Dimensions *of* quality

by Graham Gibbs

The Higher Education Academy

ISBN 978-1-907207-24-2

The Higher Education Academy
Innovation Way
York Science Park
Heslington
York YO10 5BR

www.heacademy.ac.uk

Tel: +44 (0)1904 717500
Fax: +44 (0)1904 717505

To request copies of this report in large print or in a different format, please contact the Academy.

Designed by Daniel Gray
Printed by The Charlesworth Group

DIMENSIONS OF QUALITY

by GRAHAM GIBBS

FOREWORD

The perennial debate about what constitutes quality in undergraduate education has been reignited recently, not least by a range of published research, Select Committee activity, tightening of resource, and the large-scale review by Lord Browne.

As the organisation dedicated to enhancing the quality of students' learning experiences, the Higher Education Academy is pleased, through this piece of work, to contribute further to this important debate.

Our starting-point is twofold: first, that higher education should be a transformative process that supports the development of graduates who can make a meaningful contribution to wider society, local communities and to the economy. Second, that any discussion around quality needs to be evidence-informed. As a result, we identified a need to synthesise and make sense of the scattered research in the field of higher education quality. We wanted to find out what the research evidence tells us and what further work we can do to apply the relevant findings in our quest to improve the quality of student learning in UK higher education.

Graham Gibbs states that the most important conclusion of this report is that what best predicts educational gain is measures of educational process: in other words, what institutions do with their resources to make the most of the students they have. Examining the evidence, he draws conclusions about some key topics that have been the subject of much debate around quality. For example, he concludes that the number of class contact hours has very little to do with educational quality, independently of what happens in those hours, what the pedagogical model is, and what the consequences are for the quantity and quality of independent study hours.

He also reiterates research (Nasr *et al.*, 1996) that shows that teachers who have teaching qualifications (normally a Postgraduate Certificate in Higher Education, or something similar) have been found to be rated more highly by their students than teachers who have no such qualification. I think this is a crucial point. At the Academy we believe that high quality teaching should be delivered by academic staff who are appropriately qualified and committed to their continuing professional development. To this end we will continue to provide and develop an adaptable framework for accredited teaching qualifications in HE, incorporating the UK Professional Standards Framework and other relevant teaching qualifications. We will also continue to work with HEIs to develop and manage CPD frameworks for learning and teaching.

The report also concludes that some dimensions of quality are difficult to quantify, and it is therefore difficult to see what effect they might have. Aspects of

departmental culture are one such area: whether teaching is valued and rewarded, whether innovation in teaching is systematically supported and funded, etc. The Academy has already conducted research into the reward and recognition of teaching which showed that over 90% of academic staff thought that teaching should be important in promotions. We will continue to focus on this work.

Some of the findings of this report may confirm aspects of institutional policy on enhancing quality, some of them will prompt new and different approaches to focused investment of funding and expertise in order to maximise educational gain, particularly at a time of diminishing resource. Some of them will call into question the efficacy and appropriateness of practices and policies, and cause us to look not at how much is spent per capita, but on how it is spent; less on how many contact hours are provided but with whom and with what consequences for independent learning; on the extent to which we truly support and adopt the kinds of pedagogic practices that engender students' intrinsic engagement in their learning.

Graham argues for a better focus on evidence in order to understand quality properly, to ensure that our quality process are informed to a greater extent by what we know about what constitutes effective practice and about the extent to which these practices are employed, to make better and more coordinated use of the full range of available data, and to understand the relationship between them.

This paper is primarily for an audience of senior managers of HEIs – the colleagues who develop and implement the kinds of institutional policies that have the propensity to improve student learning and who conceptualise the frameworks to support that vital process. We hope that this report will meaningfully inform both policy and practice and look forward to following up this work in the coming months by engaging with you in debates and discussions about the dimensions of quality.

Professor Craig Mahoney
Chief Executive
The Higher Education Academy.

1. EXECUTIVE SUMMARY

> *"A … serious problem with national magazine rankings is that from a research point of view, they are largely invalid. That is, they are based on institutional resources and reputational dimensions which have only minimal relevance to what we know about the impact of college on students … Within college experiences tend to count substantially more than between college characteristics."*
> —Pascarella, 2001

1.1 The focus of the report

This report has been written to contribute to the current debates about educational quality in undergraduate education in the UK, and about the need to justify increases in resources on the basis of indicators of educational quality. This report will identify a range of dimensions of quality and examine the extent to which each could be considered a valid indicator, with reference to the available research evidence. It attempts to identify which kinds of data we should take seriously and which we should be cautious of placing weight on. Some of these dimensions we might be wise to pay attention to currently lack a solid evidence base, especially in relation to research carried out in the UK context, and so the report also identifies priorities for research and for data collection and analysis.

1.2 The approach taken to considering dimensions of quality

The report identifies which dimensions of educational quality can reasonably be used to compare educational settings. It adapts Biggs's '3P' model (Biggs, 1993) of 'presage', 'process' and 'product' to categorise the variables under consideration (see Section 3.2). The report examines a wide range of potential indicators. Presage variables define the context before students start learning, process variables describe what goes on as students learn, and product variables relate to the outcomes of that learning. For presage and process variables the available evidence is examined concerning the validity of the variable: the extent to which it predicts student learning outcomes and educational gains. Each product variable is examined for its ability to indicate comparative quality.

1.3 The limitations of presage and product variables

Presage variables such as funding, research performance and, the reputation that enables an institution to have highly selective student entry, do not explain much of the variation between institutions in relation to educational gains. Measures of educational product such as grades do reflect these presage variables, but largely because the best students compete to enter the best-funded and most prestigious institutions and the quality of students is a good predictor of products. Measures of product such as retention and employability are strongly influenced by a raft of presage variables that go well beyond those used by HEFCE in setting performance benchmarks. The lack of comparability of degree standards proves an obstacle to interpretation of student performance data in the UK. This makes interpreting and comparing institutional performance extremely difficult.

1.4 The importance of process variables

What best predicts educational gain is measures of educational process: what institutions do with their resources to make the most of whatever students they have. The process variables that best predict gains are not to do with the facilities themselves, or to do with student satisfaction with these facilities, but concern a small range of fairly well-understood pedagogical practices that engender student engagement. In the UK we have few data about the prevalence of these educational practices because they are not systematically documented through quality assurance systems, nor are they (in the main) the focus of the National Student Survey.

Class size, the level of student effort and engagement, who undertakes the teaching, and the quantity and quality of feedback to students on their work are all valid process indicators. There is sufficient evidence to be concerned about all four of these indicators in the UK.

1.5 The importance of multivariate analysis

Few relationships between a single dimension of quality and a single measure of either educational performance or educational gain can be interpreted with confidence because dimensions interact in complex ways with each other. To understand what is going on and draw valid conclusions it is necessary to have measures of a range of dimensions of quality at the same time and to undertake multivariate analysis. Large-scale multivariate analyses have been repeatedly undertaken in the US, and have successfully identified those educational processes that affect educational gains, and those that do not or that are confounded by other variables. In contrast there has been little equivalent analysis in the UK. This is partly because data in the UK that

could form the basis of multivariate analysis for that purpose are currently collected by different agencies and have never been fully collated.

1.6 The importance of educational gain

Because educational performance is predicted by the entry standards of students, to compare institutional performance in a valid way it is necessary to measure educational gain: the difference between performance on a particular measure before and after the student's experience of higher education. While the most influential US studies measure educational gain in a variety of ways, there is very little evidence available in the UK about educational gain.

1.7 Dimensions of quality in different kinds of institutions

Institutions have different missions, and comparing them using product dimensions of quality that are the goals of only a subset of the institutions leads to conclusions of doubtful value. Process dimensions give a fairer comparative picture of quality than do presage or product dimensions. However, different pedagogic phenomena, and hence different process variables, are likely to be salient in different institutions. For example, only some of the very different ways in which The Open University or the University of Oxford achieve such high National Student Survey ratings are relevant to other kinds of university.

1.8 Dimensions of quality in different departments

Indicators of dimensions of quality often vary widely between departments within the same institution, for a variety of reasons. Prospective students need quality information about the specific degree programme they wish to study at an institution rather than about institutional averages or about clusters of degree programmes aggregated into 'subjects' as at present. Providing such information at a sufficient level of granularity may be impractical.

1.9 Dimensions of quality that are difficult to quantify

Studies of the characteristics of both institutions and departments that have been found to be outstanding in terms of valid dimensions of educational quality have identified process variables that would be extremely difficult to quantify or measure in a safe way, such as the extent to which teaching is valued, talked about and developed.

1.10 Evidence of the products of learning

One of the most telling indicators of the quality of educational outcomes is the work students submit for assessment, such as their final-year project or dissertation. These samples of student work are often archived, but rarely studied. There is considerable potential for using such products as more direct indicators of educational quality than proxies such as NSS scores.

1.11 The potential for improved quality, and the evaluation of improvements in quality

There is clear evidence that educational performance and educational gains can be enhanced by adopting certain educational practices. In the US the National Survey of Student Engagement (NSSE) has been used successfully by many institutions to identify where there are weaknesses in current educational processes and to demonstrate the positive impact of the introduction of certain educational practices.

Pooling data across such innovations then provides a valid basis to guide other institutions in the adoption of practices that are likely to be effective. The NSS cannot be used in the UK in the same way, despite its reliability. There is a valuable role to be fulfilled by national agencies in supporting the use of valid measures of the impact of changed educational practices, and in pooling evidence across institutions.

1.12 The potential for informing potential students about quality

It seems unlikely that comparative indicators of quality currently available in the UK could provide prospective students with a valid basis to distinguish between individual courses with regard to their educational quality. The collation of currently available data into league tables is invalid and misleading. Even in the US where a range of more valid indicators are more widely available, those responsible for collecting and interpreting the data counsel strongly against their collation into a single league table.

2. INTRODUCTION

The extent to which indicators of quality have shaped both the politics of higher education and institutional priorities is not a new phenomenon (Patrick and Stanley, 1998). However, there is currently increased emphasis on the overall quality of undergraduate education in the UK. Data from a number of recent surveys and studies have raised challenging issues about:

- differences in quality between institutions within the UK that in the past have, rightly or wrongly, been assumed to be broadly comparable;
- differences in quality between national higher education systems, to whom in the past the UK has been assumed, rightly or wrongly, to be superior, in the context of an increasingly competitive international higher education market place;
- the adequacy of national quality regimes that have emphasised scrutiny of an institution's quality assurance to a greater extent than of its educational processes or outcomes of the kind emphasised in some of the recent high profile surveys and studies.

A Parliamentary Select Committee (House of Commons, 2009) has taken evidence from a wide range of sources and reached challenging conclusions both about the quality of UK higher education and how that quality can be assured in the future.

Among all the debate there has sometimes been uncritical acceptance of some sources of evidence that cannot bear the weight of interpretation, and also rejection of evidence that deserves to be taken more seriously. Even in public reports argument has sometimes made no use of available evidence. To give one example the Quality Assurance Agency (2009) has responded to data that suggest both that UK students might study significantly less hard than their European counterparts, and that there are wide differences between institutions and subjects within the UK in relation to how many hours are studied (HEPI, 2006, 2007; Brennan *et al.*, 2009). From the perspective of the current report the key questions in this case are:

- does it matter that some students receive less class contact than others? Are class contact hours an indicator of quality?
- does it matter that some students put in less total effort than others? Are total student learning hours an indicator of quality?

In Section 5.2 below, evidence is reviewed that might inform the QAA's current position on this issue.

Similarly the findings of a study of student experience by the National Union of Students (NUS, 2008) might be interpreted differently if they were informed by the available empirical evidence on the issues it addresses, such as the effects of paid work on students' study hours.

The literature on the validity of indicators of quality is vast, widely dispersed and mostly American. It tends to be focused on specific purposes, such as critiquing a particular university league table, critiquing a particular government-defined performance indicator, establishing the characteristics of a particular student feedback questionnaire, or examining the characteristics of a particular indicator (such as research performance). Much of this literature is technical in nature and written for a specialist audience of educational researchers. The current report attempts to bring much of this diverse literature together encompassing many (though not all) dimensions of quality. It is not intended to be an exhaustive account, which would be a very considerable undertaking, and it is written for a general audience. It will not delve into statistical and methodological minutiae, although sometimes an appreciation of statistical issues is important to understanding the significance of findings.

This report is intended to inform debate by policy formers of four main kinds: those concerned about the overall quality of UK higher education; those concerned with institutional and subject comparisons; those concerned with funding on the basis of educational performance and those within institutions concerned to interpret their own performance data appropriately. It may also be useful to those directing resources at attempts to improve quality as it identifies some of the educational practices that are known to have the greatest impact on educational gains.

It is important here to be clear what this report will not do. It will not review alternative quality assurance regimes or make a case for any particular regime. In identifying dimensions of quality that are valid it will, by implication, suggest elements that should be included in any quality assurance regime, and those that should not be included.

The report will not be making overall comparisons between the UK and other HE systems, between institutions within the UK, between subjects nationally or between subjects or departments within institutions. Rather the purpose is to identify the variables that could validly be used in making such comparisons.

The report is not making a case for performance-based funding. Reviews of the issues facing such funding mechanisms can be found elsewhere (Jongbloed and Vossensteyn, 2001). However, valid indicators of quality will be identified that any performance-based funding system might wish to include, and invalid indicators will be identified that any performance-based system should eschew.

Finally, the report is not making a case for the use of 'league tables' based on combinations of quality indicators, nor does it consider the issues involved in the compilation and use of existing or future league tables. Trenchant and well-founded

critiques of current league tables, and of their use in general, already exist (Bowden, 2000; Brown, 2006; Clarke, 2002; Eccles, 2002; Graham and Thompson, 2001; Kehm and Stensaker, 2009; Thompson, 2000; Yorke, 1997). Some of these critiques cover similar ground to parts of this report in that they identify measures commonly used within league tables that are not valid indicators of educational quality.

Throughout the report there is a deliberate avoidance of using individual institutions in the UK as exemplars of educational practices, effective or ineffective, with the exception of a number of illustrations based on The Open University and the University of Oxford. Despite being far apart in relation to funding, they are close together at the top of rankings based on the NSS. They have achieved this using completely different educational practices, but these practices embody some important educational principles. They are so different from other institutions that there can be little sense in which they can be compared, or copied, except at the level of principles. It is these principles that the report seeks to highlight, because they illuminate important dimensions of quality.

3. THE NATURE OF DIMENSIONS OF QUALITY

3.1 Conceptions of quality

'Quality' is such a widely used term that it will be helpful first to clarify the focus of this report. There have been a number of attempts to define quality in higher education, or even multiple models of quality (e.g. Cheng and Tam, 1997). The most commonly cited discussion of the nature of quality in higher education in the UK is that by Harvey and Green (1993), and their helpful nomenclature will be employed here. First, quality is seen here as a relative concept – what matters is whether one educational context has more or less quality than another, not whether it meets an absolute threshold standard so that it can be seen to be of adequate quality, nor whether it is reaches a high threshold and can be viewed as outstanding and of exceptional quality, nor whether a context is perfect, with no defects. What is discussed here is the dimensions that are helpful in distinguishing contexts from each other in terms of educational quality.

Quality may also be seen to be relative to purposes, whether to the purposes and views of customers or relative to institutional missions. This report does not take customer-defined or institutionally defined conceptions of quality as its starting point. Rather an effort will be made to focus on what is known about what dimensions of quality have been found to be associated with educational effectiveness in general, independently of possible variations in either missions or customers' perspectives. The report will then return to the issue of institutional differences and will comment in passing on differences between students in the meaning that can be attached to quality indicators such as 'drop-out'.

A further conception of quality made by Harvey and Green is that of quality as transformation, involving enhancing the student in some way. This conception comes into play when examining evidence of the educational gains of students (in contrast to their educational performance). This transformation conception of quality is also relevant when examining the validity of student judgements of the quality of teaching, where what they may want teachers to do may be known from research evidence to be unlikely to result in educational gains. What is focused on here is not necessarily what students like or want, but what is known to work in terms of educational effectiveness.

It is usual to distinguish between quality and standards. This distinction is most relevant in Section 6.1 on student performance, where the proportion of 'good

degrees' can be seen to be in part a consequence of the qualities of what students have learnt and in part a consequence of the standards applied in marking the products of student learning. This report will not focus on standards that have been the subject of much recent debate; for example, concerning the operation of the external examiner system.

3.2 Categorising dimensions of quality: presage, process and product

Education is a complex business with many interacting dimensions of quality in many varied contexts. To understand what is going on it is necessary to have a way of conceiving of the variables involved and of organising and interpreting studies of the relationships between these variables. This report will adopt the commonly used '3P' model (Biggs, 1993), which approaches education as a complex system with 'Presage', 'Process' and 'Product' variables interacting with each other. The '3P' model is essentially the same as that used by large-scale studies in the US (e.g. Astin, 1977, 1993): the 'Input-Environment-Output' model. Presage variables are those that exist within a university context before a student starts learning and being taught, and include resources, the degree of student selectivity, the quality of the students, the quality of the academic staff and the nature of the research enterprise. None of these presage variables determine directly how the educational process may be conducted, although they often frame, enable or constrain the form education takes.

Process variables are those that characterise what is going on in teaching and learning and include class size, the amount of class contact and the extent of feedback to students. Process variables also include the consequences of variables such as class size for the way students go about their learning, e.g. how those variables impact on the quantity and quality of their study effort and their overall level of engagement.

Product variables concern the outcomes of the educational processes and include student performance, retention and employability. Products can also include psychometric measures of generic outcomes of higher education, such as students' ability to solve problems. In some studies the key product measure is not student performance, but educational gain: the difference between performance on a particular measure before and after the student's experience of higher education. The difference between performance and gain will be crucial in understanding dimensions of quality, as we shall see.

The categorisation of variables as presage, process or product is not always straightforward. For example, some process variables such as the level of student engagement may be related to other process variables, such class size, which may in turn be related to funding levels. Which are the presage variables and which the products? Class size is not seen as a presage variable in the 3P model as it is in part a consequence of policy decisions about how to use resources and in part

a consequence of educational decisions about teaching methods. The presage variable of resources does not necessarily predict either. Nor is student engagement conceived of in the 3P model as a product. Both class size and student engagement are conceived of as part of the processes that may influence education outcomes, which are categorised as products.

In examining the usefulness of potential performance indicators involved in 'league tables', presage, process and product variables have sometimes been subdivided into more categories within a more complex model (Finnie and Usher, 2005; Usher and Savino, 2006), but for the purposes of this report the simple 3P model will suffice.

This report examines a wide range of presage, process and product variables in turn, and in doing so identifies relationships that are known to exist between them.

4. PRESAGE DIMENSIONS OF QUALITY

This section considers four presage dimensions of quality: funding, staff:student ratios, the quality of teaching staff and the quality of students

4.1 Funding

Institutional funding predicts student performance to some extent. It predicts cohort size (Bound and Turner, 2005), and class size predicts student performance (see Section 5.1). Funding also affects the kind of teachers the institution can afford to undertake the teaching and this affects student performance (see Section 4.3). How much funding per student is allocated to the provision of learning resources also predicts student study effort, which in turn predicts student performance (see Section 5.2). However, funding predicts performance largely because the best students go to the best-resourced institutions and the quality of the students predicts their performance (see Section 4.4). A series of large-scale US studies have found little or no relationship between institutional funding and measures of educational gain (Pascarella and Terenzini, 2005).

Even the ability of institutional levels of funding to predict student performance is somewhat limited. A study in the US has compared groups of colleges with near identical funding per student and found graduate completion rates varying between 35% and 70% (Ewell, 2008), so the differences in what colleges do with their funding must be very wide. In addition, institutions with similar levels of performance display widely varying levels of funding with some receiving only 60% of the revenues per student that others receive, but achieving near identical performance on a whole range of outcome measures. Twenty institutions that had been identified as unusually educationally effective, in relation to student retention and levels of student engagement, have been matched, in terms of a whole raft of other potentially influential presage variables such as student selectivity, with a group of markedly less effective institutions. There was found to be no difference in the level of funding allocated to teaching between the effective institutions and the carefully matched less effective group (NCHEMS, 2003).

What distinguished these effective institutions was that the funding was used differently; for example, on faculty development (see Section 5.3.1), teaching and learning centres and academic support staff such as tutors and counsellors (see Section 5.7.3). These uses of funding were exactly what previous and subsequent studies have found to be characteristic of "a campus ethos devoted to student success" (Gansemer-Topf *et al.*, 2004). This issue of ethos and values is returned to in the conclusion.

4.2 Student:staff ratios

While at the level of the institution student:staff ratios (SSRs) may seem to be an inevitable consequence of funding levels, institutions in practice spend funds on buildings, on administration, on 'central services', on marketing, on teachers undertaking research, and so on, to very varying extents, rather than spending it all on teaching time. The doubling of tuition fees in the US in recent decades has not been accompanied by any overall improvement in SSRs, but has largely been used for administration and meeting accreditation requirements. Institutions spend very different proportions of their available funding on teachers. So SSRs might be seen to be a more direct indicator of educational quality than funding.

Low SSRs offer the potential to arrange educational practices that are known to improve educational outcomes. First, close contact with teachers is a good predictor of educational outcomes (Pascarella and Terenzini, 2005) and close contact is more easily possible when there are not too many students for each teacher to make close contact with. Low SSRs do not guarantee close contact, as Harvard's recent self-criticism has demonstrated, but they do make it possible.

Second, the volume, quality and timeliness of teachers' feedback on students' assignments are also good predictors of educational outcomes (see Section 5.6), and again this requires that teachers do not have so many assignments to mark that they cannot provide enough, high-quality feedback, promptly. Again, low SSRs do not guarantee good feedback or feedback from experienced teachers. In the UK turn-round times for feedback may be a matter of local policy rather than driven by SSRs and turnaround times vary enormously between institutions (Gibbs and Dunbar-Goddet, 2009).

Third, while low SSRs do not guarantee small classes, they certainly make them possible, and class size predicts student performance (see Section 5.1 below).

However, once student entry characteristics are taken into account, educational gains have been found to be largely unrelated to SSRs (Terenzini and Pascarella, 1994). This suggests either that institutions with low SSRs are not exploiting their potential advantages through the use of effective educational practices or that SSR figures hide other variations, or both.

SSRs reported at institutional level do not necessarily give a good indication of the SSRs students actually experience. Patterns of work vary; for example, academics do a greater proportion of administration, with fewer support staff, in some institutions, effectively reducing their availability to students. They undertake more research in some institutions while the proportion of their research time funded by research income varies. The difference between students' years of study can be marked, with much greater funding per student characteristically being allocated to third-year courses than to first-year courses, leading to better SSRs and smaller classes in the third year (and the National Student Survey is administered in the third year). Furthermore institutions do not allocate funding to departments

in strict proportion to student numbers, but take advantage of well-recruiting and income-generating courses, and so increase their SSRs. The overall consequence is that effective SSRs, as students experience them, within a year of their study, within a department, may bear little resemblance to SSRs reported at institutional level.

There are also commonly expressed concerns about the accuracy and meaning of SSR data as reported to institutions, and as made public by HESA, due to variations in how they are calculated, complicated by variations in the nature of employment of teachers and how they are deployed (see Section 4.3 below).

4.3 Quality of teaching staff

Bald SSR data are unhelpful in that they disguise the realities of who the staff are with whom students have contact. For example, undergraduates at Yale often do not receive feedback from tenured faculty until their third year. In US research universities the teaching undertaken by graduate teaching assistants is a constant quality concern and is regularly cited in student exit surveys as their number one complaint about the quality of teaching.

An hour of a graduate teaching assistant may cost a fraction of an hour of a tenured academic, and most institutions are quick to exploit this. Recent surveys (HEPI, 2006, 2007) reveal wide variations between institutions in the proportion of teaching that students experience that has been undertaken by research students as opposed to tenured academics. The majority of small group teaching was found to be undertaken by teachers other than academics at Russell Group and pre-1992 universities. At the University of Oxford the extent to which students take a 'surface approach' to their study, emphasising only memorisation (see Section 5.5.2 below), is linked to the proportion of their tutorials taken by teachers other than College Fellows (Trigwell and Ashwin, 2004). A much lower proportion of teaching is undertaken by research students at Oxford than at other Russell Group universities.

> *I am currently an undergraduate at a Russell Group University... I have three hour-long lectures (taught by world-class professors) and three hour-long group seminars (taught by unpaid postgraduate students) a week. That is it. If we are going to compromise social mobility and academic freedom in order to fund our universities better, we damn well ought to receive a standard of education that's worth the sacrifice.*
> —Letter to *Education Guardian*, 25 May, 2010

In teaching-oriented institutions with few research students who might do the teaching, a significant proportion of teaching may be undertaken by what the US terms 'adjunct faculty' who may have portfolio teaching careers spanning a number of institutions, with an office in none of them. In vocational and creative arts courses

students may experience a significant proportion of their teaching undertaken by professionals who may spend only a few hours a week on campus. Adjunct faculty and professionals may bring special characteristics and talents to their work, but they may never have the opportunity to develop a course over several years so that it provides a rich and effective learning environment for students. They may not be paid to meet students out of class or to provide detailed comments on their assignments. They may never attend departmental meetings or fully understand how the degree programme works to which their teaching is supposed to contribute. Students may not know who the full-time tenured academics are and who are not – they will simply experience teaching of varying quality and commitment. While the HEPI data uncover variation in the proportion of teaching undertaken by research students – who students can see are mainly young – it is difficult for student surveys to identify the proportion of teaching undertaken by adjunct faculty and HEPI data do not illuminate this issue.

In the US the proportion of part-time and non-tenure-track 'adjunct faculty' has increased enormously in public institutions. Institutions with higher proportions of adjunct faculty have lower graduation rates when other factors are held constant. An increase of 10% in part-time faculty is associated with a reduction of 3% in graduation rates. Within institutions, first-year students taught to a greater extent by adjunct faculty are less likely to persist into the second year (Ehrenberg, 2006).

4.4 Quality of students

In the US, by far the best predictor of students' educational outcomes whether the measure is grades, a psychometric test of principled reasoning, or career success, is their school SAT score when they enter college, with correlations in the range 0.85 to 0.95. In other words up to 90% of all variation in student performance at university can sometimes be explained by how they performed before they entered university. In the UK the link is less strong, but there has for decades been clear evidence of the extensive impact of schooling on student performance in higher education, both in terms of school leaving grades and type of school (Smith and Naylor, 2005). In the UK students from independent schools perform less well than do students from state schools with equivalent entry grades (Hoskins *et al.*, 1997; Smith and Naylor, 2005).

The question that then arises is whether any of this enhanced performance is due to qualities of the institution other than their ability to be highly selective on entry. Large-scale longitudinal studies of a diverse range of institutions have tested students about their academic behaviour and experience (including their engagement, see Section 5.5.3 below) from a total of nearly 300 colleges and involving data from nearly 80,000 students (for a summary of this work see Kuh and Pascarella, 2004). These studies have found very little relationship between educational selectivity (i.e. quality of student intake) and the prevalence of what are known to be educationally effective

practices. Selectivity was found to be negatively associated with some practices, such as the amount of teacher feedback to students, and even where there were found to be small positive relationships (for example with higher expectations on students), selectivity only accounted for 2% of the variance in educational practices.

It might be argued that selective institutions do not need special educational practices because their students are able enough to engage themselves. However, the degree of selectivity does not predict the degree of student engagement – students are just as engaged (or unengaged) in non-selective institutions (Pascarella *et al.*, 2006). So while league tables in the UK invariably include A-level point scores as an indicator of educational quality, if the US evidence is anything to go by they tell us almost nothing about the quality of the educational process within institutions or the degree of student engagement with their studies.

It might be argued that there are educational benefits to a student of being surrounded by other able students. This could raise students' expectations of themselves (one the of 'Seven Principles')[1], and it is known that in group work it is the previous educational attainment of the best student in the group that best predicts the group grade, not the average level of prior attainment or the level of the weakest student (Gibbs, 2010). We would then need to look at the extent to which the educational process maximises how students could gain from each other, for example through collaborative learning. The extent of collaborative learning is a good predictor of educational gains (the 'Seven Principles' again). However, it will not help a student much if the other students are highly able if they then engage largely in solitary competitive learning. The US data cited above make it clear that students are not more likely to be involved in collaborative learning, or to be engaged by it, in institutions with more selective entry, in which the students are more able.

Students bring more to higher education than their A-level scores. It is likely that their cultural capital, their aspirations, self-confidence and motivations all influence their performance and interact with teaching and course design variables.

1 The 'Seven Principles of Good Practice in Undergraduate Education' (Chickering and Gamson 1987a, 1987b, 1991) are based on a very wide review of empirical evidence, and have been used widely in the US and elsewhere as guides to the improvement of university teaching. The principles are that good practice: encourages student-faculty contact, encourages cooperation among students; encourages active learning; provides prompt feedback; emphasizes time on task; communicates high expectations; and respects diverse talents and ways of learning.

5. PROCESS DIMENSIONS

This section considers the effects on educational effectiveness of class size, class contact hours, independent study hours and total hours, the quality of teaching, the effects of the research environment, the level of intellectual challenge and student engagement, formative assessment and feedback, reputation, peer quality ratings and quality enhancement processes

5.1 Class size

Meta-analysis of large numbers of studies of class-size effects has shown that the more students there are in a class, the lower the level of student achievement (Glass and Smith, 1978, 1979). Other important variables are also negatively affected by class size, such as the quality of the educational process in class (what teachers do), the quality of the physical learning environment, the extent to which student attitudes are positive and the extent of them exhibiting behaviour conducive to learning (Smith and Glass, 1979). These negative class-size effects are greatest for younger students and smallest for students 18 or over (*ibid.*), but the effects are still quite substantial in higher education. Lindsay and Paton-Saltzberg (1987) found in an English polytechnic that "the probability of gaining an 'A' grade is less than half in a module enrolling 50-60 than it is in a module enrolling less than 20" (p218). All subsequent UK studies have reported sizable negative correlations between class size (as measured by the number of students registered on a course) and average student performance, in most but not all subjects, and in most but not all contexts (Gibbs *et al.*, 1996; Fearnley, 1995). Large classes have negative effects not only on performance but also on the quality of student engagement: students are more likely to adopt a surface approach in a large class (Lucas *et al.*, 1996) and so to only try to memorise rather than attempt to understand (see Section 5.5.2 on depth of approach to learning).

At a micro-level there is evidence that the educational process is compromised as class size increases. In higher education discussion groups, for example, a whole range of things go wrong as class size increases. There is a much lower level of participation by all but a minority of students and the contributions that students do make tend to concern clarification of facts rather than exploration of ideas (Bales *et al.*, 1951).

US research shows that higher education students give lower overall ratings to teachers of large classes (Wood *et al.*, 1974; Feldman, 1984). However, there are reasons to question the meaning of this finding. The same teachers are given higher ratings when they teach smaller classes. As such ratings of teachers are relatively reliable and stable this suggests that students' ratings of teachers in large classes are reflecting something other than the teachers themselves. A qualitative study of students' experience of large classes (Gibbs and Jenkins, 1992) has thrown light on variables other than the teaching. There may be intense student competition for limited library and other resources in large classes, and teachers may have to rely on a few textbooks if students are to read anything. The amount and promptness of feedback on assignments is likely to decline, as teacher time is squeezed. The nature of assessments may change from engaging open-ended projects to quick tests, as marking otherwise takes too long. Close contact with teachers outside of class and access to remedial tutoring and advice may be more limited. Large classes may be associated with weak social cohesion, alienation and a breakdown in social behaviour, leading to cheating, hiding library books, and so on. All this is more to do with what happens outside of class on courses with large enrolments, rather than what happens in class, but it is classroom activity that is the focus of most school-based research and US higher education research. Where out-of-class studying is the major component of student learning the crucial variable may be course enrolment rather than class size. US data show that cohort size is strongly negatively correlated with student performance (Bound and Turner, 2005).

Another difference between school and higher education in relation to class-size effects is that in higher education the range of class sizes being studied is very much wider: perhaps 20 to 1,000 instead of 10 to 40 in schools. Different variables inevitably become prominent in such very large classes. In school, students may experience all their classes as much the same size. In higher education what may matter most is not the size of the largest lecture that is attended on any particular course but the size of the smallest seminar group or problem class that they attend within the same course. Open University students may attend a course with an enrolment of over 10,000, but they usually only experience a tutor group of 24, and each tutor usually has only one tutor group so they can get to know students individually. At the Open University it would probably make a difference if this tutor group was 12 or 48 but not if total enrolment was 500 or 20,000.

Classrooms used for specialist purposes, such as laboratories and studios, usually limit the number of students it is possible to teach at once, regardless of how many students have enrolled, and although laboratories have become much larger, there are limits to class-size effects while within the lab. However, increased enrolments with fixed specialist spaces have the inevitable consequence of reducing the amount of time students have access to these specialist facilities. This has transformed art and design education. Instead of students 'owning' a permanent space they can take time to become creative and competent in, when enrolment increases they visit a shared

space occasionally. The number of students in the studio at any one time may not have changed much but art students' experience has been changed out of all recognition. Gibbs *et al.* (1996) found that in Art, Design and the Performing Arts, each additional 12 students enrolled on a course gave rise to a decline of 1% in average marks.

Negative class-size effects are not inevitable and a certain amount is known about how to support good quality learning despite large classes (Gibbs and Jenkins, 1992). The Teaching More Students initiative in the early 1990s trained 9,500 polytechnic and college lecturers on the assumption that such improvements were possible despite larger classes (Gibbs, 1995). The National Centre for Academic Transformation in the US has helped scores of institutions to redesign large-enrolment, first-year courses. They have shown that it is possible to improve student outcomes while reducing teaching contact time and reducing funding. The Open University has retained the number and nature of assignments per course, the amount, quality and turnaround time of feedback from tutors, and the small size of tutor groups, through strict course approval rules, with course enrolments that are seldom below 500.

The conundrum, of course, is that in the UK overall student performance has increased at the same time that overall class size has increased. This issue is addressed in Section 6.1.

5.2 Class contact hours, independent study hours and total hours

The number of class contact hours has very little to do with educational quality, independently of what happens in those hours, what the pedagogical model is, and what the consequences are for the quantity and quality of independent study hours. Independent study hours, to a large extent, reflect class contact hours: if there is less teaching then students study more and if there is more teaching students study less, making up total hours to similar totals regardless of the ratio of teaching to study hours (Vos, 1991). However, some pedagogic systems use class contact in ways that are very much more effective than others at generating effective independent study hours. A review of data from a number of studies by Gardiner (1997) found an average of only 0.7 hours of out-of-class studying for each hour in class, in US colleges. In contrast each hour of the University of Oxford's tutorials generate on average 11 hours of independent study (Trigwell and Ashwin, 2004) and Oxford's students have been found to put in the greatest overall weekly effort in the UK despite having comparatively fewer class contact hours (HEPI, 2006, 2007). What seems to matter is the nature of the class contact. 'Close contact' that involves at least some interaction between teachers and students on a personal basis is associated with greater educational gains (Pascarella, 1980) independently of the total number of class contact hours (Pascarella and Terenzini, 2005); the provision of close contact is one of the 'Seven principles of good practice in undergraduate education' (Chickering and Gamson,1987a, 1987b, 1991).

Seven principles of good practice in undergraduate education

Good Practice Encourages Student-Faculty Contact
Good Practice Encourages Cooperation among Students
Good Practice Encourages Active Learning
Good Practice Gives Prompt Feedback
Good Practice Emphasizes Time on Task
Good Practice Communicates High Expectations
Good Practice Respects Diverse Talents and Ways of Learning

—Chickering and Gamson, 1987a

As an illustration of the lack of relationship between class contact hours and outcomes, medical education worldwide has migrated from traditional didactic pedagogies, characterised by large numbers of large class lectures, towards problem-based pedagogies, characterised by a much smaller number of small interactive classes, usually within the same resource envelope. This change has been accompanied by a substantial increase in independent learning hours and evidence of greater pedagogical effectiveness measured in a variety of ways (Dochy *et al.*, 2003). Similarly the Open University has managed to top the National Student Survey league tables while having very much the lowest class contact hours in the UK.

This is not the same as arguing that you can cut class contact hours from an existing unchanged pedagogy without making any difference to student learning, or that increasing hours will make no difference. If students read primarily in order to discuss what they have read in a seminar, and the seminar is taken away, then they will probably read a good deal less and learn less as a consequence. Very little class contact may result in a lack of clarity about what students should be studying, a lack of a conceptual framework within which subsequent study can be framed, a lack of engagement with the subject, a lack of oral feedback on their understanding, and so on. It depends what role the class contact is performing. What matters is the quantity and quality of engagement generated by the particular uses to which class contact is put.

What seems to be more important than class contact hours is the total number of hours that students put in, both in and out of class (the quality of that effort is considered in Section 5.5 below). The number of hours that students study out of class is usually measured in one of two ways: by asking students to estimate, retrospectively, how much they think they have been studying, often 'on average', or by asking students to keep a log, as they study, of what they do with their time, over a relatively short period of time (such as a week). While logs are likely to be more accurate than retrospective estimates, they also run the risk of changing student study behaviour as students become more aware of their pattern and level of effort. Indeed keeping

a log is a common learning activity on study skills courses. When asking students to estimate their study hours retrospectively, the form of the question used varies between different surveys and the timing of the surveys varies in relation to how long ago students are attempting to remember or how wide a spread of courses they are being asked to make average estimates across. Students who attend less and study less may be missed by surveys while conscientious students who attend more and study more may be more likely to return surveys. The impact of such potential biases is not well researched and the reliability of study-hours data is not known.

The question: 'Are higher study hours associated with better student learning and performance?', can be posed in two rather different ways. First: 'Are the students who study longer hours the ones that perform best?' The answer to this question is not straightforward (Stinebrickner and Stinebrickner, 2008), because very able students may be able to meet assessment requirements without having to study very hard, while less able students may put in many hours unproductively (Ashby *et al.*, 2005). There is also evidence that students who, inappropriately, take a 'surface' approach to their studies (see Section 5.5.2 below) find this so unproductive that they gradually reduce their effort after initially working hard and end up studying fewer hours than students who take a 'deep' approach (Svensson, 1977).

If, however, the question is framed differently as: 'If a student were to study more hours, would they perform better?' or even 'If average study hours on a degree programme were higher, would average performance be higher?', the answer is much more clearly 'Yes'. 'Time on task' is one of the evidence-based 'Seven Principles of Good Practice in Undergraduate Education (Chickering and Gamson, 1987). The reasonable assumption here is that if you don't spend enough time on something then you won't learn it, and that increasing the number of hours students spend studying is one of the most effective ways of improving their performance. North American research and development work on 'student engagement' (see Section 5.5.3 below) uses student effort as an important indicator of engagement.

The Bologna process has used total student effort (class contact hours plus independent study hours) as its metric for defining the demands of a Bachelors degree programme, set at 1,500 to 1,800 hours a year: 4,500 to 5,200 hours over three years. A series of studies have found that UK students' total weekly effort in hours is lower than in comparison with either the particular European countries studied or in comparison with overall European norms (Brennan *et al.*, 2009; Hochschul-Informations-System, 2005; Sastry and Bekhradnia, 2007; Schomburg and Teichler, 2006). These findings deserve to be taken seriously because they are relatively consistent across different studies and methodologies, carried out in different countries, or across repetitions of the same study in different years.

It should be possible to iron out gross differences between institutions and subject areas, as the number of study hours per credit, and hence the number of hours required for a Bachelors programme, are clearly defined in course documentation. However, the range in weekly study effort between English

institutions, within subjects, found in the HEPI studies is wide, for example from 14 hours a week to nearly 40 hours per week within Philosophy (Sastry and Bekhradnia, 2007). Differences between subjects are also wide. Broad differences in total study hours between science and technology programmes (which tend to have both high class contact hours and weekly demands for work such as problem sheets and laboratory reports) and the humanities (which tend to have both lower class contact hours and less regular assignments such as essays) are well known and have been reported frequently over the years (e.g. Vos,1991). However, the differences between subjects identified by the HEPI surveys are substantial, with some subjects having national average weekly study efforts of only around 20 hours per week. Twenty hours per week within the comparatively short UK semesters equates to around 500 hours a year: one third of the minimum specified under the Bologna Agreement. To achieve the Bologna specification of a minimum of 4,500 hours for a Bachelors programme, students in these subjects in the UK would have to study for nine years. Differences on this scale cannot easily be argued away by claiming that UK students are somehow inherently superior or that UK educational practices are somehow inherently more efficient, in the absence of any evidence to back up such claims.

A survey of international students who have experienced both a UK higher education institution and another EU higher education institution (Brennan *et al.*, 2009) found that such students are more likely to rate UK Bachelors programmes as 'less demanding' and less likely to rate them as 'more demanding', a finding that does not justify the lower number of hours involved. UK students have been reported to have done more work than was required of them to a greater extent than in any other European country (*ibid.*). Yet the total number of hours studied in the UK is still below European norms, which suggests that the UK requirements must be lower.

If it were the case that less able students needed to study more, then one would find the larger study hours figures in institutions that have students with weaker educational backgrounds. Instead the reverse is the case, with higher weekly study hours reported in institutions with students with the strongest educational backgrounds (Sastry and Bekhradnia, 2007). The most likely explanation therefore is that the demands made on students are different in different institutions, and that even weaker students are able to meet these demands while studying, in some institutions, and in some subjects, a third of the hours the Bologna Agreement specifies.

There are a number of possible explanations of why such students might study few hours:

— Some degree programmes rely heavily on coursework assessment, with over 90% of marks on some degree programmes coming from coursework rather than on examinations. While this may have come about for sound educational reasons, it may allow students to study only one or two topics in depth per course, because that is all there are the teaching resources to mark, reducing student effort in weeks when there are no formal

assessment demands. Students have been found to work more regularly and cover the syllabus to a greater extent when there is a higher proportion of marks from examinations (Gibbs and Lucas, 1997).

— High levels of detail in course specifications, of learning outcomes and assessment criteria, in response in part to QAA codes of practice, allow students to identify what they ought to pay attention to, but also what they can safely ignore. A recent study has found that in such courses students may narrow their focus to attention to the specified assessed components at the expense of everything else (Gibbs and Dunbar-Goddet, 2007). Students have become highly strategic in their use of time and a diary study has found students to progressively abandon studying anything that is not assessed as they work their way through three years of their degree (Innis and Shaw, 1997).

— Studies of the effects of paid work undertaken by full-time students have reported a substantial reduction in study hours in relation to the extent of their paid work, although effects vary somewhat between contexts and degree programmes (Carney *et al.*, 2005; Curtis and Williams, 2002; Ford *et al.*, 1995; Hunt *et al.*, 2004; Paton-Salzberg and Lindsay, 1993). A typical study is that of Svanum and Bigatti (2006), who found that paid work reduced course effort and reduced grades. In the US, students respond to the need to 'work their way through college' by taking fewer credits at a time and gaining their degree over more years (and across several institutions) as a consequence. In the UK full-time students seem to assume that they can undertake paid work while meeting the demands of a full-time course and still graduate in three years.

— Students who live at home, rather than on a residential campus, are likely to experience competing demands on their time, and less social and academic integration (Tinto, 1975). The institutions in the UK with the lowest average study hours include universities in urban conurbations with a substantial proportion of students living at home.

— The universities with low average study hours are often also institutions with low annual investment per student in libraries and other learning resources. This would make it more difficult for students to gain access to the resources they need for their study: the book will be out and the study space with a computer will be occupied. Data from HESA, HEPI and the National Student Survey have been analysed for the purpose of the current report on this issue. They showed that institutional funds allocated to learning resources, per student, predict total student learning hours (with correlations of +0.45 for the social sciences and humanities subjects analysed). Funding for learning resources also predicts average students' responses to the NSS question on the quality of learning resources,

although less well. The institution with the highest weekly average study hours also has the greatest annual investment in learning resources and the highest National Student Survey ratings for 'learning resources': the University of Oxford.

While the focus of this report is on undergraduate programmes, there has recently been a good deal of attention paid to the relative quality of UK Masters level courses, given that they are usually planned to be considerably shorter in duration than their mainland European counterparts. For example, the Chair of the UK Council for Graduate Education has argued that these gross differences do not matter because UK Masters courses are 'more intensive', and claimed that the overall amount of learning time is roughly equal between the UK and mainland Europe (Olcott, 2010). This unsubstantiated claim could be checked by repeating the Sastry and Bekhradnia undergraduate study of study hours, cited above, in Masters courses and adding questions to measure the extent to which students take a deep approach to their studies (see Section 5.5.2 below).

5.3 Quality of teaching

5.3.1. Quality of teaching: experience and training

Teachers who have teaching qualifications (normally a Postgraduate Certificate in Higher Education, or something similar) have been found to be rated more highly by their students than teachers who have no such qualification (Nasr *et al.*, 1996). This finding was in a context where obtaining such a qualification was largely voluntary, and those who have the qualification might be considered to be different in some way from those who have not, and this could be argued to invalidate the comparison. The difference might concern the extent of professionalism or commitment to teaching, but nevertheless there was no control group in the study. A longitudinal study that overcomes this objection has examined the impact over time on students' ratings of their teachers, and on teachers' thinking about teaching, of (mainly) compulsory initial training during their first year of university teaching, in eight countries. It found improvements on every scale of the 'Student Evaluation of Educational Quality', a questionnaire developed in the US (Marsh, 1982) and tested for use in the UK (Coffey and Gibbs, 2000), and improvements in the sophistication of teachers' thinking (as measured by the 'Approaches to Teaching Inventory', a measure of teaching that predicts the quality of student learning, Trigwell *et al.*, 2004). This improvement in measures of teaching quality could not be attributed to mere maturation or experience as teachers in a control group in institutions without any initial training were found to get worse over their first year, on the same measures (Gibbs and Coffey, 2004). The quality of school teaching in the UK is in part underpinned by a belief that initial or in-service training is essential and the, admittedly limited, evidence suggests that the same belief would be justified in UK higher education.

5.3.2. Quality of teaching: research record

There is little or no relationship between measures of the quality or quantity of teachers' research and measures of the quality of their teaching (for a review of 58 studies of the evidence, see Hattie and Marsh, 1996).

> *…the common belief that teaching and research were inextricably intertwined is an enduring myth. At best teaching and research are very loosely coupled.*
> —Hattie and Marsh, 1996, p529

Some excellent researchers make excellent teachers and some do not. Despite critiques of the measures of research and teaching that are normally used, none of the critics have managed to develop or use alternative measures that demonstrate a relationship between research and teaching. A minority of undergraduate students have been reported to value their teachers being active researchers provided this does not interfere with their studies (for example, through their teacher being absent while undertaking research) (Lindsay *et al.*, 2002), but there is no evidence that this improves their learning.

5.3.3. Quality of teaching: judged by students

Despite the common disparagement of student ratings of teachers, they can be highly reliable (in the sense that students agree with each other about who the best teachers are, agree with teachers' peers, and make the same judgements on different occasions) and relatively immune from biases of various kinds. Students can readily tell the difference between teachers they like and teachers who they think are good teachers, and the common criticism that student feedback is simply a popularity parade is largely unfounded (see Marsh, 1987, for a review of the vast literature on the reliability and validity of student feedback questionnaires). These observations are based on the use of thoroughly developed and tested feedback questionnaires. However, in the UK almost all such questionnaires are 'home-grown' and are likely to be of doubtful reliability and open to all kinds of biases. Different questionnaires are used in different institutions and there is therefore no basis for comparison of the quality of teachers between institutions or subjects. If student ratings of teachers were considered a useful indicator of comparative quality then there would need to be an agreement for everyone to use the same questionnaire, such as the Student Evaluation of Educational Quality (SEEQ) (Marsh, 1982; Coffey and Gibbs, 2000).

Several well-developed, reliable, student feedback questionnaires have reasonable levels of validity in that scores on scales within the questionnaires (involving adding ratings from clusters of similar questions) predict aspects of student learning process

(such as the extent of studying following teaching), learning outcomes (such as grades) and other worthwhile consequences (such as the likelihood of students choosing to study further courses with the same teacher). The proportion of variance in such measures of the products of good teaching, that is explained by student ratings, varies across different questionnaire scales and different measures of products, but it is usually high enough to take measures of teaching based on student ratings seriously (Abrami *et al.*, 1990).

There is an important distinction to be made here between student ratings of the extent to which teachers engage in activities that are known to improve learning (such as providing enough prompt feedback on assignments), which tend to be reliable and valid, and global judgements of whether teaching is 'good', which are open to all kinds of subjective variation in the interpretation of what 'good' means. Students also change over time in their sophistication as learners, for example in their conception of learning (Säljö, 1979) and in their conception of knowledge (Perry, 1970). As they change, so their conceptions of what 'good teaching' consists of evolve (Van Rossum *et al.*, 1985). What an unsophisticated student might consider to be good might consist of the teacher providing all the content in lectures and then testing for memory of that content, while a more sophisticated student might see good teaching as involving supporting independent learning and the development of a personal stance towards knowledge. What unsophisticated students want their teachers to do is often bad for their learning and responding to their global ratings uncritically is not the way to improve quality. When a survey reports a single global rating of the extent to which student think all the teaching over three years is simply 'good', these very different student conceptions of good teaching are muddled together and the average rating is then very difficult to interpret. In contrast so-called 'low inference' questions that refer to specific teacher behaviours, such as the promptness of their feedback, are much easier to interpret.

5.4 Research environment

As we have seen above (in Section 5.3.2) there is no relationship between measures of an individual academic's research and measures of their teaching. However, it could be argued that it is not individual researchers' teaching that matters here, but the research environment generated by the majority of teachers in a department being research active. This might be considered a presage variable, but as we shall see, what matters is the educational process, not prior research performance.

At the level of departments within an institution the situation is the same as it is at the level of individual teachers. The best research departments may or may not be the best teaching departments: there is no correlation between measures of a department's research and measures of its teaching (Ramsden and Moses, 1992). There are suggestions that there may be relationships between the extent of

research intensiveness of a department and students' sense of benefit, in a way that is not mediated by the teaching, but this is in any case only for some students (e.g. Dunbar-Goddet and Trigwell, 2006), and this relationship appears to be a function of the focus of attention of the students rather than a consequence of different or better teaching: if students are interested in scholarship they will be more likely to say that they benefit from a scholarly environment.

At the level of whole institutions, those where there is a strong orientation towards research often reveal a weak emphasis on teaching, and vice versa – there is a strong negative relationship in relation to institutional priorities and this has measurable effects on educational gains:

> *… a college whose faculty is research-orientated increases student dissatisfaction and impacts negatively on most measures of cognitive and affective development.*
> —Astin, 1993, p363

There is evidence that the research environment can impact positively on undergraduate students, in a way that can be measured, where maximising the benefits to undergraduates of research strengths has been made a deliberate priority. For example, MIT enables about 80% of its undergraduates to engage in a real research project as an intern or junior research student placed with a research group, and they have good evidence concerning how students benefit (Bergren *et al.*, 2007). The National Science Foundation in the US has similar evidence across many institutions that such initiatives:

> *… have a major impact on most participants' confidence and their understanding of research-related issues, increase their interest in careers in research and science and engineering, and lead them to raise their degree expectations.*
> —Ramaley, 2004, cited in Jenkins, 2004

(See also Bauer and Bennett, 2003; Hathaway *et al.*, 2002.) The key point here is that such benefits have to be deliberately engineered – they do not accrue by magic simply because research is going on as well as teaching. The institutional indicator of quality in these studies is the existence of an undergraduate research opportunities scheme, not the strength of the institution's research. Similarly the positive relationship found at the University of Oxford between students' experience of research-active staff and the extent to which they take a deep approach to learning (Trigwell, 2005) is a consequence of the collegial system fostering active inclusion in a community of (research) practice, not simply of the existence of a research enterprise.

For these reasons departmental RAE scores or other measures of research activity or performance in the environment students study within are not, on their own, valid indictors of educational quality.

5.5 Level of intellectual challenge

It can be argued that an institution, or even an entire higher education system, is of higher quality because it offers a greater level of intellectual challenge. In this section the notion of intellectual challenge is operationalised in three ways: challenge as determined by the level of the curriculum students are studying, challenge as indicated by the depth of approach students take to their studies, and challenge as reflected in the level of student engagement with their studies.

5.5.1. Level of the curriculum

In some subject areas with relatively hierarchical knowledge, formal course documentation can be very informative about the level of the curriculum. It is then possible to compare the level of intellectual challenge by examining course descriptions to see how far through the subject students are expected to progress. It is in principle possible to study curriculum documentation within well-defined disciplines, from different periods, in order to ascertain whether there has been a watering down of intellectual demands over time. There are several limitations to this kind of study:

- documentation changes over time in what is described and in what detail, and also differs between institutions;
- fashions change over time within subjects about what should be included, which makes changes of level of demand difficult to judge;
- even in supposedly hierarchical subjects it is sometimes difficult to agree what is necessary to study first, or what counts as more advanced. In Medicine, for example, problem-based learning may nowadays confront first-year students with problems of a complexity that in past curricula, that had clear divisions between pre-clinical and clinical stages, only fourth-year students encountered. However, the first-year students now tackle these complex problems at a lower level, with a lesser knowledge base. Calibrating the level of the problem and the level at which they are tackled is not straightforward.

Nevertheless there is probably more scope for determining quality by examining curriculum documentation than is currently exploited, provided it is undertaken by those within disciplinary communities that understand the complexities involved.

There are several typologies for educational objectives or learning outcomes (for example that of Bloom *et al.*, 1956) that could be used to compare the level of difficulty of what students are supposed to do with the content (for example, remember it, apply it or critique it). It is possible to distinguish levels of intellectual demand between different levels of BTEC courses in the same subject by comparing

the proportion of educational objectives specified at each of a number of levels (such as 'knowledge' and 'synthesis'). Higher levels of courses are supposed to have a higher proportion of their curricula objectives at a higher level in the hierarchy. Proposals for new courses can be turned down if too many objectives are set at too low a level.

Where a curriculum authority has the power to impose a common set of definitions and a common format of documentation, as BTEC does, this is just about possible to implement. Where each institution and each degree programme has the freedom to specify learning outcomes independently of any common conceptual framework or set of definitions, such comparisons become impossible, except in an impressionistic way. For learning outcomes that concern generic skills, such as 'communication' and 'group work', consistent judgement of levels against standards is even more difficult, independently of exemplars that go beyond curricula documentation. Alverno College in the US has demonstrated that even for such generic outcomes or capabilities, it is possible to define standards in unambiguous ways and to train teachers to use these standards in consistent ways, but unless Alverno's approach, or something like it, were to be adopted by all, comparison across contexts would still not be possible.

Nevertheless there is plenty of scope for subject associations to make more use of curriculum documentation in judging the intellectual demands of programmes, at least in principle. To judge them in practice it is also necessary to examine the products of student learning (see Section 7.7).

5.5.2. Depth of approach to studying

In the 1970s Ferenc Marton and his colleagues in Goteborg distinguished between a 'surface approach' to learning in which students intend to reproduce material, and a 'deep approach' in which students intend to make sense of material: a distinction between a focus of attention on the sign or what is signified. To illustrate the consequences for student learning outcomes, a student who takes a surface approach to reading an article with a principle-example structure (such as a case study) may remember the example, while the student who takes a deep approach is more likely to understand the principle (Marton and Wenestam, 1978). A surface approach has been demonstrated in a wide variety of studies to have depressingly limited and short-lasting consequences even for memory of facts. A deep approach is essential for long-term and meaningful outcomes from higher education (see Gibbs *et al.* (1982) and Marton *et al.* (1984) for overviews of this literature).

Students are not 'surface students' or 'deep students' – approach to learning is in the main a context-dependent response by the student to perceived demands of the learning context (Ramsden, 1979). The relevance to dimensions of quality is that it is possible to identify those features of courses that foster a surface or a deep approach. Students tend to adopt a surface approach to a greater extent when there

is, for example, an assessment system that rewards memorisation, such as superficial multiple-choice-question tests. In contrast students tend to adopt a deep approach, for example, when they experience good feedback on assignments, and when they have a clear sense of the goals of the course and the standards that are intended to be achieved. These influential characteristics of courses are the focus of the Course Experience Questionnaire (CEQ) (Ramsden, 1999), originally developed in studies at Lancaster University in the 1970s, through which students indicate the extent to which these course features are experienced. Reasonably close relationships have been found between scores on scales of the CEQ and the extent to which students take a deep and surface approach to their studies, and so CEQ scale scores that focus on certain course features can act as a rough proxy for educational outcomes, because approach predicts outcomes to some extent. The CEQ became the basis of the questionnaire used annually throughout Australian higher education to measure comparative quality of degree programmes, published in annual reports aimed at students. It has been used for some years within some institutions as a performance indicator for allocating a proportion of funding for teaching to departments, as at the University of Sydney. It has now been adopted nationally in Australia as one component of performance indicators for allocating over A$100 million of teaching funding (in 2008) to universities each year. It has become the driving force behind evidence-based institutional efforts to improve teaching that focus on course design rather than on individual teacher's skills (Barrie and Ginns, 2007). A modified version of the CEQ (the OSCEQ) has been used annually at the University of Oxford.

It is often assumed that the validity of the National Student Survey (NSS) is based on the same research and evidence. Up to a point this is true. However, the characteristic of students' intellectual engagement with their studying that best predicts their learning outcomes, the extent to which they take a deep approach, is not included as a scale in the NSS (and nor is it in the CEQ). Some characteristics of what have been found to be effective courses, such as concerning feedback, are included in the NSS. However, most of the scales of the original version of the CEQ that relate somewhat to the extent to which students take a deep approach, such as 'Clear Goals and Standards' or 'Appropriate Workload' are not included in the NSS (and neither are they in the most recent versions of the CEQ). In fact both questionnaires lack most of the scales that would strengthen their validity. The missing scales are currently included as options in both questionnaires, but this means that comparable data are not published or available for comparison between institutions or courses.

Even some of these missing scales have a somewhat tenuous claim to validity today. For example in the 1970s it was found that if students were grossly overburdened then they might abandon a deep approach and adopt a surface approach to their studies. However, 30 years later excessive workload seems a distant memory (see Section 5.2), so the 'Appropriate Workload' scale no longer seems likely to predict, to a worthwhile extent, which students will adopt a surface approach, and hence their learning outcomes.

There have been no recent studies to confirm the original findings concerning relationships between features of courses, student responses and learning outcomes in current contexts. There have been no direct studies of the validity of the NSS in relation to its ability to predict educational gains. There have been no studies that demonstrate that if evidence-based practices are adopted, and NSS scores improve, this will be associated with improved educational gains. For that kind of evidence we have to look to measures of student engagement.

5.5.3. Student engagement

There is currently a high level of interest in the US in measuring student engagement as a crucial indicator of educational quality, embodied in the use of a questionnaire: the National Survey of Student Engagement (NSSE). The research underlying the NSSE, and its wide-scale use and perceived value, has involved three main stages.

First, very large-scale studies examined a huge range of process variables on campus and also assessed a huge range of measures of educational gain, across a very wide range of courses and institutions, to identify which process variables relate to any of the measures of gain. No research encompassing so many variables at once, or on a similar scale, has been conducted outside of the US. These vast studies have been repeated over three decades and very similar findings have identified the same few process variables that are most closely linked to educational gains (Pascarella and Terenzini, 2005). Essentially the crucial variable is 'student engagement' and it has proved possible to identify the process variables involved in engaging students, such as the level of academic challenge, the extent of active and collaborative learning and the extent and quality of student-faculty interaction. These process variables are prominent in the influential evidence-based 'Seven principles of good practice in undergraduate education' elaborated above (Chickering and Gamson, 1987, 1991).

These principles have been used by many institutions in the US as indicators of quality in reviewing and improving their educational practices, and there has been much pooling of institutional experience in using these principles. Pascarella *et al.* (2008) list a plethora of studies that demonstrate that if you act on the 'Seven principles of good practice' then this improves student outcomes.

Second, a questionnaire has been developed through which students can indicate the extent to which these crucial process variables have been experienced and the ways in which they are engaged in their studying: the National Survey of Student Engagement (NSSE, 2007). This questionnaire has been used widely – by 774 universities and colleges in 2008 alone. It has also been widely used to monitor student engagement as changes have been made to educational practices in attempts to improve student learning. Sufficient studies of appropriate rigour have been published to make it possible to bring the data together and make generalisable conclusions. Meta-analysis of studies of innovation and change support the original

identification by the earlier research of a predictable range of educational processes that, when adopted or enhanced, make a measurable difference to students engagement (Carinin *et al.*, 2006).

The third stage of this research has involved validity studies – examining if there is a relationship between engagement and educational outcomes: between scores on the NSSE and both student performance and psychometric measures such as of effective reasoning and problem solving, 'inclination to inquire' and lifelong learning. A number of studies have shown clear links between NSSE scores and outcomes such as first-year success (e.g. LaNasa *et al.*, 2007). Pascarella *et al.* (2010) have gone further and in a much larger study, involving 19 colleges of a wide variety of types, demonstrated strong relationships between NSSE scale scores and a range of educational gains, involving before and after measurements. The authors concluded:

> *NSSE results regarding educational practices and student experiences are good proxy measures for growth in important educational outcomes.*

In other words if you want to know the 'value added' by students' higher education experience then the NSSE will provide a good indication without needing to use before and after measures of what has been learnt.

It is interesting to note, with reference to the self-imposed limitations of the NSS and CEQ, that the scale on the NSSE that has the closest relationship with educational gains concerns 'deep learning' (Pascarella *et al.*, 2008).

5.6 Formative assessment and feedback

The educational intervention in schools that has more impact on student learning than any other involves improving formative assessment and especially the provision of more, better and faster feedback on student work (Black and Wiliam, 1998; Hattie and Timperley, 2007). 'Good practice provides prompt feedback' is one of the evidence-based 'Seven principles of good practice in undergraduate education' (see above). On degree programmes where the volume of formative assessment is greater, students take a deep approach to their studies to a greater extent (Gibbs and Dunbar-Goddet, 2007) and deep approach is a good predictor of learning outcomes (see Section 5.5.2 above). Enhanced feedback can also improve student retention (Yorke, 2001).

The number of occasions during a three-year Bachelors programme in the UK on which students are required to undertake an assignment purely for the purpose of learning, with feedback but without marks, varies widely between institutions. One study has found a range from twice in three years at one English university to over 130 times at another (Gibbs and Dunbar-Goddet, 2009). In another UK study using the same assessment audit methodology (TESTA, 2010), the volume of

written feedback on assignments over three years varied from below 3,000 words per student to above 15,000 words, and for oral feedback varied from 12 minutes per year per student to over ten hours per year (Jessop *et al.*, 2010). These are much wider variations between institutions than exist in their funding per student, their SSRs, their class contact hours or their independent study hours. The issue addressed by the NSS that reveals the greatest area of student disquiet is feedback.

As resources per student have declined there have been economies of scale in teaching that are difficult to achieve in assessment: assessment costs go up pretty much in proportion to the number of students. This places enormous time pressures on teachers. Quality assurance systems in most institutions have not prevented the volume of formative assessment from declining substantially, despite the QAA *Code of practice*. An exception is The Open University where the number of assignments per module, and the volume and quality of tutor feedback on all assignments have been maintained over 30 years. This has been achieved across all courses by formal requirements of their course approval process and by several quality assurance processes. The Open University has exceptionally high NSS scores for assessment and feedback.

5.7 Other process dimensions of quality

5.7.1. Reputation

Seeking the views of research peers is a common method used to judge a department or university's research quality and the same methodology could in principle be used to judge educational quality. The highly influential university ranking system in the US provided by the US News and World Report, 'America's Best Colleges', invests heavily in surveys of Deans and Presidents in establishing college reputations. However, the reputational ranking that derives from these surveys correlates closely with the size of institution's federal research grants (Graham and Thompson, 2001) and can also be predicted by undergraduate selectivity, per student expenditure and number of doctoral awarding departments (Astin, 1985), none of which predict educational gains. Reputational data have a very poor reputation as a valid indicator of educational quality.

5.7.2. Peer ratings (where these focus on process dimensions)

Many quality assurance systems make use of expert peer judgement of the quality of educational provision in a degree programme, at the time of a periodic review of some kind, based on a wide range of evidence and documentation and sometimes including observation of teaching. The relationship between these ratings and the evidence on which they are based is not easy to establish as they are inherently

subjective and global, and based on different combinations of evidence, with different weightings, in different contexts, by different groups of peers. However, there may be potential for the application of professional expertise in such subjective judgements to reach more valid conclusions than could be achieved merely on the basis of individual quantitative measures. This is what Teaching Quality Assessment (TQA) ratings attempted to provide in a quantitative way. In league tables in England the six four-point rating scales involved in TQA have usually been combined into a single score out of 24, and institutional averages out of 24 have been used as indicators of educational quality. Subsequent analysis of average TQA scores for institutions has revealed that they are very largely predictable on the basis of student entry standards (A-level points scores) and research performance (RAE scores), together or separately, without reference to any measures of educational process (Drennan and Beck, 2001; Yorke, 1997, 1998). In other words, TQA scores largely reflect reputational factors. This would not be a terminal problem if research performance and quality of students were valid indicators of educational quality but, as we have seen above, they are not. The inability of reputational factors to provide a valid indicator of educational quality is highlighted above in Section 5.7.1. The inability of peer judgements to be immune from reputational factors undermines their credibility. TQA scores were also subject to other confounding variables, such as institutional size, which have not been taken into account either in moderating overall scores, or in league tables based on TQA scores (Cook *et al.*, 2006).

5.7.3. Student support

The QAA have highlighted, in their reviews of what has been learnt from institutional audits (QAA, 2003), the important role played by adequate student support services of various kinds: study skills development, counselling, English language support, support for students with special needs, and so on. There a number of reasons why it is difficult to estimate the extent to which student services play a role in educational effectiveness or gain. Support services are configured in many different ways, for example subsumed within academic roles or centralised in generic service units. They are described using different terminology: for example, there are few US equivalents of the UK's traditional personal tutor role, and few UK equivalents of the role of 'student advising' in the US. This makes collating evidence across contexts, or comparing like with like, somewhat challenging. Data concerning the positive impact of student support from large US studies are difficult to relate to the nature of UK provision. The impact of such services also relates closely to the nature of student intake. Slender provision at one institution might be perfectly adequate because it only has a zephyr of demand to deal with, while at another institution even extensive and professionally run support services may face a gale of demand and expectations and so may fall short despite extensive institutional commitment. There is clear

evidence of the role of various kinds of student support, for example concerning the impact on student performance of the development of students' study skills (Hattie *et al.*, 1996). However, what support services are appropriate, and how they might best be delivered, can be highly context- and discipline-specific. For this reason no general empirical conclusions will be drawn here.

5.7.4. Quality enhancement processes

Much of the past focus of attention of the Council for National Academic Awards, and today the Quality Assurance Agency, has been on quality processes, such as the operation of the external examiner system and the use of student evaluation of teaching, that are intended to assure quality. The assumption is that if such processes are securely in place, then an adequate level of quality can be more or less guaranteed. There is some evidence to support this kind of assumption. As was discussed in Section 4.1 above, in institutions where student engagement is found to be high and educational gains are high, one finds a higher than average investment of resources in quality enhancement processes such as faculty development and teaching and learning centres (Gansemer-Topf *et al.*, 2004). There is also evidence that some of the prescribed quality enhancement processes have a positive measurable impact, but only under certain circumstances. For example, collecting student feedback on teaching has little or no impact on improving teaching (Weimer and Lenze, 1997) unless it is accompanied by other process such as the teacher consulting with an educational expert, especially when preceded by the expert observing teaching and meeting students (Piccinin *et al.*, 1999).

The extent of institutional adoption of quality enhancement processes through teaching and learning strategies has been documented for English institutions, (HEFCE, 2001; Gibbs *et al.*, 2000), but there is currently no evidence that the extent of adoption of these processes relates to any other measures of process or product.

6. PRODUCT DIMENSIONS OF EDUCATIONAL QUALITY

6.1 Student performance and degree classifications

In the UK the measure most commonly used to indicate the quality of the outcome of higher education is the proportion of students gaining upper second class or first class degrees. The proportion of students who gain 'good degrees' has increased very markedly over time, although unevenly across institutions and subjects (Yorke, 2009). At the same time presage and process indicators of quality (such as funding per student, the quality of student intake, class size, SSRs, amount of close contact with teachers and amount of feedback on assignments) have declined. Yorke (2009) suggests a whole list of reasons why this counter-intuitive phenomenon has occurred. For example, the proportion of assessment marks derived from coursework has increased and coursework usually produces higher marks than examinations (Gibbs and Lucas, 1997). Most of the possible explanations currently lack data through which they could be tested.

The key problem appears to be that there has been little to stop grade inflation. The external examiner system has not proved capable of maintaining the standards that are applied by markers to whatever quality of student work is being assessed. As a consequence degree classifications cannot be trusted as indicators of the quality of outcomes. A whole raft of unjustifiable variations exists in the way student degree classifications are generated. For example, a Maths students is more than three times as likely to gain a first class degree than a History student (Yorke *et al.*, 2002; Bridges *et al.*, 2002) and there are idiosyncratic institutional algorithms for adding marks from different courses (Yorke *et al.*, 2008) that can make as much as a degree classification difference to individual students (Armstrong *et al.*, 1998). The best predictor of the pattern of degree classifications of an institution is that they have produced the same pattern in the past (Johnes, 1992), and institutions' historical patterns are not easily explicable.

It has been argued that there is no longer any meaningful sense in which degree standards are comparable (Brown, 2010). There has been persistent criticism of the meaning and interpretability of degree classifications as indicators of educational outcomes (e.g. House of Commons, 2009) and these arguments have been largely accepted, e.g. by the QAA (2006), and so the arguments will not be rehearsed here. What is clear is that degree classifications do not currently provide a sound basis for indicating the quality of educational outcomes of a UK institution.

6.2 Student retention and persistence

The Open University and the University of Oxford have comparable NSS student ratings for the perceived quality of their educational provision, but are at opposite ends of rankings in terms of student retention, with about 98% of entering undergraduates completing in three years at Oxford, almost double the proportion of new students completing a ten-month course at The Open University. Student retention (in relation to persisting from one year to the next and completion rates within normal time frames) vary very considerably from one institution to another even when educational provision is judged to be similarly excellent or similarly poor. Institutional comparisons are made difficult by the varied nature of student cohorts. Broadly, national retention rates vary in inverse relation to age participation rates (OECD, 2000): the broader the range of student ability is entering higher education, the lower is the overall retention rate. In addition, different institutions take their students from different subsets of the overall ability range.

Students vary not just in terms of their record of past educational success, but in other variables known to affect retention such as whether they live on campus (Chickering, 1974) and whether they are undertaking paid work to support their studies (Paton-Saltzberg and Lindsay, 1993).

In the US it is no longer the case that the majority of students gain the credits they need for a qualification from a single institution. So 'drop-out' is not only the norm but is, for many students, expected and even planned for as they accumulate credits wherever and whenever is convenient. This is not yet the norm in the UK, but 'drop-out' does not have the same meaning or significance for an increasing proportion of students as it does for policy makers (Woodley, 2004). It is not simply that part-time students complete at different rates than do full-time students, but that 'retention' has a different significance for them.

A variable known to influence retention is whether students are socially and academically well integrated (Tinto, 1975). Social and academic integration is affected by living off campus, living at home, and taking time out to earn enough to continue studying. The prevalence of these variables is very varied across institutions, and it is difficult to take all such variables fully into account in judging institutional retention performance.

Student variables also affect retention within institutions and are so influential that in the US commercial companies (such as the Noel-Levitz organisation) offer services to institutions to collect management information and other student data concerning their educational qualifications, preparedness and attitudes, in order to predict which students are most likely to drop out so that scarce additional support can be directed at the students most likely to benefit. A mathematical, data-driven approach of this kind at The Open University has identified very wide differences between entering students in relation to the probability of them completing a single course. This prediction has been used to decide which students to contact and support, with measurable positive

consequences for overall retention (Simpson, 2003; Gibbs *et al.*, 2006). The types of student variables that predict drop-out go well beyond the kind of data that HEFCE have available to calculate institutional benchmarks for retention. So even the extent to which institutions exceed or fall short of their retention benchmarks can only be a crude and incomplete measure of their educational quality.

Not all of the institutional variation in retention is due to student variables. Efforts to improve retention have been evaluated for 30 years in the US, and while overall retention rates have remained largely static, this hides substantial progress in improving retention in some institutions. A good deal is now known about what kinds of institutional efforts are likely to improve retention and persistence in the US (Barefoot, 2004) and, with a much lesser evidence base, in the UK (Yorke, 1999). Making good use of thorough information about students so as to target timely individualised support and intervention is one of the most effective practices. Other effective practices closely resemble those identified as improving student performance and educational outcomes in general (LaNasa *et al.*, 2007), discussed in Section 5.5.3 above. In particular, collaborative and interactive learning and close contact with teachers increases social and academic integration. As pointed out above, such interventions have a greater impact on less able students.

If variations between students, and especially psychological variables such as motivation and commitment, and social variables, such as where students live and how much time they have available to study, could be fully taken into account, then retention performance could be used as an indicator of educational quality. However, with the data currently available this is not yet practicable.

6.3 Employability and graduate destinations

The extent to which graduating students are able to obtain employment reasonably quickly, in graduate jobs, in fields relevant to their degree subject, and with a salary that justifies their investment of time and money in their higher education, is a commonly used dimension of quality. The difficulty with employability data, as with retention data, is their interpretation. Different methods of collecting data, and in particular the timing of the data collection, makes a considerable difference, and the process usually relies on surveys involving student self-reporting. However employability data are collected, interpreting differences between institutions is problematic for a wide variety of reasons (Smith *et al.*, 2000):

— It is affected by degree classification, and student performance is closely linked to students' prior educational qualifications (which also affects employability), which varies greatly between institutions (see Section 4.4 above). In addition, degree classifications vary in their meanings across institutions (see Section 6.1 above).

- It is greatly affected by institutional reputation, which is a very poor indicator of educational quality (see Section 5.7.1 above). There is little evidence that employers have an accurate and up-to-date picture of the educational quality of the institutions they employ graduates from, especially if they derive this picture from currently invalid league tables.
- It is affected by region and locality, due to variations in local employment markets and the proportion of students who live at home and who are less likely to move away either to study or to find employment.
- It changes over time, due to changes in the employment market, and probably differentially between institutions, without any changes in the effectiveness of higher education institutions.
- It is affected by students' social class, and the mix of social class varies between institutions.
- It is affected by students' age, and age profiles vary between institutions.
- It is affected by student affluence, with some students needing to take any employment quickly while others can afford to wait for graduate-level and subject-relevant employment, or to continue their education.
- It is affected by subject mix, with higher levels of graduate unemployment, non-graduate-level employment, and employment in fields unrelated to the degree subject, more often in some subjects than in others.

Interpreting the differences that exist between employability indicators in the UK and mainland Europe is also difficult because the higher education systems are different in crucial ways. For example, about half of UK graduates from non-vocational subjects felt that their first jobs were not appropriate to a degree-level education compared with about a quarter of such graduates from other European countries. For vocational science graduates the proportion was only 17% and 10% respectively, but still higher for the UK than for the rest of Europe (Brennan and Tang, 2008a, 2008b). The overall difference, across all subjects, is probably because European graduates tend to be older, much more likely to have Masters-level qualifications by the time they enter the job market, and more likely to have experienced work-based education relevant to their subject during their more extended education. There is also a different subject mix between the UK and Europe with more non-vocational Arts graduates in the UK (with unemployment rates twice that of some other subjects). This employability difference between the UK and Europe disappears about five years after graduation, at which point 96% of UK graduates are in employment (*ibid.*), which supports the interpretation that there are influential differences between the UK and mainland Europe regarding graduates' age and vocational experience at the time of graduation.

There are US examples of research that validate institutional employability missions by identifying long-term consequences for the way graduates demonstrate, in their professional lives, abilities learnt at college (e.g. Mentkowski and Doherty,

1984). There are only very modest studies of this kind in the UK (e.g. Jenkins *et al.*, 2001), and certainly not enough to make institutional comparisons or even to validate institutional claims about the efficacy of their employability missions.

The Higher Education Statistics Agency is able to take into account some variables (subject of study, qualifications on entry and age on entry) in setting institutional performance benchmarks for employability, but not others. Smith *et al.* (2000) have made a more mathematically sophisticated attempt to take more variables into account, but still leave out crucial variables about which data are not easy to obtain. Interpreting an institution's graduate employment performance in relation to HEFCE benchmarks is fraught with problems.

Finally, the loose fit that characterises the UK's higher education and its jobs market has been interpreted by some commentators not as a problem, but as providing flexibility for graduates to cope with a fluid employment market that is constantly changing in relation to the capabilities that are required. This issue concerns the difference between expertise for efficiency, which is what employers recruiting graduates normally demand, and adaptable expertise, that enables an individual to operate effectively in unpredictable new situations (Schwartz *et al.*, 2005). It takes very different kinds of educational process to develop these two forms of expertise. There is a lack of evidence about the long-term consequences for graduate employment of either narrowly focused vocational education or education that emphasises efficiency in generic 'employability skills', rather than emphasising the higher order intellectual capabilities involved in adaptable expertise. This makes relying on HESA's very short-term employment data a risky thing to do.

7. SUMMARY AND CONCLUSIONS

7.1 The importance of process variables

Much of this report demonstrates what commentators in the US have been arguing for many years. Presage variables such as funding, research performance and the reputation that enables an institution to have highly selective entry, do not explain much of the variation between institutions in relation to educational gains. Measures of educational product such as grades and career earnings reflect these presage variables, because the best students compete to enter the best funded and most prestigious institutions and the quality of students is the best predictor of products. Measures of product such as retention and employability are strongly influenced by a raft of variables that make interpreting an institution's performance extremely difficult.

The most important conclusion of this report is that what best predicts educational gain is measures of educational process: what institutions do with their resources to make the most of whatever students they have. The process variables that best predict gains are not to do with the facilities themselves, or to do with student satisfaction with these facilities, but concern a small range of fairly well-understood pedagogical practices that engender student engagement.

In the UK we have few data about the prevalence of these educational practices because they are not systematically documented through quality assurance systems and nor are they (in the main) the focus of the NSS. The best measure of engagement, the NSSE, is used only to a very limited extent in the UK.

7.2 The importance of multivariate analysis

Much of the UK data about relationships between presage and process variables, or between either presage or process variables and product variables, looks at one pair of variables at a time – for example, the relationship between a measure of research performance (e.g. the RAE) and a measure of teaching quality (e.g. TQA scores). Such relationships are invariably confounded with related variables, for example with the quality of students attracted to the high-status institutions that have high research performance. As a consequence few relationships between two variables can be interpreted with confidence. The few UK studies that have examined a number of variables at a time using some form of multivariate analysis (e.g. Drennan and Beck,

2001; Yorke, 1998) have confirmed that apparently strong relationships between pairs of variables (e.g. between a measure of research and a measure of teaching) are confounded by other variables that could equally be responsible for apparent relationships (e.g. a measure of quality of student intake). In the US there have been far more, larger and more complex, multivariate analyses that take into account a whole raft of variables at the same time and which, as a consequence, are able to tease out those variables that are confounded with others and those that are not. We are therefore largely dependent on US data and analyses for our understanding of the complex relationships between dimensions of quality. Some of the necessary data that would allow a more comprehensive multivariate analysis in the UK have already been collected and collated (for example by HEFCE, HESA, the NSS and by HEPI), but it currently resides in different data-bases. It would be helpful to combine these databases so as to allow multivariate analysis, and to align data collection methods to make this easier to do.

7.3 The importance of educational gain

While some UK data include measures of educational product, there are very few UK studies that have included measures of educational gain. This matters because the best predictor of product is the quality of students entering the institution, and the quality of students varies greatly between institutions, so that if you only have a measure of product, such as degree classifications, rather than of gains, then you cannot easily interpret differences between institutions. When UK studies do attempt to measure gain they involve different measures on entry than on leaving higher education (for example A-level point scores and degree classifications, respectively). Furthermore the most common measure of product, degree classification, varies in its meaning and standard across subjects and across institutions (Yorke, 2009). It is therefore difficult to interpret even these comparative measures of gain. Studies in the US in contrast are far more likely to use psychometric measures of generic educational outcomes (such as a test of critical thinking) with the same measure, and with the same standards being used across different subjects and institutions, and also using the same measure both before and after experiencing three or four years of college. In this way a reliable measure of educational gain, and comparison between institutions in relation to educational gain, is possible. Again we are heavily dependent on US studies for evidence of which quality dimensions predict educational gain, and especially on the vast studies, and reviews of evidence, undertaken by Astin (1977, 1993) and Pascarella and Terenzini (1991, 2005).

7.4 Dimensions of quality in different kinds of institutions

Relying on US data might not matter if institutions and educational processes were essentially the same on either side of the Atlantic. However, it seems likely that

the dimensions that define quality in a valid way are different in different kinds of institutions. For example, even within the US the quality indicators that appear valid for large, national, research universities (in the sense that they predict educational performance tolerably well) do not work as well, or at all, in regional schools and non-selective colleges (Schmitz, 1993). Similarly the normal lack of a relationship between an emphasis on research and an emphasis on teaching does not seem to apply to a small group of well-endowed liberal arts colleges that emphasise close contact between teachers and students (Astin, 1993). Different pedagogical phenomena are likely to be salient in different contexts, with somewhat different patterns of relationships between process and product, dependent on context.

It is not just that different educational processes might have more influence on educational gains in some types of institution than in others. Measures of educational gain themselves might also need to be different between institutions if they are to have meaning. Institutional missions vary, particularly with regard to the relative importance of employability and subject knowledge. It would be surprising if the same measures of educational gain were equally appropriate in all UK institutions. For example, The Open University's mission, emphasising openness, means that it would not seek to increase student retention and performance through increasing selectivity because that would reduce its openness. Its own indicators of quality are distinctive, and are different even from those used by HEFCE in determining its funding. The problem here is that funding mechanisms are driven by indicators of quality that cut across institutions' missions.

US research has done well to identify any consistent patterns at all across varied contexts. However, the limits of what is possible to conclude, on average, have been highlighted by those conducting the research (Pascarella, 2001). The same caution should accompany extrapolation of findings about key indicators of quality from varied US contexts to varied UK contexts.

7.5 Dimensions of quality in different departments

Much of the literature cited above, and most of the debate, has focused on institutional differences in quality. However, it is clear that departments can differ hugely within the same institution. Regarding NSS scores, there are institutions that have the highest-rated department in England in one subject and the lowest rated in another subject, despite sharing the same institutional quality indicators. Educational leadership of departments makes a difference, creating cultures that value teaching, that engage in a constant process of improving teaching, and that create rich and engaging learning environments, to some extent whatever the institutional environment and presage variables (Ramsden, 1998; Gibbs *et al.*, 2008b).

Interestingly the two institutions frequently referred to in this report, and that appear at the top of the NSS ranking, the University of Oxford and The

Open University, have 'institutional pedagogies': common patterns of teaching and assessment across all departments. All their departments score similarly highly on the NSS. This may be in part due to deliberate policies to protect the institutional pedagogy (which is known to work well) from outside pressures, with the effect of limiting departmental variation in teaching quality. Princeton University similarly has an approach to quality assurance through its course approval process that is intended to maintain 'the Princeton model' rather than encourage diverse innovation.

Much of the emphasis in the UK on recognising and rewarding quality in teaching has focused on individual teachers (e.g. through the National Teaching Fellowship Scheme) or institutions (through league tables of one kind or another). There are examples elsewhere of national and institutional schemes to identify educational quality (e.g. in Finland and at the University of Oslo, respectively) that focus on departments, or 'learning environments' at the level of a degree programme (Gibbs, 2008).

7.6 Dimensions of quality in different subjects

The Carnegie Foundation for the Advancement of Teaching established a large-scale initiative on the assumption that the pedagogies of disciplines are different: that, for example, educational quality is achieved through different strategies and practices in the creative arts than in the sciences, and differently in English than in Law (Huber and Morreale, 2002). At a sufficient level of abstraction there are similar underpinning educational principles across all disciplines (Gibbs, 1999), but they are embodied in such varied educational practices, and are salient to such different extents in different disciplines, that disciplines in effect achieve educational quality in different ways. If you then attempt to measure quality across disciplines, for example by using the Course Experience Questionnaire, you find that some disciplines emerge consistently better than others, across different studies and different institutions. Either one has to accept that certain subjects are always taught less well than others, which seems highly unlikely, or that different measures of quality are better aligned with the consequences of some (disciplinary) pedagogic practices than with others. Product measures are particularly susceptible to disciplinary variation. The distribution of degree classifications is markedly different between subjects in the UK (Yorke *et al.*, 2002; Bridges *et al.*, 2002). Employability means different things, and comes about in different ways, following the study of different disciplines. Comparing quality between disciplines is fraught with difficulties.

It seems likely that many students choose subjects first and institutions second, not the other way round. Current institutional quality data and rankings are the wrong place to start if informing potential students is the priority. Furthermore the definition of a 'subject' in the information available to students often does not correspond very closely with the programme in which students may be interested. Brown *et al.* (2009) give an example of a 'subject category' used for reporting

NSS ratings that in one institution is an average drawn from 11 different degree programmes. Students need good data about programmes more than they do about institutions or even about broad 'subjects', and the NSS currently does not provide that, for technical reasons that will be difficult to overcome. Political demands for 'better information for customers' cannot be met with current data gathering and analysis methods partly because they aggregate data in too coarse a way. Once data are aggregated in a fine enough way to be useful, there are then bound to be problems with sample sizes. This problem may be intractable and is one of a number of similar problems that make it difficult to provide information about quality in accessible and usable forms even when it has been collated (Brown, 2007).

The quality of individual courses or modules also varies within degree programmes, and the extent of this variation may be related to degree coherence. This report has focused on institutions and degree programmes rather than on variables that primarily affect individual courses.

7.7 Dimensions of quality that are difficult to quantify

Most of this report has focused on dimensions of quality that are fairly readily operationalisable in a way that enables them to be measured quantitatively, so that statistical relationships can be established with other dimensions that are similarly easy to measure. There are other dimensions of quality that are important, at least in some contexts, but that are difficult or impossible to quantify. For example, throughout literature involving case studies of excellent teaching at department level there are references to aspects of departmental culture: whether teaching is valued and rewarded, whether teachers regularly talk to each other about teaching and its improvement, whether innovation in teaching is systematically supported and funded, whether educational effectiveness is the subject of serious scholarly evaluation, and so on (Hannan and Silver, 2000). Qualities of departmental leadership of teaching make a considerable difference (Ramsden, 1998; Gibbs *et al*., 2008a), and some efforts have been made to measure teachers' perceptions both of departmental leadership of teaching and of the teaching environment that frames the kind of teaching and learning that is likely to take place (e.g. Prosser and Trigwell, 1997; Martin *et al*., 2003).

Sometimes highly effective educational systems are driven almost entirely by values, such as 'liking young people', virtually independently of the pedagogic practices employed or the resources available. In an international study of departments that were identified by their institution as of exceptionally high quality in relation to teaching, students in one of the departments said that their teachers were not especially good *but that it didn't matter* because they felt included in an exciting community of scholars (Gibbs *et al*., 2008a). Studies at Oxford Brookes University concerning why some subjects regularly produced better student performance than others found no differences in any quantitative measure of presage variables. However, a qualitative

follow-up study found that the high performing subjects were characterised by healthy 'communities of practice' involving much discussion of how to solve teaching problems so as to make the entire programme work well for students. In contrast, subjects with consistently low average marks were characterised by a corresponding lack of talking about teaching, and a fragmented focus on individual courses (Havnes, 2008). It may be difficult or impossible to measure such influential variables in ways that allow safe comparison between contexts, although it may be possible to measure their consequences, for example in relation to student engagement.

7.8 Evidence of the products of learning

Among the most telling of all indicators of the quality of educational outcomes must be students' final-year dissertations and project reports. It is a distinctive feature of UK higher education (and in the past a requirement of the CNAA for honours degree classification) that students undertake a very substantial piece of independent study in their final year. Even at US Ivy League institutions undergraduate students would usually need to take a fourth, Honours, year to tackle such a challenging piece of work. It is often a culmination and integration of all they have learnt, especially in applied and creative fields of study. There is an almost total lack of evidence concerning the relative quality of such products across institutions, within subjects. An attempt, for this report, to obtain such evidence from subject centres elicited not a single example, and the few published studies illustrate the embryonic nature of efforts (e.g. Woolf *et al.*, 1999). Dissertations and project reports are often archived and are available for study – although currently not comprehensively across all institutions. Such products would be amenable to systematic peer review within each subject's academic community, in a way that the external examiner system signally fails to do (Warren-Piper, 1994). Such products would also be amenable to review by educational researchers using a generic framework for categorising the quality of learning outcomes such as the SOLO (structure of the observed learning outcome) taxonomy (Biggs and Collis, 1982), which is capable of distinguishing levels of quality across different forms of assessment product within subjects, and even across subjects.

7.9 The potential for improved quality, and the evaluation of improved quality

The lack of a relationship between research performance, funding, SSRs and student selectivity, on the one hand, and student engagement and educational gains on the other, that makes these presage variables such poor indicators of quality, is not inevitable – it is not like an invariant physical law, i.e. it does not apply to all circumstances, for all time. It is in part a consequence of comparatively well-funded,

selective, research-oriented institutions not exploiting their potential advantages to the full. If they were to change their practices in line with what is known about educational effectiveness, in other words emphasise process variables rather than relying on presage variables to do the job on their own, then this relationship could change, provided that their potential advantages enabled them to implement the processes more thoroughly than others.

There is a rapidly growing evidence base in the US about the impact on educational gains of changes in evidence-based educational processes (cf. Carinin *et al.*, 2006). It is becoming ever clearer, as a consequence, what educational processes it is sensible to introduce with a reasonable likelihood of improving educational gains. There is no such evidence base in the UK, or even a methodology for building one. While in England there are some deliberate institutional efforts to improve NSS scores they do not always appear to be based on any valid conceptual framework concerning what processes are likely to produce what outcomes and the NSS does not currently have validity as an indicator of quality in the sense that it predicts outcomes, let alone gains. An example of an alternative approach in the UK is the TESTA project based at the University of Winchester (TESTA, 2010), which employs established methodologies for auditing assessment practices and for measuring student learning responses to those practices, both before and after evidence-based changes in assessment, involving whole degree programmes, across a number of institutions. Progress has been made in the US partly because a large number of institutions have been prepared to adopt the same evidence-based conceptual framework and use the same valid evaluation tools. This allows meaningful evaluation data to be pooled, and so enabling conclusions to be drawn that stand out from the noise of contextual variation. The Higher Education Academy is starting to take a lead in identifying and disseminating suitable evaluation tools and methodologies, and creating an evidence infrastructure within which data from locally conducted studies could be collated.

8. ABOUT THE AUTHOR

Graham Gibbs has spent 35 years in research and development work to improve the quality of teaching, student learning and assessment in higher education.

He has been centrally involved in a series of national teaching development initiatives, including the Teaching More Students Project and HEFCE's Institutional Learning and Teaching Strategy initiative, and in the co-ordination of the Fund for the Development of Teaching and Learning. He is the founder of the Improving Student Learning Symposium and of the International Consortium for Educational Development in Higher Education. He has been awarded an Honorary Doctorate by Sheffield Hallam University for his leadership of the development of teaching in the UK, and by the University of Utrecht for his international leadership of efforts to improve university teaching.

He retired from his position as Professor and Director of the Oxford Learning Institute, at the University of Oxford, in 2007.

9. ACKNOWLEDGEMENTS

In undertaking background research for this report I received support from Bahram Bekhradnia, John Brennan, David Watson and Mantz Yorke, and also from Roger Brown who in addition provided invaluable guidance in revising drafts. They helped me to locate its focus in relation to existing literature and in relation to the nature of current debates about quality and standards in higher education. Their wisdom is greatly appreciated.

I would also like to thank Higher Education Academy staff for their support, and in particular Dr Rachel Segal for her editorial guidance

10. REFERENCES

Abrami, P.C., d'Appolonia, S. and Cohen, P.A. (1990) Validity of student ratings of instruction: what we know and what we do not. *Journal of Educational Psychology.* 82 (2), pp219–231.

'adjwilson'. (2010) If the cap fits [online]. London: Education Guardian. Available from: www.guardian.co.uk/education/2010/may/25/education-letters-gibb-adhd-tuition-fees [12 August 2010].

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

Ashby, E., Plant, K., Ericsson, A., Hill, L. and Asberg, K. (2005) Why study time does not predict grade point average across college students: Implications of deliberate practice for academic performance. *Contemporary Educational Psychology.* 30 (1), pp96–116.

Astin, A. (1977) *Four critical years.* San Francisco: Jossey-Bass.

Astin, A.W. (1985) *Achieving educational excellence: a critical assessment of priorities and practices in higher education.* San Francisco and London: Jossey-Bass.

Astin, A. (1993) *What matters in college.* San Francisco: Jossey-Bass.

Bales, R.F., Strodtbeck, F.L., Mills, T.M. and Roseborough, M. (1951) Channels of communication in small groups. *American Sociological Review.* 16 (4), pp461–468.

Barefoot B.O. (2004) Higher education's revolving door: confronting the problem of drop-out at US colleges and universities. *Open Learning,* 19, (1), pp65–77.

Barrie, S. and Ginns, P. (2007) The linking of national teaching performance indicators to improvements in teaching and learning in classrooms. *Quality in Higher Education.* 13 (3), pp275–286.

Bauer, K.W. and Bennett, J.S. (2003) Alumni Perceptions Used to Assess Undergraduate Research Experience. *Journal of Higher Education.* 74 (2), pp210–230.

Bergren, M., Snover, L. and Breslow, L. (2007) Undergraduate research opportunities at MIT. *Illuminatio.* Spring 2007, pp6–8.

Biggs, J.B. (1993) From theory to practice: a cognitive systems approach. *Higher Education Research and Development.* 12 (1), pp73–85.

Biggs, J. and Collis, K.F. (1982) *Evaluating the quality of learning: the SOLO Taxonomy (Structure of the observed learning outcome).* New York: Academic Press.

Black, P. and Wiliam, D. (1998) Assessment and classroom learning. *Assessment in Education.* 5 (1), pp7–74.

Bloom, B.S., Englehart, M.D., Furst, E.J., Hill, W.H. and Krathwohl, D.R. (1956) *Taxonomy of educational objectives. Handbook 1: Cognitive domain.* New York: David McKay.

Bound, J. and Turner, S. (2005) Cohort crowding: how resources affect collegiate attainment. *Journal of Public Economics.* 91 (5–6), pp877–899.

Bowden, R. (2000) Fantasy Higher Education: university and college league tables. *Quality in Higher Education.* 6 (1), pp41–60.

Brennan, J. and Tang, W. (2008a) *The employment of UK graduates: comparisons with Europe. REFLEX Report to HEFCE No 1.* Bristol: HEFCE.

Brennan, J. and Tang, W. (2008b) *Subject differences in graduate employment across Europe. REFLEX Report to HEFCE No 2.* Bristol: HEFCE.

Brennan, J., Patel, K. and Tang, W. (2009) *Diversity in the student learning experience and time devoted to study: a comparative analysis of the UK and European evidence.* Bristol: HEFCE.

Bridges, P., Cooper, A., Evanson, P., Haines, C., Jenkins, D., Scurry, D., Woolf, H. and Yorke, M. (2002) Coursework marks high, examination marks low: discuss. *Assessment and Evaluation in Higher Education.* 27 (1), pp35–48.

Brown, R. (2006) League Tables – do we have to live with them? *Perspectives: Policy and Practice in Higher Education.* 10 (2), pp33–38.

Brown, R. (2007) *The information fallacy.* Oxford: Higher Education Policy Institute. Available from: www.hepi.ac.uk/484-1291/The-Information-Fallacy.html [June 2010].

Brown, R. (2010) *Comparability of degree standards?* Oxford: Higher Education Policy Institute. Available from: www.hepi.ac.uk/455-1838/Comparability-of-degree-standards.html [June 2010].

Brown, R., Carpenter, C., Collins, R. and Winkwist-Noble, L. (2009) Recent developments in information about programme quality. *Quality in Higher Education.* 13 (2), pp173–186.

Carinin, R., Kuh, G. and Klein, S. (2006) Student engagement and student learning: testing the linkages. *Research in Higher Education.* 47 (1), pp1–32.

Carney, C. McNeish, S. and McColl, J. (2005) The impact of part-time employment on students' health and academic performance: a Scottish perspective. *Journal of Further and Higher Education.* 29 (4), pp307–319.

Cheng, Y.C. and Tam, W.M. (1997) Multi-models of quality in education. *Quality Assurance in Education.* 5 (1), pp22–31.

Chickering, A.W. (1974) *Commuting versus resident students: Overcoming the educational inequities of living off campus.* San Francisco: Jossey-Bass.

Chickering, A.W. and Gamson, Z.F. (1987a) *Seven Principles for Good Practice in Undergraduate Education.* Racine, WI: The Johnson Foundation Inc.

Chickering, A.W. and Gamson, Z.F. (1987b) Seven principles for good practice in undergraduate education. *AAHE Bulletin.* 39 (7), pp3–7.

Chickering, A.W. and Gamson, Z.F. (1991) *Applying the seven principles for good practice in undergraduate education.* San Francisco: Jossey-Bass.

Clarke, M. (2002) Some guidelines for academic quality rankings. *Higher Education in Europe.* 27 (4), pp443–459.

Coffey, M. and Gibbs, G. (2000) The evaluation of the Student Evaluation of Educational Quality questionnaire in UK higher education. *Assessment and Evaluation in Higher Education.* 26 (1), pp89–93.

Cook, R., Butcher, I. and Raeside, R. (2006) Recounting the scores: an analysis of the QAA subject review grades 1995–2001. *Quality in Higher Education.* 12 (2), pp135–144.

Curtis, S. and Shani, N. (2002) The effect of taking paid employment during term-time on students' academic studies. *Journal of Further and Higher Education.* 26 (2), pp129–138.

Curtis, S. and Williams, J. (2002) The reluctant workforce: undergraduates' part-time employment. *Education and Training.* 44 (1), pp5–10.

Dochy, F., Segers, M., Van den Bossche, P. and Gijbels, D. (2003) Effects of problem-based learning: a meta-analysis. *Learning and Instruction.* 13 (5), pp533–568.

Drennan, L.T. and Beck, M. (2001) Teaching Quality Performance Indicators – key influences on the UK universities' scores. *Quality Assurance.* 9 (2), pp92–102.

Dunbar-Goddet, H. and Trigwell, K. (2006) *A study of the relations between student learning and research-active teachers.* Paper presented at the 14th International Improving Student Learning Symposium, Bath, 4–6 September.

Eccles, C. (2002) The use of university rankings in the United Kingdom. *Higher Education in Europe.* 27 (4), pp423–432.

Ehrenberg, R.G. (2006) *What's Happening in Public Higher Education?* Westport, CT: Praeger.

Ewell, P. (2008) No correlation: musings on some myths about quality. *Change.* November-December 2008, 40 (6), pp8–13.

Fearnley, S. (1995) Class size: the erosive effect of recruitment numbers on performance. *Quality in Higher Education.* 1 (1), pp59–65.

Feldman, K. (1984) Class size and college students' evaluations of teachers and courses: a closer look. *Research in Higher Education.* 21 (1), pp45–116.

Finnie, R. and Usher, A. (2005) *Measuring the Quality of Post-secondary Education: Concepts, Current Practices and a Strategic Plan.* Kingston, ON: Canadian Policy Research Networks.

Ford, J., Bosworth, D. and Wilson, R. (1995) Part-time work and full-time higher education. *Studies in Higher Education.* 20 (2), pp187–202.

Gansemer-Topf, A., Saunders, K., Schuh, J. and Shelley, M. (2004) *A study of resource expenditure and allocation at DEEP colleges.* Ames, IA: Educational Leadership and Policy Studies, Iowa State University.

Gardiner, L.F. (1997) *Redesigning higher education: producing dramatic gains in student learning. ASHE-ERIC Higher Education Report 7.* Washington DC: Association for the Study of Higher Education.

Gibbs, G. (1995) National scale faculty development for teaching large classes. In: Wright, A. (ed.) *Teaching Improvement Practices.* New York: Anker.

Gibbs, G. (1999) Are the pedagogies of the disciplines really different? In: Rust, C. (ed.) *Improving Student Learning Through the Disciplines.* Oxford: Oxford Centre for Staff and Learning Development.

Gibbs, G. (2008) *Designing teaching award schemes.* York: Higher Education Academy.

Gibbs, G. (2010) *The assessment of group work: lessons from the literature.* Oxford: Assessment Standards Knowledge Exchange. Available from: www.brookes.ac.uk/aske/documents/Brookes%20groupwork%20Gibbs%20Dec%2009.pdf [June 2010].

Gibbs, G. and Coffey, M. (2004) The impact of training of university teachers on their teaching skills, their approach to teaching and the approach to learning of their students. *Active Learning in Higher Education.* 5 (1), pp87–100.

Gibbs, G. and Dunbar-Goddet, H. (2007) *The effects of programme assessment environments on student learning.* York: Higher Education Academy.

Gibbs, G. and Dunbar-Goddet, H. (2009) Characterising programme-level assessment environments that support learning. *Assessment and Evaluation in Higher Education.* 34 (4), pp481–489.

Gibbs, G. and Jenkins, A. (eds.) (1992) *Teaching Large Classes: maintaining quality with reduced resources.* London: Kogan Page.

Gibbs, G. and Lucas, L. (1997) Coursework assessment, class size and student performance: 1984–94. *Journal of Further and Higher Education.* 21 (2), pp183–192.

Gibbs, G., Habeshaw, T. and Yorke, M. (2000) Institutional learning and teaching strategies in English higher education. *Higher Education.* 40 (3), pp351–372.

Gibbs, G., Knapper, C. and Picinnin, S. (2008a) *Departmental leadership for quality teaching: an international comparative study of effective practice.* London: Leadership Foundation. Available from: www.lfhe.ac.uk/research/projects/gibbsoxford.html [May 2010].

Gibbs, G., Knapper, C. and Picinnin, S. (2008b) Disciplinary and contextually appropriate approaches to leadership of teaching in research-intensive academic departments in higher education. *Higher Education Quarterly.* 62 (4), pp 416–436.

Gibbs, G., Lucas, L. and Simonite, V. (1996) Class size and student performance: 1984–94. *Studies in Higher Education.* 21 (3), pp261–273.

Gibbs, G., Morgan, A. and Taylor, E. (1982) A review of the research of Ference Marton and the Goteborg Group: a phenomenological research perspective on learning. *Higher Education.* 11 (2), pp123–145.

Gibbs, G., Regan, P. and Simpson, O. (2006) Improving student retention through evidence based proactive systems at the Open University (UK). *College Student Retention.* 8 (3), pp359–376.

Glass, G.V. and Smith, M. L. (1978) *Meta-Analysis of Research on the Relationship of Class Size and Achievement.* San Francisco: Far West Laboratory for Educational Research and Development.

Glass, G.V. and Smith, M. L. (1979) Meta-analysis of Research on the Realationship of Class-Size and Achievement. *Evaluation and Policy Analysis.* 1, pp2–16.

Graham, A. and Thompson, N. (2001) Broken ranks: US News' college rankings measure everything but what matters. And most universities don't seem to mind. *Washington Monthly.* 33 (4), pp9–14.

Grunig, S.G. (1997) Research, reputation and resources: the effect of research activity on perceptions of undergraduate education and institutional resource acquisition. *Journal of Higher Education.* 33 (9), pp9–14.

Hannan, A. and Silver, H. (2000) *Innovating in Higher Education: teaching, learning and institutional cultures.* Buckingham: The Society for Research into Higher Education/Open University Press.

Harvey, L. and Green, D. (1993) Defining quality. *Assessment and Evaluation in Higher Education.* 18 (1), pp9–34.

Hathaway, R.S., Nagda, B.A. and Gregerman, S.R. (2002) The relationship of undergraduate research participation to graduate and professional education pursuit: an empirical study. *Journal of College Student Development.* 43 (5), pp614–631.

Hattie, J. and Marsh, H.W. (1996) The relationship between research and teaching: a meta-analysis. *Review of Educational Research.* 66 (4), pp 507–542.

Hattie, J. and Timperley, H. (2007) The power of feedback. *Review of Educational Research.* 77 (1), pp 81–112.

Hattie, J., Bibbs, J. and Purdie, N. (1996) Effects of Learning Skills Interventions on Student Learning: A Meta-analysis. *Review of Educational Research.* 66 (2), pp99–136.

Havnes, A. (2008), *There is a bigger story behind. An analysis of mark average variation across Programmes.* European Asssociation for Research into Learning and Instruction Assessment Conference. University of Northumbria.

HEFCE (2001) *Analysis of Strategies for Learning and Teaching. Report 01/37a.* Bristol: Higher Education Funding Council for England.

HEPI (2006) *The Academic Experience of Students in English Universities (2006 Report).* Oxford: Higher Education Policy Institute.

HEPI (2007) *The Academic Experience of Students in English Universities (2007 Report).* Oxford: Higher Education Policy Institute.

Hochschul-Informations-System (2005) *Eurostudent 2005: Social and Economic conditions of student life in Europe 2005.* Hannover: HIS.

Hoskins, S., Newstead, S.E. and Dennis, I. (1997) Degree Performance as a Function of Age, Gender, Prior Qualifications and Discipline Studied. *Assessment & Evaluation in Higher Education.* 22 (3), pp317–328.

House of Commons Innovations, Universities, Science and Skills Committee (2009) *Students and Universities. Eleventh Report of Session 2008-09. Volume 1.* London: The Stationery Office.

Huber, M.T. and Morreale, S. (2002) *Disciplinary styles in the scholarship of teaching and learning: exploring common ground.* Washington DC: American Association for Higher Education and the Carnegie Foundation for the Advancement of Teaching.

Hunt, A., Lincoln, I. and Walker, A. (2004) Term-time employment and academic attainment: evidence from a large-scale survey of undergraduates at Northumbria University. *Journal of Further and Higher Education.* 28 (1), pp3–18.

Innis, K. and Shaw, M. (1997) How do students spend their time? *Quality Assurance in Education.* 5 (2), pp85–89.

Jenkins, A. (2004) *A guide to the research on teaching-research relations.* York: Higher Education Academy.

Jenkins, A., Jones, L. and Ward, A. (2001) The Long-term Effect of a Degree on Graduate Lives. *Studies in Higher Education.* 26 (2), pp147–161.

Jessop, T. & El-Hakim, Y. (2010) *Evaluating and improving the learning environments created by assessment at programme level: theory and methodology.* European Association for Research into Learning and Instruction Assessment Conference, University of Northumbria, June 2010.

Johnes, G. (1992) Performance indicators in higher education: a survey of recent work. *Oxford Review of Economic Policy.* 8 (2), pp19–33.

Jongbloed, B.W.A. and J.J. Vossensteyn (2001) Keeping up Performances: an international survey of performance based funding in higher education. *Journal of Higher Education Policy and Management.* 23 (2), pp127–145.

Kehm, B.M. and Stensaker, B. (2009) *University rankings, diversity and the landscape of higher education.* Rotterdam: Sense Publishers.

Kuh, G.D. and Pascarella, E.T. (2004) What does institutional selectivity tell us about educational quality? *Change.* September-October 2004, 36 (5), pp52–58.

LaNasa, S., Olson, E. and Alleman, N. (2007) The impact of on-campus student growth on first-year engagement and success. *Research in Higher Education.* 48 (8), pp941–966.

Lindsay, R. and Paton-Saltzberg, R. (1987) Resource changes and academic performance at an English Polytechnic. *Studies in Higher Education.* 12 (2), pp213–27.

Lindsay, R., Breen, R. and Jenkins, A. (2002) Academic research and teaching quality: the views of undergraduate and postgraduate students. *Studies in Higher Education.* 27 (3), pp309–327.

Lucas, L., Jones, O., Gibbs, G., Hughes, S. and Wisker, G. (1996) *The effects of course design features on student learning in large classes at three institutions: a comparative study.* Paper presented at the 4th International Improving Student Learning Symposium, Bath.

Marsh, H.W. (1982) SEEQ: a reliable, valid and useful instrument for collecting students' evaluations of university teaching. British *Journal of Educational Psychology.* 52, pp77–95.

Marsh, H.W. (1987) Students' evaluations of university teaching: research findings, methodological issues, and directions for future research. *International Journal of Educational Research.* 1 (3), (entire issue).

Martin, E., Trigwell, K., Prosser, M. and Ramsden, P. (2003) Variation in the experience of leadership of teaching in higher education. *Studies in Higher Education.* 28 (3), pp247–259.

Marton, F. and Wenestam, C. (1978) Qualitative differences in the understanding and retention of the main point in some texts based on the principle-example structure. In: Gruneberg, M.M., Morris, P.E. and Sykes, R.N. (eds.) *Practical aspects of memory.* London: Academic Press.

Marton, F., Hounsell, D. and Entwistle, N. (1984) *The experience of learning.* Edinburgh: Scottish Academic Press.

Mentkowski, M. and Doherty, A. (1984) *Careering After College: Establishing the Validity of Abilities Learned in College for Later Careering and Professional Performance. Final Report to the National Institute of Education.* Milwaukee, WI: Alverno College.

Nasr, A., Gillett, M. and Booth, E. (1996) Lecturers' teaching qualifications and their teaching performance. *Research and Development in Higher Education.* 18, pp576–581.

NCHEMS (2003) *Do DEEP institutions spend more or differently than their peers?* Boulder, CO: National Centre for Higher Education Management Systems.

NUS (2008) *NUS Student Experience Report.* London: National Union of Students.

OECD (2000) *Education at a Glance 2000.* Paris: Organisation for Economic Cooperation and Development.

Olcott, D. (2010) Par for the course. *Times Higher Education.* 8 April, p32.

Pascarella, E.T. (1980) Student-faculty informal contact and college outcomes. *Review of Educational Research.* 50 (4), pp545–595.

Pascarella, E.T. (2001) Identifying excellence in undergraduate education: are we even close? *Change.* 33 (3), pp19–23.

Pascarella, E.T. and Terenzini, P. (1991) *How college affects students.* San Francisco: Jossey-Bass.

Pascarella, E.T. and Terenzini, P. (2005) *How college affects students: a third decade of research, Volume 2.* San Francisco: Jossey-Bass.

Pascarella, E.T., Cruce, T., Umbach, P., Wolniak, G., Kuh, G., Carini, R., Hayek, J., Gonyea, R. and Zhao, C. (2006) Institutional selectivity and good practices in undergraduate education: How strong is the link? *Journal of Higher Education.* 77 (2), pp251–285.

Pascarella, E.T., Seifert, T.A. and Blaich, C. (2008) *Validation of the NSSE benchmarks and deep approaches to learning against liberal arts outcomes.* Paper presented at the annual meeting of the Association for the Study of Higher Education, Jacksonville, FL. Available from: www.education.uiowa.edu/crue/publications/index.htm [March 2010].

Pascarella, E.T., Seifert, T.A. and Blaich, C. (2010) How effective are the NSSE benchmarks in predicting important educational outcomes? *Change.* January-February 2010, 42(1), pp16–22. Available from: www.changemag.org/index.html [March 2010].

Paton-Saltzberg, R. and Lindsay, R. (1993) *The Effects of Paid Employment on the Academic Performance of Full-time Students in Higher Education.* Oxford: Oxford Polytechnic.

Patrick, J.P. and Stanley, E.C. (1998) Teaching and research quality indicators and the shaping of higher education. *Research in Higher Education.* 39 (1), pp19–41.

Perry, W.G. (1970) *Forms of intellectual and ethical development in the college years: a scheme.* New York: Holt, Rhinehart and Winston.

Piccinin, S., Cristi, C. and McCoy, M. (1999) The impact of individual consultation on student ratings of teaching. *International Journal of Academic Development.* 4 (2), pp75–88.

Prosser, M. and Trigwell, K. (1997) Perceptions of the teaching environment and its relationship to approaches to teaching. *British Journal of Educational Psychology.* 67, pp25–35.

QAA (2003) *Learning from Subject Review 1993–2001.* Gloucester: Quality Assurance Agency for Higher Education.

QAA (2006) *Background Briefing Note: The classification of degree awards.* Gloucester: Quality Assurance Agency for Higher Education.

QAA (2009) *Thematic enquiries into concerns about academic quality and standards in higher education in England.* Gloucester: Quality Assurance Agency for Higher Education.

Ramsden, P. (1979) Student learning and perceptions of the academic environment. *Higher Education.* 8 (4), pp411–427.

Ramsden, P. (1998) *Learning to lead in higher education.* London: Routledge.

Ramsden, P. (1999) A performance indicator of teaching quality in higher education: the Course Experience Questionnaire. *Studies in Higher Education.* 16 (2), pp129–150.

Ramsden, P. and Moses, I. (1992) Associations between research and teaching in Australian higher education. *Higher Education.* 23 (3), pp273–295.

Säljö, R. (1979) Learning about learning. *Higher Education.* 8 (4), pp443–451.

Sastry, T. and Bekhradnia, B. (2007) *The academic experience of students in English universities.* London: Higher Education Policy Institute.

Schmitz, C. (1993) Assessing the validity of Higher Education Indicators. *Journal of Higher Education.* 64 (5), pp503-521.

Schomburg, H. and Teichler, U. (2006) *Higher education and graduate employment in Europe: Results from graduate surveys from twelve countries.* Dordrecht: Springer.

Schwartz, D.L., Bransford, J.D. and Sears, D. (2005) *Efficiency and innovation in transfer.* Stanford, CA: Stanford University. Available from: www.stanford.edu/~danls/Efficiency%20and%20Innovation%204_2004.pdf [April 2010].

Simpson, O. (2003) *Student retention in open and distance learning.* London: Routledge Falmer.

Smith, M.L. and Glass, G.V. (1979) *Relationship of class-size to classroom processes, teacher satisfaction and pupil affect: a meta-analysis.* San Francisco, CA: Far West Laboratory for Educational Research and Development.

Smith, J. and Naylor, R. (2005) Schooling effects on subsequent university performance. *Economics of Education Review.* 24 (5), pp549–562.

Smith, J., McKnight, A. and Naylor, R. (2000) Graduate employability: policy and performance in higher education in the UK. *The Economic Journal.* 110 (464), pp382–411.

Stinebrickner, R. and Stinebrickner, T. (2008) The Causal effect of Studying on Academic Performance. *The B.E. Journal of Economic Analysis and Policy.* 8 (1), Article 14.

Svanum, S. and Bigatti, S.M. (2006) The Influences of Course Effort and Outside Activities on Grades in a College Course. *Journal of College Student Development.* 47 (5), pp564–576.

Svensson, L. (1977) On qualitative differences in learning: III – Study skill and learning. *British Journal of Educational Psychology.* 47, pp223–243.

Terenzini, P. T, and Pascarella, E. T. (1994) Living myths: Undergraduate education in America. *Change.* 26 (1), pp28–32.

TESTA (2010) *Transforming the Experience of Students Through Assessment.* Available from: www.winchester.ac.uk/studyhere/ExcellenceinLearningandTeaching/research/Pages/TESTA.aspx [12 August 2010].

Tinto, V. (1975) Dropout from higher education: a theoretical synthesis of recent research. *Review of Educational Research.* 45 (1), pp89–125.

Thompson, N. (2000) Playing with numbers: how US News mis-measures higher education and what we can do about it. *Washington Monthly.* 32 (9), pp16–23.

Trigwell, K. (2005) Teaching–research relations, cross-disciplinary collegiality and student learning. *Higher Education.* 49 (3), pp235–254.

Trigwell, K. and Ashwin, P. (2004) *Undergraduate students' experience at the University of Oxford.* Oxford: Oxford Learning Institute. Available from: www.learning.ox.ac.uk/oli.php?page=365 [April 2010].

Trigwell, K. & Prosser, M. (2004) The development and use of the Approaches to Teaching Inventory. *Educational Psychology Review,* 16, 4, pp409–424.

Usher, A. and Savino, M. (2006) *A world of difference: a global survey of university league tables.* Toronto, ON: Educational Policy Institute. Available from: www.educationalpolicy.org/pdf/world-of-difference-200602162.pdf [12 August 2010].

Van Rossum, E.J., Deijkers, R. and Hamer, R. (1985) Students' learning conceptions and their interpretation of significant educational concepts. *Higher Education.* 14 (6), pp617–641.

Vos, P. (1991) *Curriculum Control of Learning Processes in Higher Education.* 13th International Forum on Higher Education of the European Association for Institutional Research. Edinburgh.

Warren-Piper, D. (1994) *Are Professors Professional?* The Organisation of University Examinations. London: Jessica Kingsley.

Weimer, M. and Lenze, L.F. (1997) Instructional interventions: a review of literature on efforts to improve instruction. In: Perry, R.P. and Smart, J.C. (eds.) *Effective Teaching in Higher Education: Research and Practice.* New York: Agathon Press.

Wood, K., Limsky, A.S., and Straus, M. A. (1974). Class Size and student evaluations of faculty. *Journal of Higher Education.* 43, pp524–34.

Woodley, A. (2004) Conceptualising student drop-out in part time distance education: pathologising the normal? *Open Learning.* 19 (1), pp47–63.

Woolf, H., Cooper, A., Bourdillon, B., Bridges, P., Collymore, D., Haines, C., Turner, D. and Yorke, M. (1999) Benchmarking academic standards in History: an empirical exercise. *Quality in Higher Education.* 5 (2), pp145–154.

Yorke, M. (1997) A good league table guide? *Quality Assurance in Higher Education.* 5 (2), pp61–72.

Yorke, M. (1998) The Times 'League Table' of universities, 1997: a statistical appraisal. *Quality in Education.* 6 (1), pp58–60.

Yorke, M. (1999) *Leaving early: undergraduate non-completion in higher education.* London: Falmer.

Yorke, M. (2001) Formative assessment and its relevance to retention. *Higher Education Research and Development.* 20 (2), pp115–126.

Yorke, M. (2009) *Trends in honours degree classifications, 1994–95 to 2006–07, for England, Wales and Northern Ireland.* York: Higher Education Academy.

Yorke, M., Bridges, P. and Woolf, H. (2000) Mark distributions and marking practices in UK higher education; some challenging issues. *Active Learning in Higher Education.* 1 (1), pp7–27.

Yorke, M., Barnett,G., Bridges, P., Evanson, P.,Haines, C., Jenkins,D., Knight, P., Scurry, D., Stowell, M. and Woolf, H. (2002) Does grading method influence honours degree classification? *Assessment and Evaluation in Higher Education.* 27 (3), pp269–279.

Yorke, M., Woolf, H., Stowell, M., Allen, R., Haines, C., Redding, M., Scurry, D.,Taylor-Russell, G., Turnbull, W. and Walker, W. (2008) Enigmatic Variations: Honours Degree Assessment Regulations in the UK. *Higher Education Quarterly.* 63 (3), pp157–180.

The Higher Education Academy supports the sector in providing the best possible learning experience for all students. It does this by:

- providing national leadership in developing and disseminating evidence-informed practice about enhancing the student learning experience
- operating as an independent broker, enabling expertise to be shared across institutions and subject areas
- working at multiple levels, with individual academics, subject communities, departments, faculties and institutions
- working across all parts of the UK, recognising the distinctive policy contexts and priorities of the devolved administrations but also provising opportunities to share expertise among them.

The Academy is an independent organisation funded by grants from the four UK higher education funding bodies, subscriptions from higher education institutions, and grant and contract income for specific initiatives.