visit. So great, then, is the perceived need to conform that a state of affairs is produced in which the subject benchmark statement simply states what practitioners of the subject are already doing, but those same practitioners have to go to considerable lengths to demonstrate that what they are doing is in conformity with the benchmark statement. Small wonder, perhaps, that benchmarking is viewed less positively in universities than in other sectors (Holloway and Francis (2002: 251).

Subject benchmarking and improvement

A finding emerging with striking clarity from the fieldwork is that, overall, subject benchmarking was not perceived as exhibiting any of the characteristics normally associated with benchmarking. In particular, only 2/34 answered that benchmarking was leading to quality improvement. Views on the link between subject benchmarking and best practice were more positive, but even here only just over one third of respondents thought that benchmarking would encourage the dissemination of best practice. The nature of subject benchmarking is thus called
into question, since the fieldwork indicates that the tension identified by Jackson (1999) between developmental benchmarking and regulatory benchmarking in fact dissipates, leaving only regulatory benchmarking. As Williams (2004a: 1) notes, the Academic Infrastructure supposes 'the need for common standards.'

Accountability, autonomy and academic identity

According to Brennan and Shah (2000: 131), the first concern of institutional quality management is 'the need to comply with whatever external quality demands are current in the national system.' Although these external demands were frequently identified in the fieldwork as being politically motivated, and as being a widespread feature of life in contemporary Britain, this was not typically interpreted as a lack of trust in the professionalism of the academic community and its ability to control its own activities. Rather, public accountability, whilst accepted as a principle, was criticized as over-burdensome: 'It all comes down to the old saying: You can't increase the weight of a pig by weighing it. [Benchmarking is] taking up lots of time and paper. ... We're spending too much time weighing the pig' (NC4). In the specific case of subject benchmarking, OQS reflected that there was not much evidence that it had actually led to curriculum change at OldU, so that if all that was happening was that people were simply taking their existing courses but 'changing the words a bit about them in order to show some auditor or whoever that they have indeed met the benchmark statement with their curriculum as it is, I think it's hard to see what the value of that is. A lot of effort to change nothing.' It is difficult to disagree with this assessment.
As was indicated in the Introduction, the post-Dearing quality agenda has been interpreted by some as offering a more restrictive view of university autonomy. Associated with this is not infrequently the contention that the professionalism of academics is being eroded by unwarranted external interference in the way in which they function. In the fieldwork, it emerged that subject benchmarking was considered to have almost no effect on the way respondents thought about their own position as academics or about the position of universities in today's society. Indeed, these two items in the interview schedules produced very little by way of reaction at all. It was remarkable, however, that the strongest expressions of unrest tended to come from the younger members of academic staff, even though, overall, they tended to be positive about the introduction of subject benchmarking and threshold standards. This raises interesting questions about the persistence of images of academic status and institutional autonomy which have perhaps been superseded by the realities of a mass system of HE. Time constraints made it impossible to investigate in detail the assumptions and tensions underlying this phenomenon.

Generally speaking, however, reactions to subject benchmarking tended to be somewhat subdued. Not only did the statements themselves not provoke much by way of forceful reactions, either in support of or opposed to the benchmarking regime; some respondents indicated that the benchmarking process had in effect generated nothing new, that the benchmark statements were a 'damp squib' (OC5). In practical terms, subject benchmarking had achieved little beyond setting out existing practice and had changed nothing. Given this, the future value of the
benchmark statements is difficult to discern. As OQS put it: 'Without any further impetus from government, they will just slowly be forgotten about, become just historical documents.'

*Subject benchmarking and national curriculum*

It has been seen that one of the fears voiced at an early stage in the development of subject benchmarking was that it would lead to a national curriculum for UK HE. Whilst many respondents stressed the need for flexibility to be safeguarded in programme design, there was little explicit mention of a national curriculum, though whenever a possible national curriculum for HE was mentioned, it was evaluated negatively. In assessing this, it should not be forgotten that one third of the academics interviewed claimed no familiarity with subject benchmarking and were thus unlikely to have formulated for themselves their own views on potential issues involved in the system. The repeated calls for flexibility indicate a mindset opposed to what would doubtless be seen as the unwarranted constraint represented by any development in the direction of a national curriculum. These calls for flexibility also suggest that clearer opposition to a national curriculum for HE would have emerged if subject benchmarking had been part of a system of Subject Review, as had originally been intended: Subject Review would have subjected all academic disciplines to close scrutiny and so would have necessitated direct engagement with subject benchmarking by each member of a department. Yet even if the shift from Subject Review to Institutional Audit has dulled the perceived threat of a national curriculum, the undercurrent of caution is still detectable. It is interesting that the only respondent to offer an additional comment after checking
an interview transcript observed: 'the benchmarking idea seems to imply a national curriculum in Higher Education which is (in my opinion) against the spirit of academe' (OC9).

**The appositeness of subject benchmarking**

The benchmark statements, it should not be forgotten, resulted from a major developmental effort; what Morley (2003: 45) has referred to as 'the benchmarking industry'. If, then, 'what we've ended up with is: not very much' (OQS), the value of the whole subject benchmarking undertaking is seriously undermined. Even the manner in which the benchmarking process was initiated is seriously flawed. Dearing's (NCIHE, 1997: para.10.55) starting point, for instance, is simply that, with a large increase in student numbers combined with a marked increase in the proportion of First or Upper Second class honours degrees awarded, 'many think that it is not plausible to say that standards have not declined.' To move from this observation to a recommendation (NCIHE, *op.cit.*: Recommendation 25) placing particular emphasis on threshold standards is little short of eccentric. Neither is the QAA's (QAA, 1999a: 7) stipulation helpful that the standards to be established by the benchmarking groups should not be expressed in terms of degree classifications. This dissociation of descriptors of standards from designations of award classifications sidesteps the issue of how to indicate levels of actual achievement. Indeed, several respondents in the fieldwork commented on the difficulty of interpreting 'the attainment thing' (OC6), with its lack of clear reference to degree classifications.
Moreover, the very assumption of declining standards is open to question, even Dearing (NCIHE, 1997: para.10.57) conceding that the Commission did not actually have proof that degree standards had fallen over time. For all that Furedi (2002: 21) laments 'the onward march of grade inflation' and Morley (2003: 6) records 'the moral panic over standards', Harvey (2002: 21) reports the alternative view that higher education itself has changed, meaning that current and past standards cannot be compared: 'Things are different but standards are not worse.' Supporting evidence is even to be found in the QAA's (QAA, 2003d: 4) own report on Subject Review 1993-2001, which records that, after 2 904 Subject Review reports, 'HE provision in England and Northern Ireland was generally found to be of very high quality: the overwhelming majority (99 per cent) of subject review visits resulted in the provision being approved in the first instance.' Even allowing for the fact that Dearing reported at the mid-point of the period covered by the QAA's (op. cit.) overview report, therefore without access to all its findings, it is difficult to understand why the benchmarking exercise was begun with single subject honours degrees rather than, for instance, modular degrees, picking up on Trow's (1994: 42) reminder: 'The incoherence of undergraduate studies in the United States results also from the modularization of courses, and the freedom students have to elect among these modules in their accumulation of credits towards their degree'.

Nevertheless, the QAA (2003d: 4) overview report states: 'A new era of quality assurance, including a stronger emphasis on the maintenance and enhancement of academic standards, has already begun.' On the basis of the fieldwork, it is not easy
to see what constructive contribution subject benchmarking will make to this.

6.2 The significance of the study

Brown (within Gordon, 2002: 208) notes that 'evidence that [quality assessment] is having a serious impact on the quality of teaching and learning remains patchy.' Among the reasons why it is so difficult to explain what prevents quality assessment from effectively enabling professional learning for academic staff is 'because we lack not only micro studies of the impact of quality assessment ... but also because micro studies of local experiences are comparatively rare' (Brown, op.cit.: 208). The same arguments apply in the case of subject benchmarking: the benchmarking study begins to address them.

This comparative empirical study, carried out at the micro level in two parallel pairs of academic departments and in the corresponding institutional quality administrations, considers the impact of subject benchmarking as perceived by those required to apply it in HE. As such, it begins to investigate how members of the academic community interpret the requirements and implications of an innovation representing a significant shift in the way in which academic programmes of study need to be shown to be conceived and managed. Becher and Trowler (2001: 194) describe one of the central purposes of their research as 'to emphasize the existence and to begin to identify the nature of the connections between fields of study and those who work in them'. The present study begins the process of systematic and rigorous investigation not only into reactions to subject benchmarking, but also into possibly differing reactions to subject benchmarking
based on membership of different fields of study, different types of HE institution and different academic generations. By engaging directly with those involved in the operationalization of subject benchmarking, either at the departmental or at the institutional level, the study investigates the meanings attached to this new departure, identifies layers of complexity underlying those meanings, and goes on to suggest areas for further research.

6.3 Limitations of the research

Stanley and Wise (1993: 161) emphasize that 'all research involves, as its basis, an interaction, a relationship, between researcher and researched.' Whilst every effort was made in the fieldwork to establish a relationship of openness and trust, the limited amount of time available made the success of this difficult to assess. The limited access to respondents, in some instances exacerbated by interruptions or the respondent's need to move promptly to a further appointment, also meant that there could not always be the same level of probing, or that some avenues had to be left unexplored. Here, too, the impact of the limitation is difficult to assess.

A further problem in the fieldwork, especially in view of the relatively low level of familiarity with subject benchmarking revealed, was the possibility of a variation on the Hawthorne Effect: 'experimental arrangements may induce an effect over and above the intentions of the investigator' (Bryman, 1989: 89). That is to say that it was possible that, by focusing attention on an item with which the respondent was not familiar, that item may, through the operation of 'reactivity' (Bryman, op.cit.: 65), have taken on a temporary and, in the respondent's terms unwarranted,
importance, thereby distorting responses. OC4, for instance, mentioned that he had looked at the benchmark statement immediately prior to the interview. By contrast, OC8 indicated just before the interview that he was ‘worried’ because he did not really know anything about benchmarking and that he had thought of looking at the QAA document, but then had thought that ‘that was not perhaps the point.’ The hope had, indeed, been to obtain as far as possible a spontaneous, unprepared response to the research questions.

Difficulties of three kinds arose with interview transcripts. The first respondent to look at a written version of an interview had expected features of the spoken language, such as hesitations or false starts, to be edited out, and on receipt of a document including such features, withdrew it. In this instance, a compromise was reached by telephone: a revised ‘edited’ version was submitted and approved, though clarification requested of certain points of detail was, unfortunately, not forthcoming. In all subsequent instances, respondents wishing to check a transcript were also sent a short document identifying differences between spoken and written language.

In another instance, a respondent who had given a very interesting and informative interview reacted in the same way as the first respondent. However, checks on progress with the revised transcript remaining unanswered, and bearing in mind the respondent’s doubts, it was decided to abandon this interview.

The third difficulty arose with three respondents who were interested in seeing a transcript, but who then did not respond using the printed reply form and pre-paid
envelope provided to comment on the document. All further attempts to contact these respondents remaining fruitless, each one was contacted a final time with a message to the effect that, in the absence of any request for alterations, it was assumed that permission was granted to use the data contained in the interview transcripts. This final contact produced no response, and so the three transcripts were adopted.

Morley (2003: xi) stresses that the small sample size underlying her work permits only 'limited, cautious and modest claims.' This limitation is even more acute in the case of the benchmarking study, which, while making no claim to generalizability, nevertheless offers insights into the realities informing perceptions of the introduction and operationalization of a significant innovation in UK higher education. It should also be borne in mind, however, that both OldU and NewU are well-regarded and well-established institutions, NewU having gained a solid reputation for its pre-1992 activities. Attitudes in less mature institutions may differ.

An impediment to the study was the almost total lack of documentation relating to subject benchmarking, essentially because none seemed to exist.

6.4 Subject benchmarking: the way forward

Hargreaves and Christou (2002: 190) assert: 'If the sector does not find a way to value subject benchmarks they will remain a one off, partial experiment with little expectation of continued development.' However, the fieldwork suggests that certain modifications to the existing benchmark statements could begin to improve
their practical usefulness. One would be to adopt a regular format, each statement identifying the same range of standards, from threshold upwards, but also covering the highest levels of achievement. Another would be to link these standards directly to degree classifications, or to whatever system might replace degree classifications (Elton: 2004; Morley et al., 2004; VandeLinde, 2002).

In fact, the nature of the problem may be not so much that the sector undervalues subject benchmarking, but rather that the Academic Infrastructure overvalues it. If, as has been suggested, regulatory benchmarking is essentially regulation, and if benchmarks are 'reference points' (Wright and Williams, 2001), then the benchmarking statements come to take on the air of guidelines or, in QAA terms, precepts. It is surely no accident that the subject benchmark statement for chemistry is prefaced by a letter from the chairman of the benchmarking group in which the document is described as 'the General Guidelines for the Academic Review of Bachelors Honours Degree Programmes in Chemistry' (QAA, 2000c: 1). In other words, the appropriate status for subject benchmarking would appear to be as an appendage to the QAA's (2000e) Code of practice for the assurance of academic quality and standards in higher education. Section 7: Programme approval, monitoring and review. Yet even this is not totally unproblematic, as 'there is still some uncertainty, or even confusion, about the nature and status of the Code' (QAA, 2004: 2), with perceptions that the Code of Practice is 'a series of prescriptive rules or regulations' (QAA, op.cit.: 2). Additionally, referring to the precept in the QAA's (2000a: 6) Code of practice requiring external examiners to refer to subject benchmarks when commenting on standards, Brown (2004: 153)
points out that this could turn into an indirect route to a national curriculum, even if
this currently seems unlikely. Nevertheless, by incorporating the statements into the
Code of practice they would at least be clearly situated in what may be perceived as
the most appropriate context for them.

A new dimension to the discussion surrounding subject benchmarking and a
national curriculum is potentially introduced by the proposal in the White Paper
The future of higher education (DfES, 2003: para.4.33) to award the university title
on the basis of taught degree awarding powers. Whilst it is stressed (DfES, op.cit.:
para.4.34) that this will not make it easier to achieve degree awarding powers,
Thomson (2003: 1) reminds that it will be necessary to maintain degree standards
over time. Should the proposal be implemented, it remains to be seen how these
degree standards would be set and assessed, and whether the advent of teaching-
only universities might finally open the door to a national curriculum, possibly via
an augmented benchmarking system.

Similarly, the introduction of top-up fees in universities and the associated
‘prospect of a sweeping review of the entire accountability framework’ (Baty,
2004: 6) raised by Alan Johnson, the Higher Education Minister, in June 2004 may
lead to a return to greater external scrutiny at subject level. This, in turn, may
rekindle fears of the progressive undermining of professionalism, if Elton (2004:
421) is correct in identifying ‘large governmental pressures to see students as the
customers of higher education in a system where surveillance and performance
indicators are replacing trust’. The incoming Chairman of the QAA (Younger,
warns of the need ‘to protect the [HE] sector from the demands for statutory external regulation’.

The key to a dynamic and constructive future for subject benchmarking may be for ownership of it to be claimed by the academic profession. Holloway and Francis (2002: 249) argue that, ‘the perceived value of any sort of benchmarking might have been irreparably damaged by its association with costly and cumbersome audit process being imposed from “above”.’ Even the QAA (2003d: 56) acknowledges that ‘long-term curriculum leadership in the subject areas’ was to come from the LTSN. Rather than constituting part of an external control mechanism, subject benchmarking could be integrated into the dynamic process of academic development by responsibility for it passing into the domain of the Higher Education Academy, which now subsumes the LTSN (Ramsden, 2004: 1). As explained by Wagner (2004), Academy Chair: ‘The purpose of the Academy is to enhance the student experience and raise the status and standing of teaching and learning in higher education by enabling staff to operate to the highest professional standards.’

Perhaps above all, if the designation ‘subject benchmarking’ is to be maintained, efforts should be made to exploit the enhancement potential of benchmarking. Reference has already been made to the challenge of modularization: it is perhaps ironic that the QAA’s (1999e: para.6.2) own advisory group on multidisciplinary and modular programmes should state: ‘The Advisory Group ... recommends that the standards of higher education in the UK could be better assured through the
adoption of an enhancement rather than referent benchmarking process and philosophy.

6.5 **Future research**

The present study suggests the interest of taking the work further in a project seeking to be both representative and diachronic, monitoring changing perceptions over time.

According to Holloway and Francis (2002: 246): ‘the benefits [of subject benchmarking] to stakeholders other than funding bodies remain relatively unclear.’ The constraints of the present study made it impossible to include students; neither does the National Union of Students (NUS: e-mail) have a policy on subject benchmarking. Future research could profitably investigate student perceptions of subject benchmarking. Indeed, Hill *et al.* (2003) indicate that student perceptions of quality generally are under-researched.

Similarly valuable would be work to ascertain whether or not massification is leading to declining standards, and so whether or not the starting point for the subject benchmarking system is valid.

Underlying the normal understanding of benchmarking is the quest for improvement; but according to Morley (2003: 13): ‘the concept of improvement in quality assurance can often seem under-theorized in the context of sophisticated analyses in the academy of measurement techniques.’ However, the QAA (2003b) in Scotland has adopted a model of enhancement-led institutional review (ELIR), in
which enhancement is defined as ‘taking deliberate steps to bring about continuous improvement in the effectiveness of the learning experience of students’ (QAA, op.cit.: para.16). Subject benchmarking information is listed as one of the common reference points most likely to be included in ELIR (QAA, op.cit.: para.19). These considerations suggest two valuable areas of research: on the one hand, the development of a theoretical model of quality enhancement; and on the other hand a comparative study of perceptions of subject benchmarking in England and Scotland.

Beyond the UK, work should be done to evaluate the status of subject benchmarking within the Bologna process (Williams, 2004c: 10).

Freed et al. (1997: 137) ask: ‘Even though it is hard to define, who can be against quality?’ Quality in HE is complex, fraught with ambiguity, and contested at every turn. So imbricated is subject benchmarking with the convoluted structures, value systems, power relationships and sensitivities contributing to the construction of the idea of quality that the researcher may be forgiven for exclaiming with Melville (1992: 157): ‘This whole book is but a draught – nay, but the draught of a draught. Oh, Time, Strength, Cash, and Patience!’
Appendix 1: Interview schedule used with academic staff

**Academic Staff**

Institution: ___________________________  Department: ___________________________

Name: ________________________________

Date: ___________________________  Location: ___________________________

Time: ___________________________  Identifier: ___________________________

**Statement of informed consent**

☐

**Statement of purpose**

☐

Do you mind if I record our conversation?

Yes ☐  No ☐

Would you mind telling me how long you have been teaching in higher education?

________________________________________________________________________

Would you mind telling me how long you have been in this department?

________________________________________________________________________

________________________________________________________________________

1. Could you tell me a little about your experience to date of quality assurance and enhancement?

2. Are you used to working with criteria set by an external agency, such as a professional body?

   yes  1  ☐  no  2  ☐  other  3  ☐

3. How do you feel about the introduction of benchmarking?

   Prompt  Is it  very good  4  ☐  good  5  ☐  neither good nor bad  6  ☐  bad  7  ☐  very bad  8  ☐  other  ☐
4. How do you react to the

*(Chemistry)* performance criteria and statement of threshold performance

*(History)* statement of the threshold standard

established for single honours degrees in Chemistry/History?

| Prompt 1 | Is it | very good | 4 |
|          |      | good      | 5 |
|          |      | neither good nor bad | 6 |
|          |      | bad       | 7 |
|          |      | very bad  | 8 |
|          |      | other ?   | 3 |

| Prompt 2 | Copy of text/s. |

5. Would you say that benchmarking has led you to change your approach to your courses?

| yes       | 1 |
| no        | 2 |
| other     | 3 |

| Prompt | Course design | 9 |
|        | content       | 10 |
|        | delivery      | 11 |
|        | other         | 3 |

6. Would you say that benchmarking has had an impact on your professional life?

| yes       | 1 |
| no        | 2 |
| other     | 3 |

7. In your view, is benchmarking leading to improvements in quality?

| yes       | 1 |
| no        | 2 |
| other     | 3 |

*how?*

*why not?*

*what would?*

8. Do you think that benchmarking will encourage the dissemination of best practice?

| yes       | 1 |
| no        | 2 |
| other     | 3 |

*how?*

*why not?*
9. In your view, was it necessary to introduce benchmarking?

<table>
<thead>
<tr>
<th>Option</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>yes</td>
<td>1</td>
</tr>
<tr>
<td>no</td>
<td>2</td>
</tr>
<tr>
<td>other</td>
<td>3</td>
</tr>
</tbody>
</table>

10. In your view, why was benchmarking introduced?

11. Has benchmarking led to changes in the way you think about your own position as an academic?

<table>
<thead>
<tr>
<th>Option</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>yes</td>
<td>1</td>
</tr>
<tr>
<td>no</td>
<td>2</td>
</tr>
<tr>
<td>other</td>
<td>3</td>
</tr>
</tbody>
</table>

12. Has benchmarking led to changes in the way you think about the position of universities in today’s society?

<table>
<thead>
<tr>
<th>Option</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>yes</td>
<td>1</td>
</tr>
<tr>
<td>no</td>
<td>2</td>
</tr>
<tr>
<td>other</td>
<td>3</td>
</tr>
</tbody>
</table>

13. Is there anything you would like to add concerning benchmarking or threshold standards? [eg. Impact of new QAA methodology?]

<table>
<thead>
<tr>
<th>Option</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>yes</td>
<td>1</td>
</tr>
<tr>
<td>no</td>
<td>2</td>
</tr>
<tr>
<td>other</td>
<td>3</td>
</tr>
</tbody>
</table>

Many thanks for agreeing to take part in this project.

Would you like to check a transcript of our conversation?

<table>
<thead>
<tr>
<th>Option</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>☑</td>
</tr>
<tr>
<td>No</td>
<td>☑</td>
</tr>
</tbody>
</table>
Chemistry
Subject benchmark statements

Subject benchmark statements provide a means for the academic community to describe the nature and characteristics of programmes in a specific subject. They also represent general expectations about the standards for the award of qualifications at a given level and articulate the attributes and capabilities that those possessing such qualifications should be able to demonstrate.

This subject benchmark statement, together with the others published concurrently, refers to the bachelors degree with honours.

Subject benchmark statements are used for a variety of purposes. Primarily, they are an important external source of reference for higher education institutions when new programmes are being designed and developed in a subject area. They provide general guidance for articulating the learning outcomes associated with the programme but are not a specification of a detailed curriculum in the subject. Benchmark statements provide for variety and flexibility in the design of programmes and encourage innovation within an agreed overall framework.

Subject benchmark statements also provide support to institutions in pursuit of internal quality assurance. They enable the learning outcomes specified for a particular programme to be reviewed and evaluated against agreed general expectations about standards.

Finally, subject benchmark statements are one of a number of external sources of information that are drawn upon for the purposes of academic review* and for making judgements about threshold standards being met. Reviewers do not use subject benchmark statements as a crude checklist for these purposes however. Rather, they are used in conjunction with the relevant programme specifications, the institution's own internal evaluation documentation, together with primary data in order to enable reviewers to come to a rounded judgement based on a broad range of evidence.

The benchmarking of academic standards for this subject area has been undertaken by a group of subject specialists drawn from and acting on behalf of the subject community. The group's work was facilitated by the Quality Assurance Agency for Higher Education, which publishes and distributes this statement and other benchmarking statements developed by similar subject-specific groups.

The statement represents the first attempt to make explicit the general academic characteristics and standards of an honours degree in this subject area, in the UK.

In due course, but not before July 2003, the statement will be revised to reflect developments in the subject and the experiences of Institutions and academic reviewers who are working with it. The Agency will initiate revision and, in collaboration with the subject community, will establish a group to consider and make any necessary modifications to the statement.

This statement is © The Quality Assurance Agency for Higher Education 2000.

It may be reproduced by educational Institutions solely for educational purposes, without permission. Excerpts may be reproduced for the purpose of research, private study, or review without permission, provided full acknowledgement is given to the subject benchmarking group for this subject area and to the copyright of the Quality Assurance Agency for Higher Education.

Electronic storage, adaptation or translation of the statement is prohibited without prior written agreement from the Quality Assurance Agency for Higher Education.

* academic review in this context refers to the Agency's new arrangements for external assurance of quality and standards. Further information regarding these may be found in the Handbook for Academic Review, which can be found on the Agency's web site.

ISBN 1 85824 482 X
AR 002 4/2000

© Quality Assurance Agency for Higher Education 2000

Published by
Quality Assurance Agency for Higher Education
Southgate House
Southgate Street
Gloucester GL1 1UB
Tel 01452 557000
Fax 01452 557070
Web www.qaa.ac.uk

Text printed by
Kall Kwik, Gloucester

The Quality Assurance Agency for Higher Education is a company limited by guarantee
Dear Mr Randall

I have pleasure in forwarding to you the General Guidelines for the Academic Review of Bachelors Honours Degree Programmes in Chemistry, which constitutes the outcome of the work of the Agency's Chemistry Benchmarking Group.

As you will be aware, the Guidelines received wide circulation in draft form. The draft was circulated to universities and colleges offering degree courses in chemistry, to trade associations and a selection of large and small chemical and pharmaceutical companies, and to professional bodies and a number of selected individuals in various occupations. The response was encouraging both in terms of the exceptionally high response rate and in the positive tenor of comments and constructive suggestions for change that we received and to which we endeavoured to do justice in finalising the document.

We are conscious that these Guidelines are among the first of many which the Agency will be issuing. As such, they must be viewed as experimental and will no doubt need modification in the light of the trialling to be undertaken during 1999. I and the other members of the Group look forward to these developments with interest.

As required by the Agency, the Group has prepared benchmarking guidelines for BSc Honours courses in chemistry. As well as applying to broadly based specialist programmes in chemistry, we consider that they are relevant to programmes in specialised applications of chemistry (for example, medicinal chemistry, analytical chemistry, etc) and of value for courses in which chemistry constitutes the major study (which are often entitled "Chemistry with..."). The fact that the Agency is addressing the problem of benchmarking joint honours and multidisciplinary studies is well known; the outcome will have implications for a further range of chemistry provision.

In our judgement, confirmed by the consultation process, priority as far as the preparation of further guidelines of direct relevance to chemistry is concerned, should be given to those for MChem/MSc Chemistry courses (Dearing's "higher honours" courses). These courses are now beginning to enrol significant numbers of students. In some, albeit a minority of departments they are becoming the main undergraduate provision; hence the need for the production of these guidelines at the earliest opportunity.

Yours sincerely

Professor E W Abel CBE CChem FRSC
Chairman, Chemistry Benchmarking Group
Academic standards - Chemistry

Introduction

In its consultation paper entitled Developing the Quality Assurance and Standards Framework for UK Higher Education, the Quality Assurance Agency for Higher Education (QAA) advocates the development of 'benchmark Information on subject threshold standards' which articulates the abilities and skills expected of bachelors honours graduates in different subjects.

The purpose of such information is to assist:
- higher education Institutions in designing and approving programmes of study;
- external examiners and academic reviewers in verifying and comparing standards;
- where appropriate, professional bodies in their accreditation and review of programmes relating to professional competence;
- students and employers when seeking information about higher education provision.

This particular document sets out proposals for the benchmark Information in chemistry. It focuses on four major aspects concerning programmes leading to bachelors honours degrees:

1. The major aims and purposes that may be associated with bachelors honours degree programmes in chemistry.
2. The essential subject-matter components that may be expected to be covered in all study programmes leading to such degrees.
3. The abilities, competencies and skills to be developed in students through the study of chemistry at bachelors honours degree level.
4. Recommendations concerning procedures appropriate for the assessment of the knowledge, abilities and skills set out above and the criteria for different levels of attainment.

The specifications and criteria concerning the foregoing points, set out in the following sections of this document, are intended to provide a broad framework within which course providers can develop purposeful and challenging chemistry programmes that respond to the needs of their students, as well as to the changing nature of chemistry. Their purpose is not to impose on institutions a set of rigid conditions that would stifle innovation in programme development and in the design of learning experiences. However, in so far as they seek to articulate the key qualities expected of bachelors honours chemistry students at the end of their degree programme, it is hoped that they will make a valuable contribution to the definition of 'graduateness' in chemistry and the maintenance of the standard of chemistry degrees.

Details of the aims, objectives and content of Individual programmes will be found in the programme specifications or other documentation issued by higher education Institutions.

Main aims of degree programmes in chemistry

The main aims of bachelors honours degree programmes in chemistry should be:
- To instil in students a sense of enthusiasm for chemistry, an appreciation of its application in different contexts and to involve them in an intellectually stimulating and satisfying experience of learning and studying.
- To provide students with a broad and balanced foundation of chemical knowledge and practical skills.
- To develop in students the ability to apply their chemical knowledge and skills to the solution of theoretical and practical problems in chemistry.
- To develop in students, through an education in chemistry, a range of transferable skills, of value in chemical and non-chemical employment.
- To provide students with a knowledge and skills base from which they can proceed to further studies in specialised areas of chemistry or multi-disciplinary areas involving chemistry.
- To generate in students an appreciation of the importance of chemistry in an Industrial, economic, environmental and social context.
**Subject knowledge**

Each institution providing bachelors honours degree programmes in chemistry is free to decide on the content, nature and organisation of its courses or modules. Therefore, chemistry degree programmes offered by individual institutions will have their own particular characteristics. While it is acknowledged that the depth in which individual aspects are treated may vary with the nature of specific chemistry programmes, it is expected that all programmes will ensure that students become conversant with the following main aspects of chemistry.

- Major aspects of chemical terminology, nomenclature, conventions and units.
- The major types of chemical reaction and the main characteristics associated with them.
- The principles and procedures used in chemical analysis and the characterisation of chemical compounds.
- The characteristics of the different states of matter and the theories used to describe them.
- The principles of quantum mechanics and their application to the description of the structure and properties of atoms and molecules.
- The principles of thermodynamics and their applications to chemistry.
- The kinetics of chemical change, including catalysis; the mechanistic interpretation of chemical reactions.
- The principal techniques of structural investigations, including spectroscopy.
- The characteristic properties of elements and their compounds, including group relationships and trends within the Periodic Table.
- The properties of aliphatic, aromatic, heterocyclic and organometallic compounds.
- The nature and behaviour of functional groups in organic molecules.
- The structural features of chemical elements and their compounds, including stereochemistry.
- The relation between bulk properties and the properties of individual atoms and molecules, including macromolecules.
- Awareness of major issues currently at the frontiers of chemical research and development.

**Abilities and skills**

At bachelors honours level, students are expected to develop a wide range of different abilities and skills. These may be divided into three broad categories, viz:

a. Chemistry-related cognitive abilities and skills, i.e., abilities and skills relating to Intellectual tasks, including problem solving;
b. Chemistry-related practical skills, e.g., skills relating to the conduct of laboratory work;
c. Transferable skills that may be developed in the context of chemistry and are of a general nature and applicable in many other contexts.

The main abilities and skills that students are expected to have developed by the end of their bachelors honours degree programme in chemistry, are as follows.

a. Chemistry-related cognitive abilities and skills

- Ability to demonstrate knowledge and understanding of essential facts, concepts, principles and theories relating to the subject areas identified above.
- Ability to apply such knowledge and understanding to the solution of qualitative and quantitative problems of a familiar and unfamiliar nature.
- Ability to recognise and analyse novel problems and plan strategies for their solution.
- Skills in the evaluation, Interpretation and synthesis of chemical information and data.
- Ability to recognise and implement good measurement science and practice.
- Skills in presenting scientific material and arguments clearly and correctly, in writing and orally, to a range of audiences.
- Computational and data-processing skills, relating to chemical information and data.
b. Chemistry-related practical skills

- Skills in the safe handling of chemical materials, taking into account their physical and chemical properties, including any specific hazards associated with their use.
- Skills required for the conduct of standard laboratory procedures involved in synthetic and analytical work, in relation to both inorganic and organic systems.
- Skills in the monitoring, by observation and measurement, of chemical properties, events or changes, and the systematic and reliable recording and documentation thereof.
- Competence in the planning, design and execution of practical investigations, from the problem-recognition stage through to the evaluation and appraisal of results and findings; this to include the ability to select appropriate techniques and procedures.
- Skills in the operation of standard chemical instrumentation such as that used for structural investigations and separation.
- Ability to interpret data derived from laboratory observations and measurements in terms of their significance and the theory underlying them.
- Ability to conduct risk assessments concerning the use of chemical substances and laboratory procedures.

c. Transferable skills

- Communication skills, covering both written and oral communication.
- Problem-solving skills, relating to qualitative and quantitative information, extending to situations where evaluations have to be made on the basis of limited information.
- Numeracy and computational skills, including such aspects as error analysis, order-of-magnitude estimations, correct use of units and modes of data presentation.
- Information-retrieval skills, in relation to primary and secondary information sources, including information retrieval through on-line computer searches.
- Information-technology skills such as word-processing and spreadsheet use, data-logging and storage, Internet communication, etc.
- Interpersonal skills, relating to the ability to interact with other people and to engage in team-working.
- Time-management and organisational skills, as evidenced by the ability to plan and implement efficient and effective modes of working.
- Study skills needed for continuing professional development.

Assessment procedures and performance criteria

a. Assessment procedures

It is essential that the procedures used for the assessment of students' achievement in chemistry should correspond to the knowledge, abilities and skills that are to be developed through their degree programme.

Evidence on which the assessment of student achievement is based, should include:

- Formal examinations, including a significant proportion of 'unseen' examinations
- Laboratory reports
- Problem-solving exercises
- Oral presentations
- Planning, conduct and reporting of project work.

Additional evidence of use for the assessment of student achievement may be derived from:

- Essay assignments
- Portfolios on chemical activities undertaken
- Literature surveys and evaluations
- Collaborative project work
- Preparation and displays of 'posters' reporting project work
- Reports on external placements (where appropriate).
b. Performance criteria
Although all students graduating at bachelors honours level in chemistry are expected to demonstrate that they have acquired knowledge, abilities and skills in the areas identified in the foregoing sections, it is accepted that there will be significant differences in their attainment. The following criteria are suggested as indicators of different levels of attainment in these performance areas. These attainment levels are not meant to reflect the traditional degree classifications.

Attainment Level A (highest):
- Knowledge base is extensive and extends well beyond the work covered in the programme. Conceptual understanding is outstanding.
- Problems of a familiar and unfamiliar nature are solved with efficiency and accuracy; problem-solving procedures are adjusted to the nature of the problem.
- Experimental skills are exemplary and show a thorough analysis and appraisal of experimental results, with appropriate suggestions for improvements.
- Performance in transferable skills is generally very good.

Attainment Level B:
- Knowledge base covers all essential aspects of subject matter dealt with in the programme and shows some evidence of enquiry beyond this. Conceptual understanding is good.
- Problems of a familiar and unfamiliar nature are solved in a logical manner; solutions are generally correct or acceptable.
- Experimental work is carried out in a reliable and efficient manner.
- Performance in transferable skills is sound and shows no significant deficiencies.

Attainment Level C:
- Knowledge base is sound, but is largely confined to the content of the programme. Level of conceptual understanding is generally sound.
- Problem-solving ability is sound in relation to problems of a familiar type or those that can be tackled through the straightforward application of standard procedures and/or algorithms.
- Experimental work is generally satisfactory and reliable.
- Performance in transferable skills is largely sound.

Attainment Level D:
- Knowledge and understanding of the content covered in the course are basic.
- Problems of a routine nature are generally adequately solved.
- Standard laboratory experiments are usually carried out with reasonable success though significance and limitations of experimental data and/or observations may not be fully recognised.
- Transferable skills are at a basic level.

Attainment Level E (lowest)
- Knowledge base is acceptable in relation to some of the content covered in the programme.
- Problem-solving ability extends to simple 'standard' problems, following routine procedures.
- Experimental skills are rudimentary.
- Transferable skills are rudimentary.

c. Threshold performance for bachelors honours degrees
Students who are awarded a bachelors honours degree in Chemistry are expected to demonstrate knowledge, abilities and skills corresponding on balance to at least attainment level D.
**Chemistry benchmarking group membership**

Professor E Abel (Chair)  University of Exeter  
Professor P Atkins  Lincoln College, University of Oxford  
Professor I Haines  University of North London  
Professor R Jones  The Open University  
Professor R Kempa  University of Keele  
Professor M Page  University of Huddersfield  
Professor B Parsons  NEWI  
Professor D Phillips  Imperial College, London  
Professor D Rice  University of Reading  
Professor K Smith  University of Wales, Swansea  
Professor A Townshend  University of Hull  
Professor P Tasker  
Professor J Winfield  University of Glasgow  

Dr S Gruber (ex-officio)  The Royal Society of Chemistry
Appendix 3: QAA (2000d) Subject benchmark statement for history

History
Subject benchmark statements

Subject benchmark statements provide a means for the academic community to describe the nature and characteristics of programmes in a specific subject. They also represent general expectations about the standards for the award of qualifications at a given level and articulate the attributes and capabilities that those possessing such qualifications should be able to demonstrate.

This subject benchmark statement, together with the others published concurrently, refers to the bachelors degree with honours.

Subject benchmark statements are used for a variety of purposes. Primarily, they are an important external source of reference for higher education institutions when new programmes are being designed and developed in a subject area. They provide general guidance for articulating the learning outcomes associated with the programme but are not a specification of a detailed curriculum in the subject. Benchmark statements provide for variety and flexibility in the design of programmes and encourage Innovation within an agreed overall framework.

Subject benchmark statements also provide support to Institutions in pursuit of internal quality assurance. They enable the learning outcomes specified for a particular programme to be reviewed and evaluated against agreed general expectations about standards.

Finally, subject benchmark statements are one of a number of external sources of Information that are drawn upon for the purposes of academic review and for making judgements about threshold standards being met. Reviewers do not use subject benchmark statements as a crude checklist for these purposes however. Rather, they are used in conjunction with the relevant programme specifications, the Institution's own Internal evaluation documentation, together with primary data in order to enable reviewers to come to a rounded judgement based on a broad range of evidence.

The benchmarking of academic standards for this subject area has been undertaken by a group of subject specialists drawn from and acting on behalf of the subject community. The group's work was facilitated by the Quality Assurance Agency for Higher Education, which publishes and distributes this statement and other benchmarking statements developed by similar subject-specific groups.

The statement represents the first attempt to make explicit the general academic characteristics and standards of an honours degree in this subject area, in the UK.

In due course, but not before July 2003, the statement will be revised to reflect developments in the subject and the experiences of Institutions and academic reviewers who are working with it. The Agency will initiate revision and, in collaboration with the subject community, will establish a group to consider and make any necessary modifications to the statement.

This statement is © The Quality Assurance Agency for Higher Education 2000.

It may be reproduced by educational Institutions solely for educational purposes, without permission. Excerpts may be reproduced for the purpose of research, private study, or review without permission, provided full acknowledgement is given to the subject benchmarking group for this subject area and to the copyright of the Quality Assurance Agency for Higher Education.

Electronic storage, adaptation or translation of the statement is prohibited without prior written agreement from the Quality Assurance Agency for Higher Education.

* academic review in this context refers to the Agency's new arrangements for external assurance of quality and standards. Further information regarding these may be found in the Handbook for Academic Review, which can be found on the Agency's web site.
Academic standards - History

Initial statement

Introduction

1. In the first instance, the Quality Assurance Agency for Higher Education defined the task of benchmarking groups as producing "broad statements which represent general expectations about standards, particularly at the threshold level, for the award of honours degrees in the field." Following discussions with the Agency, it was agreed that the History Group should focus upon providing a framework for judging programmes which set an acceptable level in principle attainable by the typical History student. We have not seen it as our task to provide a basis for judgements to be made about a particular student's learning achievement, though we do provide an Appendix which makes reference to judgements in this respect. Our assumption has been that academic reviewers will be making judgements about a department's capacity to deliver on standards.

2. We use the term programme to refer to courses of study leading to a degree award. We use the term course to refer to modules or units within a degree programme. Our present concern is with the criteria relating to the work of students completing honours degree programmes (level H4), not with those relating to earlier stages and exit points (H1, H2). However, the criteria relating to programmes are relevant to every stage of progression to the honours degree, from the first year onwards. We recognise that provision in the subject is very extensive through single, combined honours and interdisciplinary programmes. Our principal concern at this stage is with single honours programmes leading to an award in History but our recommendations often relate more broadly. We imagine it will be seen as sensible, in programme approval and review of joint degrees, to take notice of the general tenor of the Benchmarking Statements for the two subjects concerned. We note that the QAA has established an advisory group to consider the implications for benchmarking academic standards of modular and multidisciplinary programmes. Pending further work in this field we hope that our statement will be of value to all those concerned with History in universities and colleges. We accept that organisational patterns vary across the sector. Where we refer to departments, this is a shorthand for history subject groups however organised.

Guiding assumptions

3. History differs from many subjects in that we do not recognise a specific body of required knowledge nor a core with surrounding options. We take it as self-evident that knowledge and understanding of the human past is of incalculable value both to the individual and to society at large, and that the first object of education in History is to enable this to be acquired. We accept variation in how the vast body of knowledge which constitutes the subject is tackled at undergraduate degree level. This entails an approach which concentrates on using knowledge in order to develop certain skills and qualities of mind. The focus in this Statement is on how knowledge is used to acquire these skills and qualities. The form of our argument follows from this. Throughout our work we have been guided by the belief that we should refer to everything that is crucial and integral to the issue of standards. In other words, we take the view that it will not be possible for academic reviewers to make judgements about academic standards in History without some consideration of every aspect of a degree programme which we consider here. Moreover, we believe that departmental statements about the framework of programmes, if they are to be properly useful to both staff and students, will need to cover all the ground that we have covered here. Our final paragraph attempts to summarise the standards which should be sought and achieved by the typical student in key areas of the discipline.

4. We have seen our task as the following: to lay out criteria for judging the suitability and adequacy of single-honours degree courses in History; to do this in a way that is as specific as possible without undermining the principle that there are many different suitable and adequate ways of constructing and making available the great richness and diversity of History; to do it in a way that recognises also the need for adaptability to new academic developments in the field, and innovations in course structures and teaching methods. We insist that teaching and learning are evolving processes and that it not our intention to freeze the teaching of History in a particular model. Our benchmarking statement should be seen as a starting point: departments and subject groups will have the chance to demonstrate how benchmarking standards can be built on by the provision of additional or perhaps alternative opportunities.

5. We have taken full account of the particular characteristics of History as a discipline. Its subject matter, distinguishing it from other humanities and social sciences, consists of the attempts of human beings in the
past to organise life materially and conceptually, individually and collectively, while the object of studying
these things is to widen students' experience and develop qualities of perception and judgement. History
provides a distinctive education by providing a sense of the past, an awareness of the development of
differing values, systems and societies and the inculcation of critical yet tolerant personal attitudes. History's
reciprocal relationship with other disciplines can have an important influence on the experience of the
student of the subject.

6. We recognise that the concepts, theories and methodologies of the social sciences are themselves used by
many historians, most obviously but by no means exclusively within courses in economic and social history.
There has been a long and important tradition in the United Kingdom of teaching and writing history within
a social science framework, which continues both within distinct degree programmes in economic and social
history and as an important feature in many degree programmes in History. We recognise that where history
is taught within the context of the social sciences, students will need to devote considerable time to acquiring
a knowledge of one or more social science. In general, students of all types of history - cultural and political as
well as economic and social - should have an awareness of relevant and appropriate concepts and theories.

7. We are convinced that particular types of skill, quality, and accomplishment are not connected solely to
particular types of course provision or subject matter. Just as there is no one model for a programme, there
will be no one model for the relationship between course provision and students' attainments. We reject any
idea of mechanical progression in History. Skills and qualities are acquired cumulatively and iteratively.

8. Good undergraduate history teaching takes a variety of forms, and programmes quite legitimately
combine different teaching methods in a number of ways. This variety arises from the different interests and
abilities of individual scholars, from the requirements of different areas of the field of history, and from the
fact that departments or subject groups in different institutions have access to different combinations of
teaching resources. Each programme should define its own desired outcomes in ways that command general
credibility, and departments. In designing their teaching to fulfill those outcomes, should recognise the need
to assure their standards by means of the professional external scrutiny provided by institutional peer
review, external examiners and the QAA's scheme of academic review.

9. Assessment is a critical element in the educational process and an essential element in effective
learning. We therefore recommend that all departments should develop a clear assessment policy which is
consistent with the learning outcomes of its degree schemes. They should specify clearly in their
documentation what students are expected to learn, how their work will be assessed, and the relationship
between the two. At the same time, we believe departments should think carefully about and explain the
relationship between the functions of formative and summative assessment, ie between assessment designed
as feedback on progress and assessment for degree award and classification purposes.

10. Important abilities and qualities of mind are acquired through the study of History. They are
particularly valuable for the graduate as citizen and are readily transferable to many occupations and
careers. Some of these qualities and abilities are generic, in that they are imparted by most degree
programmes in the Humanities and Social Sciences. But degree-level study in History also instils ways of
thinking which are intrinsic to the discipline while being no less transferable. These include a respect for
historical context and evidence, a greater awareness of the historical processes unfolding in our own time,
and a deeper understanding of the varied traditions current today. These qualities of mind and abilities are
most effectively and economically developed by deep and prolonged immersion in, and engagement with
the practice, methods and material of the subject itself. The cumulative acquisition of, and ability to apply
transferable skills, and the development of students as competent historians thus necessarily proceed hand-
in-hand. The link between the two lies ultimately in the habits of mind or intellectual approach developed
by students who have been trained as capable practising historians. These will continue to inform the
application of their minds to other matters later in life.

11. We take it as axiomatic that students must progress and that well-designed programmes facilitate their
progression. History programmes do not impart knowledge and skills to be passively absorbed: reading,
discussion and writing, and engagement, exploration and discovery are essential. But we stress the
importance of historical knowledge. The historian's skills and qualities of mind are developed through the
processes of acquiring, evaluating and discussing historical knowledge in the courses and the independent
study that History degree programmes demand. Although we prescribe no particular diet of historical
knowledge, programmes need to impart such knowledge and also to encourage students to acquire more.
The learning outcomes of a History degree programme have to be seen in terms of particular pieces of
student work - either written or spoken - in which crucial tests are understanding texts and marshalling an
argument. Accordingly, the ability to deploy ideas and information, to show conceptual grasp and to shape
argument becomes difficult to separate in assessment practice from the ability to display appropriately
relevant, wide and diverse historical knowledge.
The historian's skills and qualities of mind

12. We recommend that History degree students should undertake programmes which foster and inculcate the following skills and qualities:

I. The ability to understand how people have existed, acted and thought in the always different context of the past. History involves the cultural shock of encountering and sensing the past's otherness and of learning to understand unfamiliar structures, cultures and belief systems. These forms of understanding also shed important light on the influence which the past has on the present.

II. The ability to read and use texts and other source materials, both critically and empathetically, while addressing questions of genre, content, perspective and purpose.

III. The appreciation of the complexity and diversity of situations, events and past mentalities. This emphasis is central to History's character as an anti-reductionist discipline fostering intellectual maturity.

IV. The understanding of the problems inherent in the historical record itself: awareness of a range of viewpoints and the way to cope with this; appreciation of the range of problems involved in the interpretation of complex, ambiguous, conflicting and often incomplete material; a feeling for the limitations of knowledge and the dangers of simplistic explanations.

V. Basic critical skills: a recognition that statements are not all of equal validity, that there are ways of testing them, and that historians operate by rules of evidence which, though themselves subject to critical evaluation, are also a component of intellectual integrity and maturity.

VI. Intellectual independence: a History programme is not simply or even primarily a preparation for research in the subject, but it should incorporate the general skills of the researcher, namely the ability to set tasks and solve problems. This involves: bibliographic skills; the ability to gather, sift, select, organise and synthesise large quantities of evidence; the ability to formulate appropriate questions and to provide answers to them using valid and relevant evidence and argument. It should develop reflexivity, i.e. an understanding of the nature of the discipline including what questions are asked by historians, and why.

VII. Marshalling of argument - in written and oral form drawing on and presenting all the above skills. Such argument should have structure: it should be relevant and concise. In the case of written argument it should be expressed in clear, lucid and coherent prose. Orally it should involve the capacity to sustain a reasoned line of argument in the face of others, to listen, to engage in sustained debate, and amend views as necessary in the light of evidence and argument.

13. Some programmes, e.g. in economic and social history, incorporate the methodologies of other Humanities and Social Science disciplines. A number of specific skills are thus essential to particular types of programme, and desirable though not obligatory in others. Departments or institutions are strongly recommended to make provision, where appropriate, for the development of at least one of these: the use of C and IT in learning or analysis; numeracy and quantitative methods; archaeological fieldwork; archival study; or skills associated with the study of other disciplines with which History has close links. We also note the capacity of overseas exchanges or study-abroad programmes to enrich students' intellectual and personal development.

14. The generic skills acquired through the study of History are:

I. Self-discipline;

II. Self-direction;

III. Independence of mind, and Initiative;

IV. Ability to work with others, and have respect for others' reasoned views;

V. Ability to gather, organise and deploy evidence, data and information; and familiarity with appropriate means of identifying, finding, retrieving, sorting and exchanging information;

VI. Analytical ability, and the capacity to consider and solve problems, including complex problems.

VII. Structure, coherence, clarity and fluency of oral expression;

VIII. Structure, coherence, clarity and fluency of written expression;

IX. Intellectual integrity and maturity;

X. Empathy and Imaginative Insight;
Criteria for content

15. History provision across the sector is characterised by a diversity of periods, cultures, methodologies and conceptual assumptions. Nevertheless a number of central requirements can be specified. These six requirements do not point to a particular combination of courses or a particular programme structure:

16. **Time depth:** Awareness of continuity and change over an extended time-span is central to an historical awareness. It leads to an understanding of historical process, and it opens the way to the insights which stem from a juxtaposition of past and present. Programmes should introduce students to the issues of continuity and change and give them experience of the intellectual benefits accruing from the study of History over an extended period of past time. Programmes which cover a relatively short time-span should demonstrate how they provide students with a long-term perspective on their subject-matter.

17. **Geographical range:** For good reasons it has become an established convention that students study the history of more than one society or culture. Among other things this opens the way to comparative perspectives. Where a single country is the focus of the programme, that programme should incorporate serious and sustained comparison with others. The student whose prime interest lies in Britain is in no way exempt from this requirement: the possibilities of enhanced objectivity which flow from studying other countries are particularly important, not to mention the implications of Britain’s Imperial past and multicultural present. History’s ability to promote understanding between cultures and between national traditions remains as important as ever.

18. **Contemporary sources:** Opportunity for close work on source material originating in the period studied is essential. This will often comprise written documents, but when appropriate will include artefacts, visual evidence etc. Students should carry out intensive critical work on such source material. This may take place in a “Special Subject” course, in other courses or in independent work. In many instances the work done by students approximates to historical research. We note that most students do not expect a career in research, but we nevertheless regard documentary work as a necessary part of learning some of the characteristics of the discipline.

19. **Reflexivity:** All History students should be expected to reflect critically on the nature of their discipline, its social rationale, its theoretical underpinnings and its intellectual standing. This may take place in a course labelled historiography or historical method, in other courses or in independent work.

20. **Diversity of specialisms:** History comprises many varieties, each with its distinctive focus and theoretical orientation (for instance, economic, social, political, cultural, environmental history, the history of women, and gender). Students should be introduced to some of these varieties of approach. The aim is not comprehensiveness, but a critical awareness that there are many principles of selection and modes of enquiry. Where a programme is strongly based on one variety, serious comparisons must be made with the contribution of others to historical understanding.

21. **An extended piece of written work:** This allows the student to formulate, execute and complete an independent extended piece of written work with appropriate supervision. In most cases this will be based on contemporary source materials. Alternatively it may take the form of an in-depth historiographical enquiry, as in the critical evaluation of a particular historical controversy, or a particular historian’s oeuvre. It may be free-standing, or it may arise from - and be linked to - a taught course. Where this is not offered at all, departments should demonstrate and justify an alternative and commensurate experience of work in depth.

Progression

22. Students progress through History degree programmes largely by gaining experience and knowledge as they take successive courses over a period normally of three - four years. It is a cumulative process of “learning by doing”. Subject matter varies, and courses may make heavier and/or more sophisticated demands on students over time, but the general process is one of developing and reinforcing similar skills and qualities throughout the programme. Qualitative advances may be achieved in a number of ways, for instance through increasing conceptual sophistication, increasing interpretative skill, increasing capacity for sustained written and/or oral analysis, greater independence of learning, and so on. Departments are not therefore expected to conform to any one model, but they should show how their particular programmes are designed to provide students with the means to gain in insight, competence and performance over 3/4 years. Some programmes may in effect give students nine terms (or six semesters) of doing the same kind of thing with a variety of subject matter but with a growing competence. Others may attach particular skills and attributes to particular courses and prescribe how students shall move through them. Whatever the structure of the programme, students are expected to achieve higher standards at the end of their degree studies than at the beginning; departments will need to show how their particular programme facilitates this
process. Because History is a non-sequential discipline, there is no fixed order of progression from one type of course to another. There is no reason in principle why survey courses should be more strongly represented in year 1, or close documentary study in year 3, if departments are able to demonstrate how their particular provision fosters progression.

**Teaching and learning**

23. **Students should be provided with documentation for each individual course** which explains what the course is designed to achieve, and the means to its attainment. Documents should also include an outline of the course structure, information about the nature and amount of assessment, and a bibliography. These course guides should be designed to be read by students in relation to departmental documentation which includes details of the degree scheme, criteria for all levels of classification and all forms of assessment in use, the range of available courses, course structures, assessment methods and weightings, and advice about plagiarism. Both individual course and departmental documentation should make the teaching and learning available to students in as clear and straightforward a form as possible.

24. **Programmes should offer students regular formal contact with tutors and other students in a variety of structured settings.** The purpose of these engagements is to deepen their research, oral and communication skills. They should also inculcate the qualities of self-discipline, which are necessary for the successful pursuit of the discipline. The precise form and nature of these engagements within departments will, to some extent, be shaped by circumstances. But they will need to be the result of an internal planning process which determines and reviews the match between standard learning outcomes for the degree and the content and teaching/assessment methods employed in the department's individual courses and degree programmes. It is vital to note that teaching methods/learning opportunities should not be thought of as fixed categories. They must be kept under review by departments, with due consideration being given for instance to tutors' self-assessments, student course review questionnaires, and the dissemination of good practice from other departments and Institutions.

25. **There should be opportunities in the degree scheme for all students to experience lecture or lecture-type arrangements which capture their interest and excite their curiosity.** Presentations by lecturers stamp the course or subject with the imprint of personality and enable students to reflect on and to respond to an individual's particular interpretative approach. Lectures provide a broad framework which helps define the course, while also introducing students to its main themes, debates and interpretations. At their best they offer direct entry into a range of information and ideas which students may never, or only very rarely, be able to gather for themselves; they thus serve as a launching-pad into the heart of new subject matter. They should enable the students to develop their skills in listening, selective note-taking, and reflection. Lectures will be more appropriate to certain kinds of courses than to others and it is for course teams to decide on their optimum deployment. We do not suggest that they should be employed in all courses. We also recognise that they may take many different forms including the use of AV, CAL etc.

26. **There should be a requirement during the degree scheme for all students to engage in seminars and forms of group work.** In these sessions students should be expected to participate in group discussion, give presentations and jointly explore themes and arguments. These group discussions should be aimed at improving students' understanding rather than at the acquisition of knowledge per se and should be structured in such a way as to maximise effective student participation. They will normally be preceded by a prescribed programme of reading. Such work should be seen as both deepening students' understanding of a theme or subject and developing oral communication skills. It encourages a critical, as well as self-critical but tolerant, approach to historical discussion and builds students' self-confidence. It improves their abilities to marshal historical evidence and to summarise historical arguments, as well as to think quickly on their feet, to communicate articulately and persuasively with others and to recognise the value of working closely with others.

27. **Most of a history student's time is spent working independently, reading, thinking and writing.** Course bibliographies and other reading advice will provide students with the necessary starting points, but they should be encouraged by tutors to make imaginative use of the library, the web etc. to expand their knowledge base and their range of historical approaches. History is largely a text-based discipline which requires students to learn to read widely, rapidly and critically, to take good notes, to digest arguments and to synthesise information quickly and intelligently. It also requires them to construct arguments in writing.

28. **Students should be expected to undertake a wide range of assignments (such as seminar and group presentations, reports, reviews, gobbets or document papers, essays of varying lengths, C & IT projects, dissertations).** It should be explained to students how such assignments enable them to improve their
writing and oral-communication skills, as well as those of evidence-handling, the critical treatment of themes/historical arguments and the thoughtful, persuasive presentation of their work. Assignments should be appropriate to the aims and intended outcomes of the course, though equity in the treatment of students and a balanced range of assessment across the whole programme must remain important considerations. We do not support the view that all courses should necessarily be of one term/semester duration. Certain types of historical course have been most effectively taught on a year-long basis. Where academic judgement indicates this to be desirable, this model should continue to be followed.

29. All students should receive critical and constructive comment on their progress as an integral part of teaching and learning. There should be adequate discussion of, and response to, a student's individual work. As a basic minimum, all tutors should specify in writing a period or periods each week when a given course is being taught when they will be available for academic consultation. Individual encounters with tutors, whether as part of a regular teaching arrangement, feedback on written work or more occasional meetings, are essential to helping students clarify areas of confusion and gain a better understanding of their own performance as well as of historical themes and issues.

Assessment

30. Diversity in assessment is vital for two main reasons. First, the full range of a student's abilities is most unlikely to be revealed through any single mode. Second, the increasingly diverse educational background and formal qualifications presented on entry suggest that the degree programme should afford all students the opportunity to show what they know, understand and can do.

31. Assessment of undergraduate performance is diverse. However, the essay remains a central component. The essay is a piece of written work in response to a particular question or issue, done either under examination conditions or as coursework. Essays require students to demonstrate a number of skills in combination. Because of the integrative high-order skills which they develop, they are an essential element of all History assessment at this level. We recommend that all single-honours students should be assessed in significant part on their essay-writing skills. We recommend that all departments should give serious consideration to the provision of opportunity for single-honours students to be assessed by essays of various types (as, for example, 'long' essays reflecting depth of scholarship, 'short' essays requiring precision of focus; essays focusing on different historical concepts - change, cause, similarity and difference etc; essays written to a target length and essays written to time). We also recommend that departments give serious consideration to requiring students to write at least some essays under exam conditions which afford safeguards against plagiarism and the use of inappropriate outside assistance. This also gives students the opportunity to develop relevant life-skills such as the ability to produce coherent, reasoned and supported arguments under pressure.

32. We recommend that all single honours students should be assessed in some way or another on their understanding of and their ability to handle primary source material.

33. We recommend that all single honours history students should be assessed on their ability to address historical problems in depth. Students should have the chance to pursue a historical enquiry, sustaining and developing it through several stages. Such an exercise involves both task-setting and problem-solving. This is normally done through an extended piece of written work, usually of at least twice the length required for standard coursework essays.

34. Departments should also consider whether single-honours students should be given the opportunity to have their critical and communication skills assessed in other forms. The development of oral communication skills is important in the process of educating a historian and students should, where practicable, have opportunity to be assessed on this skill. Oral presentations can be of different types including, for example, formal paper delivered to a group; general contribution to seminar discussion; chairing or otherwise leading seminar discussion, and response to contributions made by others.

35. We recommend that all departments give serious consideration to ensuring that single honours students also have the opportunity to have their critical and communication skills assessed in some of the following ways:

i. team working and collaborative activity: group projects, fact-finding, evidence-processing work, etc;
ii. shorter written tasks, including historical literature reviews and reports;
iii. use of information technology to answer questions about historical data, including statistical and/or graphical analysis of historical data sets and to present findings in a variety of appropriate forms (bar graphs, pie charts, etc);
iv. use of information technology for bibliographic and archive searches;
v. practical experience in the use of archival material.

36. It is not the intention of the History Benchmarking Group to prescribe any one assessment strategy. Establishing criteria for classification is the business of departments and institutions. Different modes of assessment will be appropriate to different schemes of study, and will reflect the particular emphases and concentrations in those programmes. In order properly to evaluate the range of undergraduate study, an honours degree in history should be awarded on the basis of more than one form of assessment.

Assessment criteria

37. Departments should operate, and publish to their students, descriptors which characterise levels of performance characteristic of first, upper second, lower second, and third classes. This might take the form of a template setting out assessment criteria, and an example of such a template is annexed to this document. Where different modes of assessment privilege different qualities (for example essay writing in examination conditions, extended essays, dissertations, oral competence) different templates may need to be produced. Such templates or their equivalent should be published in student handbooks.

38. Given that the qualities of the graduate historian will differ from those who have pursued honours degrees in different disciplines, procedures for assessing performance may be different in history from those appropriate to other disciplines. History departments, sections, or subject groups should have the autonomy to establish criteria for classification appropriate to the discipline and to this benchmark statement. Benchmark standards should inform, but not unduly circumscribe, the approach to determining modes of assessment, and the precise weightings given to assessing different historical skills, methodologies, and learning outcomes.

39. Published criteria should be available for all forms of assessment, including the assessment of long essays, dissertations, oral presentations, and vivas where any or all of these modes of assessment are used. Criteria at all levels of classification should give predominance to positive achievement, although below the first class they should also indicate the kinds of limitations which disqualify a candidate from achieving a mark in a higher class.

40. Individual Institutions might wish to develop new methods for describing undergraduate achievement. The basic threshold for achievement of Honours (H4) must remain the standard required to achieve a Third Class in traditional systems of classification. Describing performance above this level might be done through issuing transcripts rather than by classification. Institutions might issue transcripts containing assessment marks for all courses or modules and an overall percentage mark (which might be weighted). Such a procedure would discriminate more precisely between different candidates' performance and would enable Institutions to show how a student performed across a range of assessments. Outstanding performance might be rewarded by graduating with distinction. Such a candidate would have achieved the same overall standard as a student graduating first class in a traditional system of classification.

41. An honours degree in History should normally reflect at least two years (or four semesters) of work beyond a previously qualifying standard. The qualifying standard would normally be one year's study at or equivalent to H1.

42. All graduates in history should demonstrate competence in the discipline and the purpose of schemes of assessment is to evaluate the level of competence achieved. In establishing and maintaining history degree programmes, departments should take into account the following summary of learning outcomes. They will not necessarily wish to include assessment of all these learning outcomes in degree classification:

i. command of a substantial body of historical knowledge;
ii. the ability to develop and sustain historical arguments in a variety of literary forms, formulating appropriate questions and utilizing evidence [12 (vi), 12 (vii)];
iii. an ability to read, analyse, and reflect critically and contextually upon historical texts [12 (ii), 18];
iv. an appreciation of the complexity of reconstructing the past, the problematic and varied nature of historical evidence [12 (iv)];
v. an understanding of the varieties of approaches to understanding, constructing, and interpreting the past; and, where relevant, a knowledge of concepts and theories derived from the humanities and social sciences [7, 20];
vi. the ability to read, analyse, and reflect critically and contextually upon historical texts and other source materials [12 (ii), 17, 32];
vii). the ability to gather and deploy evidence and data to find, retrieve, sort and exchange new information [14 (v), 14 (vi), 35 (iii), 35 (iv), 35 (v)];

viii. a command of comparative perspectives, which may include the ability to compare the histories of different countries, societies, or cultures [17];

ix. awareness of continuity and change over extended time spans [16];

x. an understanding of the development of history as a discipline and the awareness of different historical methodologies [19];

xi. an ability to design, research, and present a sustained and independently-conceived piece of historical writing [21, 31, 33];

tax. the ability to address historical problems in depth, involving the use of contemporary sources and advanced secondary literature [33];

taxi. clarity, fluency, and coherence in written expression [12 (vi), 14 (vii), 27, 28, 31, 35 (iii)];

axi. clarity, fluency, and coherence in oral expression [12 (vii), 14 (vii), 26, 28, 34, 35];

axii. the ability to work collaboratively and to participate in group discussion [26, 34, 35 (i)];

axiii. competence in specialist skills which are necessary for some areas of historical analysis and understanding, as appropriate [13, 35 (iii), 35 (iv)].

Recommendations

1. The Group recommends that all students studying History as part of their degree:

i. Undertake a programme which fosters the skills and qualities of mind listed in paragraphs 12 and 13 of this statement.

ii. Be provided with comprehensive course and department documentation

iii. Be provided with opportunities to participate regularly in a variety of structured settings with tutors and other students.

iv. Receive good diagnostic feedback on their progress as an integral part of teaching.

2. The Group recommends that all single history honours students:

i. Follow a programme which gives them practical experience of the intellectual benefits occurring from studying the subject over an extended period of historical time.

ii. Study the history of more than one society or culture.

iii. Carry out intensive critical work on source materials generated by the period under study.

iv. Be expected to reflect critically on the nature of their discipline.

v. Be introduced to some of the many varieties of History.

vi. Be involved in lecture or lecture-type arrangements which capture their interest and excite their curiosity.

vii. Engage in seminars and forms of group work.

viii. Undertake a wide range of assignments.

ix. Be assessed in significant part on their essay-writing skills.

x. Be assessed on their understanding of and ability to handle contemporary source material.

xi. Be assessed on their ability to address historical problems in depth.

3. The Group recommends that all departments should give serious consideration to requiring that all single history honours students will:

i. Formulate, execute and complete an independent extended piece of written work, with appropriate supervision on which they are assessed.

ii. Write at least some of their essays under exam conditions.

3. Departments will also wish to consider the desirability of providing the opportunity for all single honours students to be assessed on:

i. Varying types of and extended writing.

ii. Oral communication.

iii. Other forms of presentation.
Annex 1

Assessment criteria for examination by essays written under timed conditions

First class

Structure and focus
- Work which engages closely with the question set, and shows a mature appreciation of its wider implications.
- The structure of the answer will facilitate a clear, coherent, and compelling development of the writer’s argument.
- Descriptive material and factual evidence will be deployed in order to support and develop the writer’s argument, and will be deployed with a vigorous sense of relevance and an appropriate economy of expression.

Quality of argument and expression
- The writing will be clear, fluent, and accurate. The range of vocabulary and linguistic idioms will be appropriate to the case being developed.
- The answer will go well beyond the effective paraphrasing of other historians’ ideas, and demonstrate conceptual command of the historical (and, where appropriate, historiographical) issues at stake.
- The answer may develop ideas which are original, and may be structured in a way which enables the writer to develop independent lines of thought in compelling and coherent ways. Intellectual independence, when grounded in a mature consideration of available evidence, should take the candidate into the highest markbands.

Range of knowledge
- Relevant knowledge is both broad and deep. This will include knowledge of contemporary sources, historiography, secondary literature. The range of reading implied by the answer will be extensive.
- The writer will show an ability to move between generalization and detailed discussion, and will be able to synthesize as well as particularize.
- Writers will show an ability to evaluate the nature and status of information at their disposal, and where necessary identify contradiction and attempt a resolution.
- The answer will demonstrate an informed and secure understanding of the historical period or periods under discussion.

Upper Second class

Structure and focus
- Work which displays an understanding of the question, shows an appreciation of some of its wider implications, and makes a serious attempt to engage with the question set.
- The structure of the answer will facilitate a clear development of the writer’s argument, towards the lower end of this markband candidates will not sustain an analytical approach throughout.
- Descriptive material and factual evidence will be deployed relevantly. Towards the lower end of this markband candidates may not always bring out the full implications of evidence cited.

Quality of argument and expression
- The writing will be clear and generally accurate, and will demonstrate an appreciation of the technical and advanced vocabulary used by historians.
- The answer will deploy other historians’ ideas and seek to move beyond them. The answer will also show an appreciation of the extent to which historical explanations are contested.
- Although the answer might not demonstrate real originality, the writer will present ideas with a degree of intellectual independence, and will demonstrate the ability to reflect on the past and its interpretation.
Range of knowledge

- Knowledge is extensive, but might be uneven. Demonstrated knowledge will include reference to relevant contemporary and historiographical sources. The range of reading implied by the answer will be considerable.
- The answer will demonstrate a sense of the nature of historical development.
- The writer will show an ability to move between generalization and detailed discussion, although there may be a tendency towards either an over-generalized or an over-particularized response.
- Writers will reflect on nature and status of information at their disposal, and will seek to use it critically.
- The answer will demonstrate a secure understanding of the historical period or periods under discussion.

Lower Second class

Structure and focus

- Work which displays some understanding of the question set, but may lack a sustained focus and may show only a modest understanding of the question's wider implications.
- The structure of the answer may be heavily influenced by the material at the writer's disposal rather than the requirements of the question set. Ideas may be stated rather than developed.
- Descriptive material and factual evidence will be deployed, but not necessarily with the kind of critical reflections characteristic of answers in higher markbands.

Quality of argument and expression

- The writing will be sufficiently accurate to convey the writer's meaning clearly, but it may lack fluency and command of the kinds of scholarly Idioms used by professional historians. In places expression might be clumsy.
- The answer will show some understanding of historians' ideas, but may not reflect critically upon them. The problematic nature of historical explanations may be imperfectly understood.
- The answer is unlikely to show any originality in approach or argument, and may tend towards assertion of essentially derivative ideas.

Range of knowledge

- Knowledge will be significant, but may be limited and patchy. There may be some inaccuracy, but basic knowledge will be sound. The range of reading implied by the answer will be limited.
- The answer will show some limited awareness of historical development.
- The writer might be prone to being drawn into excessive narrative or mere description, and may want to display knowledge without reference to the precise requirements of the question.
- Information may be used rather uncritically, without serious attempts to evaluate its status and significance.
- The answer will demonstrate some appreciation of the nature of the historical period or periods under discussion.

Third Class

Structure and focus

- Work which displays little understanding of the question, and may tend to write indiscriminately around the question.
- The answer will have structure but this may be underdeveloped, and the argument may be incomplete and unfold in a haphazard or undisciplined manner.
- Some descriptive material and factual evidence will be deployed, but without any critical reflection on its significance and relevance.
Quality of argument and expression

- The writing will generally be grammatical, but may lack the sophistication of vocabulary or construction to sustain a historical argument of any complexity. In places the writing may lack clarity and felicity of expression.
- There will be little appreciation of the problematic or contested nature of historical explanations.
- The answer will show no intentional originality of approach.

Range of knowledge

- There will be sufficient knowledge to frame a basic answer to the question, but it will be limited and patchy. There will be some inaccuracy, but sufficient basic knowledge will be present to frame a basic answer to the question. The answer will imply relevant reading but this will be slight in range.
- There will be understanding of historical development but it will be underdeveloped, and the ideas of historians and other writers may be muddled or misrepresented.
- There will be an argument, but writer may be prone to excessive narrative, and the argument might be signposted by bald assertion rather than informed generalizations.
- There will be sufficient information to launch an answer, but perhaps not to sustain a complete response. Information will be used uncritically as if always self-explanatory.
- The answer will demonstrate appreciation of the nature of the historical period or periods under discussion, but at a rudimentary level.
Annex 2

A statement of the threshold standard

The Benchmarking statement for History has sought to provide, for all stakeholders, the general expectations about standards in the Single Honours History degree. This statement offers, in particular, Information and advice about the content and progression of an Honours History degree programme, the specific historical and general transferable skills that such a degree programme will impart, and the means by which such qualities of mind can be assessed. It also, of course, indicates the abilities that an Honours graduate in History can be expected to possess.

The Benchmarking statement has recognised that the historical content in the many Single Honours Degree programmes on offer in the United Kingdom will vary in detail, although they are likely to share certain general characteristics. While the specific content of History degree programmes will undoubtedly vary, all will teach a substantial body of historical knowledge and all will develop the particular historical skills and the general transferable skills expected of an Honours graduate in History. These historical skills will include: an appreciation of the complexity and diversity of situations, events and mentalities in the past and of the surviving evidence about them; the ability to read, analyse, and reflect critically and contextually upon, a wide range of source materials; an awareness of the varieties of approaches to understanding, constructing and interpreting the past; an understanding of history as a discipline and of different historical methodologies; an awareness of continuity and change over time; an ability to gather evidence to develop and sustain historical arguments; and the ability to marshal an argument and to express it with clarity, fluency and coherence. An Honours History programme may expect students to employ other cognate skills in, for example, languages, computing and quantitative methods, and will certainly seek to develop such generic or transferable skills as self-discipline, self direction, independence of mind, empathy and imaginative insight, and the ability to work with others and to have respect for the reasoned view of others.

A student who has graduated with an excellent performance in Single Honours History can confidently be expected to have mastered a very extensive range and depth of historical knowledge in particular areas of the past and to have demonstrated a very superior command of nearly all the historical and transferable skills outlined above. The typical or modal History graduate will have acquired a considerable range and depth of historical knowledge in particular areas of the past and will be able to show a very sound competence in nearly all of the historical and transferable skills outlined above. A student who has shown a weak command of historical knowledge and a limited command of the historical and transferable skills expected of an Honours History graduate will not have succeeded in graduating with an Honours degree, but will have failed at an earlier stage in the Single Honours degree programme. Those students graduating in Single Honours History at the threshold level may well have demonstrated an unevenness of performance in the various courses or modules, and in the assessed work, in their degree programme. The unevenness can be detected by an examination of the range and diversity of marks awarded for their various courses or modules. Some students may perform consistently at the threshold level, but most students who graduate at the threshold level, will, none the less, have shown in parts of their degree programme the level of competence shown by the typical or modal student. At their best, they will have displayed sound historical knowledge and competence in many of the historical and transferable skills expected of an Honours History graduate, but they have graduated at the threshold level because they have not been able to sustain that level of performance across all of their courses and assessed work. Only a part of their performance is likely to have been assessed at the lowest threshold level. Even those students who have performed consistently at this level, however, will have demonstrated a basic understanding of the evidence, will have presented a structured, if underdeveloped and incomplete argument or have presented a thorough narrative with insufficient analysis, and will have expressed themselves in a manner that is generally grammatical and intelligible, if somewhat lacking in clarity and fluency.
**History subject benchmarking group membership**

<table>
<thead>
<tr>
<th>Name</th>
<th>Institution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dr M Arnot</td>
<td></td>
</tr>
<tr>
<td>Professor D Bates</td>
<td>University of Glasgow</td>
</tr>
<tr>
<td>Professor C Clark</td>
<td>University of Warwick</td>
</tr>
<tr>
<td>Professor M Daunton</td>
<td>Churchill College, University of Cambridge</td>
</tr>
<tr>
<td>Professor H Dickinson</td>
<td>University of Edinburgh</td>
</tr>
<tr>
<td>Dr Susan Doran</td>
<td>St Mary's College, Twickenham</td>
</tr>
<tr>
<td>Professor W Doyle</td>
<td>University of Bristol</td>
</tr>
<tr>
<td>Professor D Eastwood</td>
<td>University of Wales, Swansea</td>
</tr>
<tr>
<td>Professor E Evans</td>
<td>University of Lancaster</td>
</tr>
<tr>
<td>Professor A Fletcher (Chair)</td>
<td></td>
</tr>
<tr>
<td>Professor A Jones</td>
<td>University of Wales, Aberystwyth</td>
</tr>
<tr>
<td>Mr R Lloyd-Jones</td>
<td>Sheffield Hallam University</td>
</tr>
<tr>
<td>Dr E McFarland</td>
<td>Glasgow Caledonian University</td>
</tr>
<tr>
<td>Professor A Porter</td>
<td>King's College London</td>
</tr>
<tr>
<td>Professor P Stafford</td>
<td>University of Huddersfield</td>
</tr>
<tr>
<td>Professor J Tosh</td>
<td>University of North London</td>
</tr>
</tbody>
</table>
References


AUT (2001) *HEFCE 01/45, Quality assurance in higher education: a response from the Association of University Teachers*, London: AUT


Barnett, R. (1998) ‘“In” or “For” the Learning Society?’, *Higher Education Quarterly, 52*, 1, pp. 7-21


Baty, P. (2001b) ‘LSE leads revolt against QAA’, *THES*, 23 March

Baty, P. (2001c) ‘Gang of five plans to escape QAA grip’, *THES*, 6 April, p.3


Baty, P. (2002) ‘”Guidance” ousts “compliance” at QAA’, *THES*, 10 May, p.4


Becher, T. and Trowler, P.R. (2001) *Academic Tribes and Territories. Intellectual enquiry*
and the culture of disciplines, 2nd ed, Buckingham: SRHE & Open University Press


Canovan, C. (2001) ‘Randall stresses the need for accountability’, THES, 19 October, p.8


& Lybrand and CBI


Dean, J.P. and Whyte, W.F. (1958) ‘How Do You Know If the Informant is Telling the Truth?’, Human Organization, 17, 2, pp.34-38


DES (1994) Education Secretary John Patten’s Speech to HEFCE Conference, 12 April 1994


Doherty, G.D. (1997) ‘Quality, standards, the consumer paradigm and developments in
higher education', *Quality Assurance in Education*, 5, 4, pp.239-248


Furedi, F. (2002) ‘Would you fail the Trades Description Act?’, *THES*, 16 August, p.21


Harker, B. (1995)'Postmodernism and Quality', *Quality in Higher Education*, 1, 1, pp.31-39


HEFCE (2000) Better accountability for higher education. A review for the HEFCE by PA Consulting, Bristol: HEFCE [HEFCE 00/36]

HEFCE (2001) Quality assurance in higher education. Proposals for consultation, Bristol: HEFCE [01/45]


Inglis, F. (2004) ‘Cold heart needs to be revived’, *THES*, 13 February, p.16


Kenyon, C. (1997) 'QAA Chairman's Speech to the CVCP Conference, University of Strathclyde, 18 September 1997', *higher quality*, 2, pp.18-20


King, R. (2002) 'Institutional diversity is a chimera best put away', *THES*, 19 April, p.14


218


219


Nisbett, J. and Watt, J. (1978) *Case Study*, Nottingham: Nottingham University, School of Education


Platt, Jennifer (1981a) ‘On interviewing one’s peers’, *British Journal of Sociology, 32*, 1, pp.75-91


Pratt, J. (2002) 'Status: more than just a name', *THES*, 20 September, p.25


QAA (1998a) ‘Consultation: developing the quality assurance and standards framework for UK higher education’, *higher quality*, 3, pp.3-23


QAA (1998c) ‘Responses to an Agenda for Quality’, *higher quality*, 4, pp.13-21

QAA (1999a) ‘Subject benchmarking: update’, *higher quality*, 5, pp.4-7


QAA (1999c) ‘Quality assurance: the new approach’, *higher quality*, 6, pp.3-8

QAA (1999d) ‘Illustrative report on quality and standards’ *higher quality*, 6, pp.16-22


QAA (2000b) ‘Subject benchmarking: publication of statements and formation of new benchmarking groups’, *higher quality*, 7, p.4

QAA (2000c) *Chemistry [Subject benchmark statement]*, Gloucester: QAA [AR 002 223]


QAA (2001a) ‘Further progress in benchmarking academic standards’, higher quality, 8, p. 4

QAA (2001b) Quality assurance in UK higher education: a brief guide, Gloucester: QAA [QAA 016 05/01]


QAA (2002a) QAA external review process for higher education in England: Operational description, Gloucester: QAA [QAA 019 03/02]


QAA (2002e) ‘Information for students’, www.qaa.ac.uk

QAA (2003a) A brief guide to quality assurance in UK higher education, www.qaa.ac.uk


Newsletter, Summer, pp.1-2


Randall, J. (2000b) 'Quality street', *Guardian Education*, 25 April, p.11


Schuman, Howard (1982) 'Artifacts are in the Mind of the Beholder', *The American Sociologist*, 17, pp.21-28

225


