Durham E-Theses

A case study of a policy decision: anonymous marking with special reference to the university of Durham in 1994

Yeo, Jake

How to cite:

Yeo, Jake (1995) *A case study of a policy decision: anonymous marking with special reference to the university of Durham in 1994*, Durham theses, Durham University. Available at Durham E-Theses Online: http://etheses.dur.ac.uk/5116/

Use policy

A Case Study of a Policy Decision: Anonymous marking with special reference to the University of Durham in 1994

Abstract

Broken down into four separate Chapters, this thesis presents a study of anonymous marking, with special reference to the experience of the University of Durham which introduced a policy of anonymous marking in 1994. It is a study of a single university, and of a very recent change. The thesis adopts a range of different approaches.

Chapter One identifies the principles underlying anonymous marking and examines them in the context of the mission, aims and objectives of a university.

Chapter Two reviews the literature which has sought various explanations for women achieving a lower proportion of first and third class degrees compared to men, suggesting that more attention should be paid to investigating the differential performance of men and women in individual subjects rather than for degrees as a whole; the pattern of results in the one not being the same as the pattern of results in the other. The hypothesis that sex bias in marking lies behind this pattern is shown to be inconclusive. As a result, the positions of a number of high profile individuals and educational organisations, which advocate the widespread introduction of anonymous marking based on the fact that sex bias in marking has been proven to exist, are shown to be misplaced and possibly premature.

In Chapter Three, the practical operation of a newly implemented system of anonymous marking, and the processes involved in the run up to, and after, the policy decision to introduce the system at the University of Durham, are viewed in a case study. The university is analysed as a political system in which the policy process is understood by identifying the different interests involved. The practical implications - in terms of costs, time, administration, anonymity and morale - for all members of the university are considered against the criteria for anonymous marking, and the aims and objectives of the university.

In the concluding Chapter it is recommended that higher education institutions contemplating introducing anonymous marking take a close look at the practicability and desirability of such a system, particularly in the light of other developments in higher education, and the feelings of some of the academic staff and of the student community who did not like to feel anonymous. If a system of anonymous assessment is to be introduced it should be done carefully, with wide consultation and with a clear view as to what the system is being designed to achieve, heeding the lessons and the practical recommendations of this study. Otherwise good objectives will be poorly served by poor policy changes.

CONTENTS

ACKNOWLEDGEMENTS

A CASE STUDY OF A POLICY DECISION: ANONYMOUS MARKING WITH SPECIAL REFERENCE TO THE UNIVERSITY OF DURHAM IN 1994.

Introduction

The aim of this thesis is to study the introduction of anonymous marking, its effects and implications in higher education with special reference to the University of Durham experience. This will be done from three perspectives: a philosophical perspective; a social-psychological perspective; and a practical perspective. The latter will form a case study. In the conclusion the policy process of an educational organisation will also be set within a political framework.

Anonymous marking, a system of marking in which the names, sex and ethnicity of the candidates under examination are unknown to the examiner, is a fairly recent development in the history of changes to affect higher education in Britain. It is only since the 1960s that traditional examination practices, or modes of assessment, have come under serious public scrutiny. As J. Heywood points out, the term 'assessment' did not seem to have much use before 1969, largely because, as he describes, the higher education system was "dominated by written terminal examinations".[1]

The practice of marking anonymously is now widespread, particularly in the former polytechnics. In a number of the older universities, like the University of Durham, for example, the change has been more recent. In

---

[1] Heywood, J., 'Enterprise Learning and its Assessment in Higher Education', A Technical Report published by the Employment Department's Learning Methods Branch, Report No. 20, April 1994, p.28.

Durham's case, anonymous marking has only been introduced across all departments in the last year; this is why it has been chosen for a case study in Chapter Three.

It is important to stress at the outset that this thesis will not be concerned with the wider debate on assessment, which has concentrated most recently on issues arising from initiatives such as enterprise learning, nor will it be concerned with multiple strategies for assessment. Inevitably the discussion will touch on a number of the issues at the core of such debate, for example the issues of quality, reliability and validity in assessment. However, this research will be more specifically targeted at the processes of marking within assessment; an area which has been largely ignored by scholars. It will do this in the context of the recent change in policy in relation to marking in the University of Durham.

The thesis is divided into four sections. Chapter One will consider the general philosophy underlying anonymous marking and other methods of assessment/marking, appreciating that no one method of assessment is perfect, each having its own strengths and weaknesses, advocates and critics. This will be important as it will permit marking to be reviewed in a university-wide context. Here a discussion of the objectives of a university, and where and how the role of assessment fits in with these objectives will be relevant. Definitions of the main concepts being used will be offered.

Chapter Two will consider the specific arguments and allegations for and against anonymous marking in more detail, assessing the validity of the claim that where anonymity does not exist some persons - for example, women - are prejudiced against. Evidence from a number of

psychologists, sociologists and bodies like the National Union of Students will be critically analysed. Particular attention will be paid to the debate in the 1980s between Dr. E. Rudd[2], Dr. C. Bradley[3] and Dr. S. Clarke[4], which focused on the issue of whether sex bias could be said to exist in examining.

Chapter Three will bring together lessons from the earlier sections and apply them to a case study. The practical experience and implications of the introduction of anonymous marking in the University of Durham in 1994 will be studied; Durham being chosen because it provides the most recent and easily accessible material for analysis. By examining the effects of anonymous marking with regard to **all** aspects of work in the university, right through from that of students, academics, their support staff, the university administration staff (particularly in the Examinations Department) and ancillary staff, the extent to which anonymous marking can be said to contribute towards achieving certain university objectives can be evaluated. The situation before and after the introduction of anonymous marking in this institution will be compared recognising that the time since 1994 is not long enough to isolate the effects of anonymity.

---

[2] Rudd, E., 'A comparison between the results achieved by women and men studying for first degrees in British universities', Studies in Higher Education, 9, 1984, pp. 47-57.

-------, 'A reply to Bradley', Studies in Higher Education, 10, 1, 1985, pp. 95-96.

-------, 'Reply to Clarke', Studies in Higher Education, 13, 1988, pp. 333-336.

[3] Bradley, C., 'Sex bias in the evaluation of students', British Journal of Social Psychology, 23, 1984, pp. 147-153.

-------, 'Sex bias in examining reconsidered: A rejoinder to Rudd', Studies in Higher Education, 10, 1, 1985, pp. 91-94.

[4] Clarke, S., 'Another look at the degree results of men and women', Studies in Higher Education, 13, 1988, pp. 315-331.

The research findings will be drawn together in a conclusion which considers anonymous marking in relation to the mission, goals and objectives of a university; anonymous marking, degree results and bias; policy and practice; and the relevance of political concepts and analysis. The thesis ends with a suggestion as to areas for further study. Recommendations for future practice will be offered where appropriate.

CHAPTER ONE: ANONYMOUS MARKING IN A THEORETICAL CONTEXT.

What is assessment ?

Before the general philosophy underlying anonymous marking and other methods of assessment can be understood, it is important to be clear in what sense the term 'assessment' will be used here. The meaning of the term has been used in an umbrella-like way, as T. W. Hartle indicates:

> *What assessment appears to have become in higher education is a catch-all phrase that refers to a wide range of efforts to improve educational quality. This tendency to use one concept to refer to a handful of different (if related) things means that there are few shared meanings and little agreement about the nature, purpose or content of appropriate public policies.*[5]

Various distinctions and categories have been applied to the concept of assessment. For example, D. Rowntree[6] identifies the following modes of assessment: formal and informal; formative and summative; continuous and terminal; course work and examination; process and product; internal and external; and convergent and divergent. The thesis will have cause to consider some of the above. The ways in which assessment is used can also be grouped in a number of ways. N. Harris draws on the distinctions between what he calls 'comparative', 'diagnostic', 'absolute' and 'predictive'

---

[5] Hartle, T. W., 'The growing interest in measuring the educational achievement of college students' in Adelman, C. (ed.), <u>Assessment in Higher Education: Issues and Contexts,</u> US Department of Education: Washington DC, 1986, quoted in Heywood, 'Enterprise Learning', p. 53.

[6] Rowntree, D., <u>Assessing Students: How Shall We Know Them?</u>, Harper & Row, London, 1977.

assessment.[7] The lists are endless, as are the labels describing the particular features of assessment. The important point to recognise is that assessment may be about different things for different people. This is why a common meaning of the term is difficult to identify. The same will apply to the objectives of assessment , which when considered together are neither unconnected nor entirely compatible. Having said all this, such distinctions, though extremely interesting, do not convey the underlying essence of assessment.

There are two things common to all definitions of assessment. First, some notion of *measurement* of achievement or performance and second the idea of *feedback*. Not all works on assessment recognise this. For example the glossary in the pamphlet entitled 'Assessment' produced by the National Association of Teachers in Further and Higher Education, which itself is taken from another reputable source, defines *assessment* as "the placing of an individual upon some given or accepted scale set up to portray his achievement or ability as accurately as possible".[8] This definition is fine as far as it goes, but assessment would be worth nothing if its results weren't communicated to somebody (feedback), even if only to oneself, as with self-assessment. Without feedback it would be impossible to modify the practices and performances of all those involved in the assessment process. Furthermore, the motivation that feedback provides would be lost.

---

[7] For an elaboration on these distinctions see Harris, N. D. C., 'What is assessment?', Assessment in Higher Education, 1, 1975, pp. 5-12.

[8] National Association of Teachers in Further and Higher Education (NATFHE), 'Assessment', a pamphlet produced by the Examinations Panel of the Association of Teachers in Technical Institutions, (no date given), p.42. [This glossary is taken from Nutall, D. L. and Willmott, A. S., British Examinations, Techniques of Analysis, National Foundation for Educational Research in England and Wales, Windsor, 1972.]

One of the most lucid explanations of assessment can be found in a book (published research report) by C. M. L. Miller and M. Parlett.[9] In simple language they describe the characteristics common to any system of assessment. Importantly, both the measurement of educational ability and the role of feedback are featured. The series of stages involved in assessment are as follows:

a   *A specification of a task or tasks is prepared by one or more members of the departmental staff.*

b   *This 'assessment task' is presented to students for completion within certain boundary conditions; usually the task is compulsory.*

c   *The work completed in connection with this task is taken away for 'examination' and appraisal by particular staff members.*

d   *Judgements are made by those evaluating the work and these judgements are communicated to selected others (the student himself, the board of examiners, etc.).*[10]

This is a simple but useful way of understanding assessment in very broad terms.

---

[9] Miller, C. M. L. and Parlett, M., Up to the Mark: A Study of the Examination Game, Society for Research into Higher Education, London, 1974, p. 14.

[10] From this breakdown, Miller and Parlett show that it is possible to distinguish between assessment tasks along five lines: task complexity; weight (how much the assessment task counts for); time allowed; predictability of what the task involves (e.g. whether the questions are known beforehand); and task distribution (how spaced out the tasks are), Up to the Mark, pp. 14-15.

What is a university all about ?

Assessment forms a key part of what a university is all about publicly and privately. To illustrate this one need only look at the mission statements (statements of a university's purpose, usually condensed into an easy-to-read form[11]) of various higher education institutions. The University of Durham, for example, states as one of its goals, "to maintain and monitor excellence in teaching and examining".[12] The objectives or goals of a university are likely to vary according to which part of a university is being analysed. However, the overall mission of most universities will make some reference to achieving the highest possible standards in teaching and research with some sort of public service commitment built in.[13]

A university is very much a system made up of a number of sub-systems, each with its own priorities and objectives. For instance, the academic staff community, the student community and the administration of a university all share in the idea of a university, yet each will have different outlooks as to what they consider is important. This idea of a university as a micro-society rather than as an institution or building is an idea endorsed by Bruce Truscot in his book, Redbrick University.[14] In this section, the objectives of, and the demands on, each of these communities

---

[11] For an example of what a mission statement might include see Allen, M., The Goals of Universities, The Society for Research into Higher Education and Open University Press, Milton Keynes, 1988, pp. 147-154.

[12] University of Durham Mission Statement, 'Investing in Excellence', No date given.

[13] Bruce Truscot identifies the twofold aim of universities as teaching and research taking the somewhat controversial view that research should come first. Red Brick University, Faber and Faber Ltd., London, 1943, p. 105.

[14] Truscot, Red Brick University, pp. 45-46.

will be considered, because this will have a bearing on what is required of assessment. Thus by appreciating what a university is all about from these different perspectives, an understanding will be gained of the sorts of demands placed on any system of assessment.

Before continuing let us be clear about the difference between goals and objectives so that possible ambiguity is avoided. In educational terms, a goal can be seen as being in between objectives on the one hand and a mission on the other; goals being more specific than missions and objectives being more specific than goals. In this way goals often relate to the whole university, whilst objectives will vary enormously between courses, departments and Faculties. Objectives tend to be the responsibility of lower level management and are derived from the goals of an institution. They are more specific and measurable. Goals are less easily measurable and are derived from an institution's overall mission. Interestingly, R. H. Fenske makes the point that goals usually make reference to a clientele, a process and an outcome.[15] More will be said of goals and objectives later, in relation to the concepts of efficiency and effectiveness.

The administration of a university has a number of responsibilities, worth noting for their influence on assessment practice. Primarily, the role of the administrative staff is to facilitate the highest standards in teaching and research of their educational institution. Quality, in recent years, has become the buzz-word for higher education. Quality assessment, quality assurance and audit, and concepts like 'Total Quality Management' are becoming integral features of a university's work, though much debate

---

[15] Fenske, R. H., 'Setting institutional goals and objectives', in Jedamus, P. *et al.*, Improving Academic Management, Jossey-Bass, London, 1981, p. 179.

surrounds their actual worth. The aims and direction of higher education is reflected in the fact that universities are now rated for their teaching and research, the outcome having a direct effect on their level of funding. There is much current argument over the role and value of regulatory bodies like the Higher Education Quality Council and the funding councils which carry out this work. There is increasing pressure from certain quarters to establish one body of a more independent nature to do this work. Much of the academic community currently feels that an excessive amount of time and resources is spent on the quality of documentation, much to the detriment of the quality of teaching, learning (of which assessment is an integral part) and research. Some, like Dr. Mike Fitzgerald, the Vice-Chancellor of Thames Valley University, give credence to the estimation that for each university to meet the demands of quality management, an extra two full-time members of staff (or the equivalent of) might be required.[16] This, therefore, will have a bearing on a university's decision as to which mode of assessment it should adopt.

Democratic accountability and value for money are linked to this notion of quality. They, too, have become increasingly important to the continued survival of higher education institutions, whose funding has become inextricably linked to explicit, rationalised measures which serve as performance indicators for the purpose of outside evaluation. Such considerations are now at the heart of the work of university managers. The definition and pursuit of university goals and objectives reflect this. Among the goals derived from the University of Durham's mission to "achieve and sustain the highest standards of excellence in teaching and research as a collegiate university" is the aim to, "strengthen the

---

[16] Fitzgerald, M., 'Assessment and audit in disrepute', The Times Higher Education Supplement, Oct. 28, 1994.

management of our University [Durham] at all levels to ensure that our goals are achieved efficiently and effectively".[17] Underneath this heading are the sub-headings efficient, effective, excellent and economic, each with a sentence explaining the particular ways in which the University of Durham has sought to attain such values. This is common right across the higher education sector. Thus any system of assessment must, in line with this, be seen to be worthwhile in terms of value for money.

It is important to recognise that the concepts of efficiency, effectiveness and economy, which now lie at the heart of public sector management, are highly ambiguous terms which are often used in misleading ways. As Richard A. Chapman indicates, the terms efficiency and effectiveness often become synonymous with each other, and amount to little more than 'hurrah' words.[18] To avoid falling into the same trap these concepts must be defined. Economy can be understood as using fewer resources, or "input reduction".[19] Using a Treasury definition, effectiveness is "essentially concerned with objectives and can be measured in terms of the extent to which objectives are achieved". Here objectives are taken to mean ends that are more specific and measurable than goals, as outlined above. An example of an objective by which the effectiveness of universities might be measured is that of student admission or completion rates.

---

[17] University of Durham Mission Statement, 'Investing in Excellence', No date given.

[18] Chapman, Richard A., 'Ethics in the public sector', Politeia, vol. 12, 2, 1993, p.32 (pp. 28-42). See also Chapman, Richard A., 'Efficiency and effectiveness in the civil service', Appendix 25, Eleventh Report from the Expenditure Committee Session 1976-7, vol. III, Appendices, HC 535-III, HMSO, London, 1977, pp. 957-959.

[19] Greenwood, John and Wilson, David, Public Administration in Britain Today, 2nd edition, Unwin Hyman, London, 1989, p. 12.

Efficiency, according to John Greenwood and David Wilson, is "concerned with the relationship between the inputs (resources) into a particular activity and the outputs (goods or services) produced by it".[20] In this way efficiency relates to the extent to which goals set by higher level management are achieved. Such ends tend to be more general, like for example wanting to improve the standard of education across Britain. Pursuing such a goal in the name of efficiency will involve dealing with many unquantifiable elements, as well as having to incorporate the unintended consequences of objectives. These might include factors like a drop in the morale of teachers due to the extra burden of apparently unnecessary work being placed upon them to achieve the set goal. Clearly, therefore, effective management and efficient management are not always coexistent as many assume, and what can be said to be efficient might be misleading with respect to the public sector.

It is actually extremely difficult to determine exactly what efficiency (and good performance) is, because of the nature of the environments within which universities work. Identifying and defining objectives and goals in a non-profit driven sector, where much of the work is qualitative rather than quantitative, is not straightforward and can lead to a situation where universities are driven to pursue that which can be seen and can be quickly measured, at the cost of other equally important long-term things, like for example the quality and equity of the education being provided. Some contend that quality and equity in education can be sufficiently measured by indicators and in this way guaranteed. Either way, one must be careful to recognise that targets can always be set to be met. Thus claims to be effective must be viewed cautiously with a critical eye. This

---

[20] Greenwood and Wilson, Public Administration, p. 12.

must be borne in mind when analysing the assessment procedures of Durham University later in the thesis.

There are various other demands on the management of a university which have some relevance to assessment. The need for universities to be publicly accountable as part of the quality assurance package necessitates that certain information is made available. For instance, student success and failure rates are readily accessible at most higher education institutions, serving the dual purpose of feedback for the university and an indication of performance for outsiders. This sort of information is the end product of assessment, because without formal assessment there would be no measure of success and failure.

There is increasing concern that standards in universities are maintained. As a result an institution's assessment procedures are constantly under scrutiny. It has become important for courses to be evaluated, prestige of departments and universities to be promoted and for equal opportunities to be guaranteed. Being able to show the value added nature of the work universities do has become a requirement, even though the measures often used to arrive at this information are highly questionable. For example, the assumption commonly made is that 'A' level scores correlate with degree honours results. In this way a student with high 'A' level grades is expected to achieve high honours results. Whether or not he or she does achieve high honours results is often interpreted as an indication of the level of value added by the university. However, such a measurement takes no account of the differences between the type of learning involved in the study of 'A' levels compared to degrees, nor the differences in the nature of the courses offered by the university (some courses require prior knowledge of the subject, like for instance Mathematics and French, other

courses accept applicants with no prior knowledge, three examples being Politics, Sociology and Anthropology). With the latter courses it is argued that a student with high grades in GCE 'A' levels in quite different subjects should not be expected to get a higher class of degree than a student with different grades in other 'A' level subjects. Hence, assessing the amount of value that is added by a university is not as straightforward as one might think. These are all pressures which need to be accomodated in a system of assessment, from a managerial perspective.

A university also exists for the purposes of individuals and society as a whole. Much of the above has focused on what a university is all about with respect to its administration, and how examination results may be interpreted in terms of quality assessment, although plainly the responsibilities involved in providing quality assurance span all levels of university work. Nevertheless, a student perspective as to what he or she wants to get out of his or her time at university will differ somewhat. The majority of students would probably say that their main reason for coming to university is to get a good degree and an acceptable job. Other reasons might include wanting to increase their own learning and development. Immediately one can see the importance of assessment in the achievement of both of these objectives. Getting a good degree and job involves some formal indication of an individual's competence. Degree classification is the last stage of an assessment process that serves this purpose. Similarly, one cannot learn and develop without feedback, and this involves some form of assessment. This might only be an informal form of assessment like, for instance, a personal tutor's comments on an individual's progress. Again, therefore, assessment has an integral part to play in what a university is all about, this time from a student perspective.

Universities also serve important purposes for employers. The role of assessment acts as a selection device for sifting prospective candidates. Assessment provides an indication of a student's ability to handle pressure and stress, as well as testing other skills like the degree to which an individual can reason analytically and work alongside others. These are transferable skills relevant to the work place.

Tutors and lecturers have a need for the assessment of their students. Assessment forms an important part of what a university is all about for them; it helps academics plan, review and update their own courses and the way they teach them. The students' results is one factor, among others, that a member of staff can be judged by. Thus assessment has a role to play in both a student's and an academic's personal learning and development.

Universities have an important role with respect to the community and society at large. Not only do they promote enterprise in the community (the work of the Business Schools of many universities bears this out), but they carry out research of social value which is increasingly being privately funded. Universities validate courses at other educational bodies, making education more accessible in the locality. The validation of such courses includes overseeing the assessment practices of these courses. In terms of society at large, higher education plays an important role in encouraging citizenship and imparting social values. Radicals might argue that in this respect universities are little more than tools for social engineering and the preservation of middle class values and, accordingly, the preservation of the social and economic structure of society. Examinations and formal assessment contribute to this stratification within society. Further than this, the argument would

probably extend to the view that the system of higher education and grants or loans for students, as it stands, ensures that only those who can afford access will have it. Whatever the view, the importance of assessment is self evident.

This section has, so far, done three things. It has emphasised the importance of assessment. It has helped to place assessment in a university-wide context, illustrating where and how the role of assessment fits in with the many objectives and the overriding mission of a university. It has also provided an indication of the many demands likely to be put upon any particular system of assessment. How far any one system of assessment can satisfactorily meet all of the requirements placed upon it is debatable, and is something on which the discussion will focus when analysing the practical experience of the University of Durham in Chapter Three.

Assessment criteria

It is now necessary to consider some of the concepts associated with assessment, to ensure that their meanings are understood and how, if it all, they relate to anonymous marking. When discussing marking as an aspect of assessment it is essential to be clear about the criteria upon which different modes of assessment/marking are based. Here the notions of fairness, objectivity, reliability and validity need to be explored, as do some of the wider objectives of universities.

Fairness

Fairness is an objective of any system of assessment and university. With regard to marking, fairness is often equated with objectivity, such that the more objective a mode of marking is, the fairer it is deemed to be. This idea is based on objectivity minimising the possible influence of bias and marker inconsistency, for which there is more scope where marking is subjective. Here 'bias' is taken to mean preferential treatment. In this way an examiner would be biased if he or she showed an intentional preference towards a particular student or group of students or towards a particular approach, which would be the case with ideological bias. However, this is not the only conception of fairness with regard to assessment.

One could argue that total objectivity in assessment is too inflexible and therefore unfair. If one considers assessment by multiple choice which is as objective a form of assessment as one can hope to achieve, in the sense that multiple choice tests clearly enable answers to be marked as correct or incorrect, there are many criticisms which suggest it is a highly unfair process. The worth of a multiple choice exam depends solely on the nature of the questions and the subject being tested. Generally speaking, multiple choice questions are only effective where factual knowledge is involved. Where a subject under examination places less emphasis on factual recall, the task of setting multiple choice questions becomes extremely complex. This point is made in a booklet on the setting and evaluation of multiple choice questions which explains that "to spend too much time devising ingenious Multiple Choice Questions to test

understanding is as wasteful as using vivas only to test factual knowledge".[21]

It is highly contentious whether fairness should amount to the complete elimination of value judgements from the examination process. The debate about what is fair is, as C. M. L. Miller and M. Parlett put it, "part of a wider, more philosophical question: one can argue that someone with a close-up knowledge of an individual is better able to make a judgement concerning his endeavours; or alternatively one can question this, on the grounds that personal feelings are likely to cloud a dispassionate judgement."[22] Is it fairer to account for, or ignore, factors like a student's work throughout the year? One can imagine a scenario where a good student fails his or her degree because of one awful examination paper. In this case should any consideration be given to the student's performance beyond the particular examination? Equally, one can question whether it is fairer that an examiner should or should not know beforehand that the script he or she is marking has been written by a dyslexic student or a foreign language student? Here the relevance of anonymous marking is evident: is it fairer that a student be known to his or her marker or that he or she remain anonymous? Such issues need to be addressed in order that a fair and equitable system of assessment can be achieved. It should be pointed out that there are procedures within a system of anonymous marking which allow the special circumstances of students to be accounted for. With most schemes anonymity is lifted when the Board of Examiners meets to agree the final marks of students. However, one

---

[21] Lennox, B., Hints on the Setting and Evaluation of Multiple Choice Questions of the One-from-Five Type, Booklet No. 3, Association for the Study of Medical Education, Dundee, 1974, quoted in Cockburn, B. and Ross, A., Inside Assessment, Teaching in Higher Education Series No. 7, School of Education, University of Lancaster, p. 25.

[22] Miller and Parlett, Up to the Mark, p. 44.

should also recognise that this is after and not before marking has taken place, thus there is the influence (if any) of the examiner's original mark.

Clearly, therefore, it is important to decide what the concept of fairness means in particular cases because in general theoretical terms fairness may be conceived in very different ways. This section has not provided the definitive answer as to what is fair, it has merely raised some important questions. It is up to university departments and Boards of Examiners to determine what is fair by considering the sorts of issues raised above, and by making sure that the method of assessment adopted suits the nature of their particular subject or discipline. Having said this, most scholars would agree that two ideals of a fair system of assessment are reliability and validity. These concepts shall now be examined.

Reliability

There is debate as to whether reliability in itself is a necessary condition of assessment or whether it is simply the consequence of other more important ideals. Many scholars in the field of educational research tend to use the notion of consistency in place of reliability. For example, J. M. Thyne in <u>Principles of Examining</u> poses the question, "why use even the conventional term, 'reliability', if what it means is indicated better by 'consistency'?", for he is quick to point out, through drawing on Wesman's[23] imaginative example of an elastic ruler, that the "essence of reliability is consistency". Thyne argues that with an elastic ruler we cannot assume that two measures, one measuring 16 units and the other 8

---

[23] Wesman, A. G., 'Reliability and Confidence', <u>Test Service Bulletin</u>, No. 44, The Psychological Corporation, New York, 1952, in Gronlund, N. E., <u>Readings in Measurement and Evaluation</u>, The Macmilllan Company, New York, 1968.

units, are comparable, because such a ruler (being elastic) can change the meaning of its units. Thus the same thing is just as likely to measure 16 units as it is 8 units, and so on.[24] In this sense the measures produced are not reliable because they are not consistent. Thyne therefore cites marking-consistency (reliability following naturally from this), as one of four necessary conditions that must hold if examination results are to be of maximum validity. Validity will be examined in more detail in the next section.[25]

For the purposes of this discussion consistency will be used in place of reliability. In terms of assessment and marking this means that for a given group of students the outcome of assessments should be consistent from one occasion to another. Thus if consistency was attained the same results would be achieved if, first, the assessment process was repeated; second, the same students sat a similar examination on the same work; and third the examination was re-marked either by the same or new examiners. In practice it is very difficult to put these tests for assessment consistency to the test themselves, in order to establish their validity. Nevertheless, most would agree that the practice of double-marking and external marking, on top of assessment tasks of other sorts (e.g. assessed essays and dissertations) and questions that are carefully set, provide satisfactory checks for consistency.

---

[24] Thyne, J. M., Principles of Examining, University of London Press, London, 1974, pp. 8-9.

[25] For further discussion concerning the value of using the term 'consistency' over 'reliability' see Cox, R., 'Reliability and Validity of Examinations', The World Year Book of Education, Evans Bros, London, 1969; and Wiseman, S., Examinations and English Education, Manchester University Press, Manchester, 1961.

Validity

Assessment must be valid. In other words it must do what it is designed to do. Two sorts of validity are mentioned in NATFHE's pamphlet on assessment: *content validity* which is "the extent to which an examination comprehensively samples all the content and objectives of the course"; and *predictive validity* which is "the degree to which an examination or test predicts success in some future field of education or employment".[26] This discussion is concerned with content validity.

According to J. M. Thyne there are four necessary conditions of a valid examination. The first, marking-consistency, has already been dealt with. The remaining three are 'mark-relevance', 'question relevance' and 'balance'. Mark relevance specifies that all performance that is required and relevant must be marked. That which is not relevant must not be marked, otherwise the validity of the assessment will be lowered. This would be the case if an examiner awarded marks for spelling when spelling was not included in the examination criteria. In order to ensure that only that which is required and relevant be marked, examiners need to agree on a guide to relevant performances and accordingly the criteria that will be used for the purposes of marking.

Question relevance works on the basis that irrelevant questions (with respect to the syllabus and objectives of the course) invite irrelevant answers, therefore, they must be avoided if an examination is to be valid. Furthermore, it is important to make sure that the *relevant* questions are

---

[26] NATFHE, 'Assessment', p. 44. [Glossary taken from Nutall and Willmott, British examinations, Techniques of Analysis]

asked. It is virtually impossible to design questions that will cover all the relevant performances required by a course, thus it should be the aim of a valid assessment to ensure that a representation of the relevant performances are attained. Accorrding to Thyne, the best way of doing this is to draw up a sample of the <u>kinds</u> of relevant performances desired, and make sure that the examination questions include each kind.

Balance, as a condition of examination validity, refers to the weighting of marks that are given for different parts of an examination which go to make up the final mark. The correct balance must be decided upon in accordance with the purposes of the examination.

There are practical problems in testing these conditions; nevertheless, conceptually, most definitions of validity incorporate these conditions in one form or another. Thyne argues that these conditions are both necessary and sufficient for maximum validity.[27] The concepts that have been discussed in this section need to be borne in mind in later Chapters when the practice of anonymous marking comes under evaluation.

Marking and assessment

Having understood the concept of assessment in general terms the discussion will now focus on marking within assessment. It is important to be clear about the difference, and the interplay, between modes of marking and modes of assessment with regard to higher education. When discussing different forms of assessment one usually thinks of traditional terminal examinations, continuous assessment and other methods of

---

[27] See Chapter 2 of Thyne's <u>Principles of Examining</u> for a fuller exposition of these four conditions of examination validity, pp. 8-28.

formative assessment, which might include student profiling or the grading of oral performance. When considering marking within assessment the practice of double marking, double marking blind, anonymous marking and the role of external examiners is important. These two lists are certainly not mutually exclusive, yet it is immediately obvious that certain forms of marking are not compatible with certain modes of assessment. For example, it is impossible to grade a student's oral performance - a practice common to assessment within the legal profession and in some undergraduate degree courses such as music - on an anonymous basis.

What is anonymous marking ?

Before a discussion of each form of assessment is possible it is important to be absolutely clear what we mean by the different systems of marking. *Marking*, as defined by K. Lovell, is:

> *The process whereby a differentiation is made between the various levels of performance among the various scripts before the examiner. It is done by awarding qualitative or numerical grades or by placing the scripts in rank order. After the marking has been carried out and not before, value judgements can be made regarding whether the scripts are 'good' or 'bad' or whether they should 'pass or 'fail'.*[28]

The only qualification that needs to be made to this definition with respect to marking in higher education is that any value judgements (e.g. 'good', 'bad', 'pass', 'fail', 'upper second', 'first', etc.) are made against explicit

---

[28] Lovell, K., 'Examinations and Marking', in Layton, D. (ed.), <u>University Teaching in Transition,</u> Oliver & Boyd, Edinburgh, 1968, p. 142.

mark scales or boundaries which would have to be set **before** marking took place.

Among the different systems of marking there is what is commonly referred to as 'anonymous marking'. The National Union of Students defines anonymous marking as "a system of assessment where the identity of the candidate is not known by the marker". They go on to stress that "while it [anonymous marking] is largely restricted to written exams, other forms of assessment can be carried out anonymously".[29] The practice of marking anonymously can be conducted in a number of different ways. However, what is common to all systems is that a student is known to an examiner only by a personal number or pseudonym, not by his or her actual name. The candidates names are disclosed only after their scripts have been marked. Anonymous marking is not exclusive to higher education though, for the purposes of this discussion, this is the context in which it will be examined.[30]

Double marking

Double marking can be operated on an anonymous basis or with the names of students on their examination scripts. Virtually all universities mark in this way or are encouraged to do so by regulatory bodies like the funding councils.[31] Double marking involves two markers. Both markers

---

[29] National Union of Students Scotland Campaign for Anonymous Marking, 'A guide to anonymous marking', after 1992.

[30] For an indication of how widespead the practice of anonymous marking is in higher education see the 1989 City University Survey in Appendix I and an extract from a report produced by the University of Sunderland Students' Union in Appendix III.

[31] Professor Graeme Davies, Chief Executive of the Higher Education Funding Council for England, makes this point in correspondence with myself, 26 September, 1994.

mark the same scripts and then compare the grades awarded. In a system sometimes referred to as 'double marking blind', the first examiner's mark is not made known to the second examiner; their marks are completely independent. Where double marking is not blind the second examiner knows the grade awarded by the first examiner. Whether blind or not, this practice is designed to ensure consistency in marking and to safeguard against any favouritism or prejudice that might manifest itself in the final grade awarded.

There are a number of important questions surrounding the workings of double marking. First, it is not clear that double marking in any form, provides the perfect consistency much sought after. Many university departments do not have more than one member of academic staff with the same specialist knowledge in a particular field. In this situation, the second examiner marking a script will have inferior knowledge of the technical or factual content of a course he or she is unfamiliar with. For instance, this would be the case if a specialist in Russian Politics was required to double mark scripts in Middle Eastern Politics. Here the second marker will be marking with a different perspective compared to the first examiner, whose knowledge of the subject area is far more extensive. It has also been suggested that double marking might narrow the range of marks awarded to candidates because in situations where there is disagreement between the markers, compromises are usually struck. One can therefore question how far such a practice is a test of consistency, even if both markers arrive at the same marks independently.

Should second markers in this situation have access to the first examiner's marks and comments, as a marking guide, or does this defeat the whole purpose of double marking? Clearly most examiners will share common

ground over what is a good or bad essay, or indeed what is upper second class or lower second class, but is this a sufficient guarantee of marking consistency when there exists such an imbalance of expertise between the two examiners? In some respects this particular shortfall is redressed by students scripts also being referred to an external examiner who generally shares knowledge of the subject area. Furthermore, attention may be directed to a script with marks significantly out of line with the run of other scripts.

External examiners

An external examiner, as the term suggests, is an examiner who acts from outside a particular institution to moderate the assessments of markers internal to it. External examiners are usually academics of senior standing based at another university. The external examiner system is a way for universities to maintain comparable academic standards and to ensure fairness in the award of course results. In this way, the system acts as a control on quality whilst also providing reassurance to students, staff and institutions. Whether or not the external examiner system is succeeding in this respect is a matter for current debate, as publications like the Reynolds Report[32] indicate.

As with the practice of double marking there are no specific and generally applied guidelines issued about how external examiners should go about their work. Most institutions produce their own guidelines or codes of practice.[33] The Quality Assessment Division of the Higher Education

---

[32] CVCP, 'Academic standards in universities', July 1986. The report was prepared by the group on academic standards under the chairmanship of Professor P. A. Reynolds, and included "A code of practice on the external examiner system for first degree and master's courses".

Funding Council for England encourages universities to adopt such practices, but they offer no guidance on whether external examiners should mark "blind" or not, and how many and which scripts they should check. This is left very much up to the individual external examiner's discretion and the requirements of the university.

An example of the widely different approaches that can be adopted is illustrated by the practice in certain universities. Until about ten or fifteen years ago the Theology Departments at Edinburgh and St. Andrews sent all their examination scripts to the external examiner before they had even been marked internally.[34] Generally speaking, however, an external examiner's work is the last stage in the assessment process, for the specific reason that then he or she can help resolve problematic cases and arbitrate when internal marks are not agreed. External examiners will usually concentrate on scripts that are first class, failures and borderline, as well as a sample of other scripts. Some examiners prefer to mark blind, others like to know the comments of the internal examiners, especially when examining in areas where they are not specialists. Whether or not a university has in place a system of anonymous marking will have some bearing on this. With cases that are particularly problematic an external examiner may arrange to hold a *viva voce* examination to help make a judgement on the final grade a student should receive.

---

[33] For example, the University of Essex makes it quite clear that "where examination scipts are double-marked, the first marker must not write the marks on the script so that the second marker can mark the work independently". The University's position on external examiners is less informative requiring only that external examiners work in accordance with their equal opportunities policy. 'Code of Practice on Equal Opportunities for Students', University of Essex, 1993, p. 6. The Council for National Academic Awards and the Committee of Vice-Chancellors and Principals have each produced guides for external examiners: CNAA, 'Notes for guidance for external examiners', May 1980, CNAA Handbook, 1984; CVCP, 'The external examiner system for first degree and taught master's courses', April 1984.

[34] Information from Reverend Professor John Heywood Thomas, University of Nottingham.

In addition to the basic functions of guarantor of comparability and fairness, the external examiner has an important advisory function which, particularly since the 1980s, has started to gain official recognition.[35] This involves offering comments to internal examiners on the content and structure of the course being assessed, on students work and any suggestions about marking and assessment. There is an argument that the advisory function of the external examiner reduces his or her capacity to be objective because he or she will have a close personal knowledge of the academic members of staff, and the courses, he or she is advising. Whether or not the functions of an external examiner are inconsistent in this respect is debatable; however, in the aftermath of the Reynolds report, which was significant for its recommendation that all external examiners should make regular written reports on any aspect of the teaching and courses they wished (a practice then carried out at some universities but not all), there does seem to be a general level of agreement in higher education circles that both the basic and the advisory functions of the external examiner system are valued, and do not present a threat to the ideals of objectivity and fairness.

There is much current debate about how the external examiner system needs reforming. A recent report by the Higher Education Quality Council (HEQC) entitled, Learning From Audit,[36] criticised some aspects of the external examiner system, and has stimulated public debate on the value of external examiners and where the system can be improved. Based on the findings of 69 quality audits of higher education institutions, the report

---

[35] The advisory function of external examiners is recognised in the CNAA's 'Notes for guidance for external examiners', May 1980, CNAA Handbook, 1984 Section E; and in the CVCP's code of practice on external examiners, 'The external examiner system for first degree and taught master's courses', April 1984.

[36] As cited in the T.H.E.S., 6 January, 1995.

was concerned with the lack of agreed standards that existed across the academic community over the selection of external examiners, what their role actually is, what sort of impact they have on programmes of study, and the usefulness of their reports in assessing these programmes. The key problem identified by the report seemed to be the diversity or lack of consistency that existed across institutions and across external examiners in these respects. There was, therefore, felt to be a need to clarify the functions of external examiners in order to strengthen the system.[37] Incidentally, a further review of the external examiner system has been commissioned by the HEQC and is currently in progress.

Suggestions as to how the system can be made more effective have become the subject of public debate. In the T.H.E.S., Gordon Kirk, Principal of Moray House Institute of Education at Heriot-Watt University, suggests a whole host of measures which he feels would improve the effectiveness of the system like, for example, induction training for external examiners and a national code of practice to set standards for sampling students' work and the format of reporting, as well as the establishment of a single independent national agency to manage the appointment and affiliation of external examiners to particular institutions.[38] In a reply to Kirk's suggestions, Professor Richard A. Chapman of the University of Durham, reminds Kirk of the cost implications of these proposals, money which in the current climate could perhaps be better spent, and of the importance of issues such as the low

[37] In this respect the Report of the HEQC echoed the earlier findings of the 1985 Report of the Committee of Enquiry into the Academic Validation of Degree Courses in Public Sector Higher Education, 'Academic Validation in Public Sector Higher Education', HMSO, Cmnd. 9501, London, April 1985, see chapter on 'External Examiners', pp. 25-32.

[38] Kirk, Gordon, 'Practice with a clearer purpose', Times Higher Education Supplement, 6 January 1995.

pay of examiners against the work already demanded of them (the one not matching the other).[39] Also, the professional integrity of external examiners, who are almost always academics of senior standing, is in itself an assurance across the academic community that high and comparable standards are being maintained.

A consensus of opinion appears to exist over the view that the external examiner system should continue but that it needs strengthening. Evidence of this is clear in the comments of Professor Harold Silver who, reporting on the findings of a survey conducted by the Open University's Validation Service (OUVS) entitled, Using External Examiners,[40] stated that "at no point in this project was any reservation expressed about the importance, now and in the future, of the system. There was widespread approval of the contribution made by external examiners, and an adamant concern that the system should continue". Silver was one of the co-authors of the survey of the 50 institutions which the OUVS accredits (external examiners at these institutions are employed on the same basis as higher education institutions).[41] The survey reportedly found that the external examiner system was valued for its "emphasis on justice, standards and comparability." Clearly, therefore, any reform must seek to achieve a balance between strengthening the external examiner system as it now stands which, among other things, must include addressing the proper remuneration of examiners for the level of work done (a concern voiced in the Silver survey which was felt to have a direct impact on the

---

[39] Chapman, Richard A., 'Hostage of exam fortune' (letter of reply to Kirk's article), Times Higher Education Supplement, 20 January 1995.

[40] As cited in T.H.E.S., 4 November 1994.

[41] Professor Harold Silver is also the author of the report looking into the reform of the external examiner system commisioned by the HEQC which is currently in progress.

effectiveness of the system), as well as a clarification of the consultative function of examiners. At the same time, careful atttention must be paid to the functions of examiners so that they do not become overburdened by the consultative demands placed upon them to the extent that they become in effect 'one-man validators'.[42]

Conclusion

This Chapter has provided an essential foundation for discussion in later Chapters. It has demonstrated the importance of understanding assessment methods in the context of the mission, aims and objectives of a university. The Chapter has also defined the sorts of concepts associated with assessment/marking and identified some of the ambiguities inherent in these concepts; fairness, reliability, objectivity, validity, effectiveness, efficiency, and quality and equal opportunities assurance, among others, are concepts which need to be carefully considered because of their relevance to assessment and, more particularly, to anonymous marking. In theory, at least, anonymous marking is deemed to make assessment practice fairer because it ensures that only what is on the page is assessed (the scope for bias in marking is thought to be eliminated by anonymity). Accordingly, anonymous marking could be considered to increase the reliability of assessment in the sense that marker consistency should be improved. One might also contend that the validity of assessment is increased by anonymity because it is more likely that only that which is required and relevant is marked. However, with each of these points there is evidence that the theory of anonymous marking does not always live up to the reality. In the following Chapters a

---

[42] A term used in the Report of the Committee of Enquiry into the Academic Validation of Degree Courses in Public Sector Higher Education, para. 6.11, p. 28.

more detailed, practical examination of how far a system of anonymous marking advances a university's mission to attain the highest standards in teaching and research, and the objectives which are derived from this mission will be conducted. There will also be discussion of how far anonymous marking secures the aforementioned concepts (fairness, etc), in its own right, and over and above the practices which already exist, like double and external marking.

The next Chapter focuses on an important debate that has arisen in higher education in recent years. Assessment and marking procedures have been propounded as an explanation for the differential performance of men and women at degree level, resulting in some individuals and groups advocating the widespread introduction of anonymous marking. The evidence for this claim will be examined, as will the effect of anonymous marking on degree results.

CHAPTER TWO: ANONYMOUS MARKING IN A SOCIAL-PSYCHOLOGICAL CONTEXT.

Anonymous marking and bias

The issue of anonymous marking is often linked with the issue of bias. Student union literature, whether it is from the National Union of Students or from unions of particular universities, always refers to the "reduction of gender bias" or the "counteraction of racial bias" as two benefits of adopting a system of anonymous marking. However, this literature never establishes what is meant by its use of the term 'bias', nor does it satisfactorily question whether it can be said to exist. The existence of bias is too readily associated with groups such as women receiving lower degree results compared to their male counterparts. This alone, however, is no proof of bias because in any one group of students it is possible that the male students may be more able than the female students and vice versa.

What is meant by bias? This is an important question to address because the term has been used in different ways to convey very different ideas. A distinction which needs to be made, and was referred to in the Minutes of Evidence taken before the General Sub-Committee of the House of Commons Expenditure Committee in 1975-76 and 1976-77, is the difference between bias as used in a statistical sense and bias in the sense of some sort of prejudice or intentional preference on the part of the evaluators[43]. Clearly, if a sample of women gain fewer first class degrees

---

[43] This distinction in the sense in which 'bias' can be used arose out of an allegation made by Lord Crowther-Hunt that three biases existed in the selection of civil service administration trainees: a

than a sample of men this does not necessarily mean they are being prejudiced against. However, where the proportion of women gaining firsts is not the same as the proportion of total firsts awarded a statistical bias can be said to exist. Plainly, this form of bias is distinct from a set of examiners imparting a prejudice/bias because they know the gender of a student. The bias that is being considered in this context is that linked with prejudice.

It is important to recognise that there are many different forms of bias or prejudice, thus to talk only in terms of gender and race is limiting the value of the discussion as there may be hundreds of deeply felt prejudices at work in each person. For example, the debate concerning bias in marking often fails to mention the conscious or unconscious bias that might disadvantage students who are foreign, illiterate, dyslexic or who have poor handwriting (factors that cannot so easily be eradicated by anonymising examination candidates). Gender or racial bias are not the only possible biases; the issue of ideological bias has always surrounded education and has been a noticeable feature of higher education debates since at least the beginning of the twentieth century (the forming of the London School of Economics and comments about its partisan sympathies is a good example of this).

---

preference for Oxbridge graduates; a preference for former pupils of public rather than state schools; and a preference for graduates with arts rather than natural or social science backgrounds. Sir Douglas Allen, the Head of the Home Civil Service, agreed that a bias existed in these cases, but only in a statistical sense in that "the proportion of applicants coming from these various areas that you have indicated does not happen to be exactly the same as the proportion of successful people coming from these areas." He later adds, "it is not a bias at that stage which is attributable to the people who serve the selection boards imparting a bias because they know where candidates come from." The Civil Service, Eleventh Report from the Expenditure Committee (General Sub-Committee), sessions 1975-76 and 1976-77 and Appendices, HC 535-I and HC 535-II, p. xviii and pp. 808-809 (2 May 1977).

There is also the issue of bias in a more technical sense where the assessments themselves might disadvantage some groups. For example, the content or language of some tests may be more familiar to a particular group, and as a result the performances of individuals outside of this group is likely to be affected, irrespective of whether they are of equal ability.

A further distinction is made by Stephen Newstead and Ian Dennis between 'stereotypical bias' and 'subjective bias'.[44] The latter being more likely if a candidate is known personally to his or her examiner (a like or dislike), and the former might operate if the candidate is known personally or not. If the candidate is unknown, details - e.g. sex, ethnicity - may be inferrred from his or her name and may form a stereotypical expectation on the part of the examiner. Hence, there are various different forms of bias which could exist. It is also very important to keep in mind, as Newstead and Dennis point out, that "more than one bias may be at work in any situation, and in some cases they may work in opposite directions".[45] The extent to which any or all of these factors can be identified as having an influence on the outcome of examination results is open to question. The attempts that have been made by scholars to prove the existence or absence of such biases will be examined in this section, noting the limitations already identified.

---

[44] Newstead, Stephen E. and Dennis, Ian, 'Blind marking and sex bias in student assessment', Assessment and Evaluation in Higher Education, vol. 15, 2, Summer 1990, p. 132.

[45] Newstead and Dennis, 'Blind marking', p. 132.

The bias debate

The essence of the bias debate over the last decade focuses on the attempts made by various scholars to find an explanation for why women students in higher education gain fewer firsts and fewer thirds than their male counterparts. Various explanations and hypotheses have been put forward; bias being only one possible explanation.

Before a study of the literature which specifically explores the pattern of results in higher education is embarked upon, it might be useful to gain some sense of what relevance, if any, research that focuses on pre-university education has to this discussion. This will provide the opportunity to do two things. First, to establish and categorise the types of explanation that scholars have put forward to explain sex differences in performance across the education sector generally. Second, to discover if there is any evidence of there being intellectual differences between the sexes, and if there is, whether these differences remain constant throughout the different age groups. If the differences do remain constant in this respect, then they will be relevant to the debate in higher education.

Caroline Gipps and Patricia Murphy in <u>A Fair Test?</u> provide an extremely useful summary of the types of hypotheses that researchers have used to explain the differences in male and female performance. Three distinct categories emerge: biological or physiological explanations; environmental hypotheses; and assessment related factors which might have a bearing on differential performance.

Biological explanations include those that claim that sex differences can be put down to the differences in the male and the female brain, or in their differing genes or hormones. The acceptability of these sorts of explanation is decreasing in the present research climate, yet the possible influence of such theories should not be dismissed or ignored altogether as they remain unproven. This applies to theories that claim that intelligence is in some sense hereditary and that particular abilities can be traced back to the make up of genes. Although there are studies which make a case for this type of explanation,[46] most scholars tend to agree that cognitive differences are due to more than a particular gene, or set of genes.

Theories which explore the effects of hormonal differences between the sexes are perhaps most interesting in the possible conections they draw between cognitive ability and male and female maturation. Some scholars have argued that the differences in the rate of physical development of males and females can explain why one sex is stronger in a certain cognitive ability area than the other. However, research findings conflict here and tend to be complex.[47] Scholars who have studied hormonal changes might be regarded as most plausible in the connections they draw between male and female maturation and the sex differences at particular ages. Having said this, there is still a lot of doubt surrounding the validity of such theories. It is not clear what impact, if any, hormonal change has on male and female cognition beyond the age when physical development

---

[46] In <u>A Fair Test? Assessment, achievement and equity</u>, Open University Press, Buckingham, 1994, C. Gipps and P. Murphy cite the studies of R. E. Stafford as possible evidence for this. See 'Sex differences in spatial visualization as evidence of sex linked inheritance', in <u>Perceptual Motor Skills</u>, 13, 1961, p. 428; and <u>An Investigation of Similarities in Parent-Child Test Scores for Evidence of Hereditary Components</u>, NJ: Educational Testing Service, Princeton, 1963.

[47] See Halpern, D. F., <u>Sex Differences in Cognitive Abilities</u>, Hillsdale, NJ: Lawrence Erlbaum Associates Inc., 1992, for a survey of studies focusing on the effects of hormonal differences.

has reached its height. For this reason, one must guard against the temptation to make generalisations about the impact of hormonal differences on the cognitive differences of adult samples. Studies which concentrate on the impact of any functional or organisational differences in the male and female brain are equally problematic and inconclusive as D. F. Halpern, and Gipps and Murphy illustrate.[48]

Other scholars like, Dr. S. Clarke (whose arguments will be examined later in the Chapter), put forward what Gipps and Murphy refer to as 'environmental' hypotheses. Increasingly, environmental explanations have become more popular amongst scholars. These include "cultural, social and psychological influences (sometimes called 'psycho-social' variables) that affect the development of individuals within specific groups".[49] Here the influence of factors like gender role expectations is thought to shape the perceptions and performance of men and women. From an early age the socialisation of gender roles and domains, reinforced by parents, teachers and society at large, manifests in certain subjects and activities being considered more masculine or feminine. This affects the expectation of success of, and in, certain groups and the self-image and attitudes of the sexes, such that fewer women will be inclined to study engineering, for example, because it is considered a male domain, and those that do will be more highly motivated.

It has been suggested that institutional factors, such as whether a college is single sex or mixed, might have a bearing on performance. Similarly, institutions vary in their approach to teaching and learning, in their level

---

[48] Gipps and Murphy, A Fair Test?, p. 59; see also Halpern, Sex Differences.

[49] Gipps and Murphy, A Fair Test?, p. 6.

of entry requirements and in their type of student intake. These environmental factors could also affect the performance of men and women differently. For example, N. G. McCrum considers whether women are disadvantaged at Oxford by, among other things, the tutorial system because of the differences in the male and female approach to learning.[50]

Assessment based factors can also have an influence on the performance of groups. This might apply to the content of the test itself or to the administration of a test. The administration of assessment is particularly relevant to this Chapter because bias in marking has been claimed by some scholars to be influential in the differential performance of male and female undergraduates. Equally, however, the introduction of anonymous marking, which is thought by some people to overcome a potential problem that might exist in relation to gender marking bias, has its own problems in disadvantaging groups whose first language is not English or who are dyslexic. This is because no account of these circumstances can be taken at the marking stage where a student must remain anonymous. Special circumstances are considered only at a very late stage in the assessment process. Also, test content may favour one group over another, irrespective of the cognitive abilities of the groups, through the language or the cultural slant it adopts.

---

[50] McCrum, N. G., 'A fair admissions system', Oxford Magazine, 72, 1991, pp. 16-17. For further contributions on the institutional impact of Oxbridge on its results see, McCrum, N. G., 'The second sex', T.H.E.S., 31 March 1995; McCrum, N. G., 'Sixties peak of female performance', T.H.E.S. (letter), 21 April 1995; 'Of motes and beams', T.H.E.S. editorial, 31 March 1995; Goodhart, C. B., 'Women's examination results', The Cambridge Review, 109, 1988, pp. 38-40; Goodhart, C. B., 'Examination results in single sex and mixed colleges at Cambridge', The Cambridge Review, 109, 1988, pp. 139-141; Goodhart, C. B., 'Sex and class in examinations', The Cambridge Review, 113, 1992, pp. 43-44; Stewart, D., 'Women and men', Oxford Magazine, 39, 1988, p. 16; Hannabus, K. C., 'Mixed results', Oxford Magazine, 74, 1991, pp. 4-5; Hannabus, K. C., 'Mixed results', The Cambridge Review, 113, 1992, pp. 40-42; and Marston, P., 'Oxbridge women losing battle of sexes', The Daily Telegraph, 21 June 1995.

Research has indicated that there may be differences in the intellectual/cognitive abilities of males and females. However, the multitude of studies that have been produced by scholars in this area have also shown that there is little agreement about the extent of these differences, or even whether these differences exist at all. Another important point to appreciate is that 'intelligence' can mean very different things in different contexts and to different people. For example, mathematical and analytical ability are different yet both, when considered individually or together, are components of what is commonly refered to as intelligence. Intelligence can be broken down into different areas of ability with some research showing that females are more able in certain ability areas than males, and vice versa for other intellectual activities.

Consider, for example, just three areas of ability where intellectual differences between the sexes have been claimed: verbal, numerical and spatio-temperal ability. Research has shown that with respect to verbal ability, females are more able than males, particularly in adult samples and samples of boys and girls of a pre-school age.[51] Equally, males have been thought to demonstrate greater mathematical ability compared to females which research has shown to emerge in adolescence and continue beyond. The male dominance in mathematical ability is most evident at

---

[51] For examples of studies that produce evidence of this pattern see Halpern, Sex Differences in Cognitive Abilities (Halpern's findings point to consistent sex differences in verbal exercises which involve anagrams and mixed verbal ability tests); Hines, M., 'Gonadal hormones and human cognitive development' in Balthazart, J. (ed.) Hormones, Brain and Behaviour in Vertebrates, Vol. 1, Sexual Differentiation, Neuroanatomical Aspects, Neurotransmitters and Neuropeptides, Basel: Karger, 1990 (Hines found that in tests involving synonym generation females were far superior to males); and Block, R. A., Arnott, D. P., Quigley, B. and Lynch, W. C., 'Unilateral nostril breathing influences lateralized cognitive performance', Brain and Cognition, 9, 1989, pp. 181-190 (Block et. al found that in exercises involving consonant-vowel matching tests females were more able than males).

the highest end of the ability range, with differences between males and females becoming less prominent further down the ability scale.[52] Spatio-temperal or visual-spatial ability also appears to be a domain where males are more able.[53] In their review of the research, Murphy and Gipps suggest that the area of visual-spatial ability shows the largest and perhaps the most consistent intellectual differences between the sexes.[54]

However, not all the research in this area agrees with the cognitive ability trends described above. For example, in the SAT-Maths scores there is little evidence of there being intellectual differences in the mathematical performance of males and females. Similarly, the mathematical performances of the most gifted pre-adolescents are stable according to the findings of C. P. Benbow's study.[55] Conflicting research findings also exist in relation to the verbal abilities of the sexes, with a number of studies showing either that there are no significant differences between the sexes in this cognitive area, or that males are more able than females

---

[52] Stones, I., Beckman, M. and Stephens, L., present findings which show that males significantly outperformed females in geometry, measurements, probability and statistical tests, in 'Sex differences in mathematical competencies of pre-calculus college students', School Science and Mathematics, 82, 1982, pp. 295-299. Also, Hyde, J. S., Fennema, E. and Lamon, S. J., concluded that in mathematical problem-solving test, males perform better than females in high school and college, but that females have a slight advantage in this respect during primary and middle school age. See 'Gender differences in mathematics performance: A meta analysis', Psychological Bulletin, 107, 1990, pp. 139-153.

[53] For evidence of male superiority in tasks involving mental rotation and spatial perception see Linn, M. C. and Peterson, A. C., 'A meta analysis of gender differences in spatial ability: Implications for mathematics and science achievement' in Hyde, J. S. and Linn, M. C. (eds) The Psychology of Gender: Advances Through Meta-Analysis, Baltimore, MD: John Hopkins University Press, 1986, pp. 62-101. See also Schiff, W. and Oldak, R., 'Accuracy of judging time to arrival: Effects of modality, trajectory and gender', Journal of Experimental Psychology: Human Perception and Performance, 16, 1990, pp. 303-316; and, Smith, G. A. and McPhee, K. A., 'Performance on a coincidence timing task correlates with intelligence', Intelligence, 11, 1987, pp. 161-167, for similar findings involving spatio-temporal tasks.

[54] Gipps and Murphy, A Fair Test?, pp. 54-55.

[55] Benbow, C. P., 'Sex differences in mathematical reasoning ability in intellectually talented preadolescents: their nature, effects, and possible causes', Behavioural and Brain Sciences, 11, 1988, pp. 169-232.

at many verbal tasks.[56] D. F. Halpern, in her own review, identifies research which comes to very different conclusions about the visual-spatial abilities of the sexes compared to the studies already mentioned in the last paragraph.[57] Interestingly, H. Fairweather concluded that the literature in this area did not provide enough evidence of intellectual gender differences to warrant scholars theorizing about them. It is, therefore, far from clear what, if any, the differences are in the intellectual abilities of men and women.

An additional problem in drawing satisfactory conclusions from the mass of research in this area surrounds the difficulty in establishing whether a study is valid in terms of sample size and representation; the statistical methods used; the accurate definition of the tests which are used to assess the cogntive ability, and whether the content of the tests properly measured what is claimed and nothing else (i.e. a test may be assessing more than one ability area without the researcher recognising the fact); and whether the study takes sufficient account of the context in which the tests are performed. These are just some of the limitations that Gipps and Murphy refer to which could undermine the conclusions of scholars in this area, and which make lessons to be learnt for the purposes of this discussion extremely difficult to draw.[58]

---

[56] Halpern, for example, in Sex Differences in Cognitive Abilities, concludes that males are more able than females in certain verbal tests like, solving analogies and the verbal part of the SAT. Hyde, J. S. and Linn, M. C., arrive at the conclusion that gender differences no longer exist across verbal abilities, in 'Gender differences in verbal ability: A meta analysis', Psychological Bulletin, 104, 1988, pp. 53-69.

[57] Halpern, Sex Differences in Cognitive Abilities.

[58] Gipps and Murphy, A Fair Test?, pp. 45 and 55.

It is clear, therefore, that this area of research which focuses on the cognitive differences between the sexes is a minefield of conflicting evidence; for every study purporting to have found a pattern in a particular cognitive ability area, there is another which can be found claiming the opposite. Whether any such differences are rooted in genetic or social factors is also contentious and, according to Gipps and Murphy, depends in part on the social and political context of opinion of the day.[59] One feasible explanation of the development of male and female intelligence and performance is that both environmental and biological factors determine the differences between individuals, if not also between genders.[60] The broad similarities, as much as the differences, in the cognitive abilities of men and women are important, if not more important. The point here is that the differences between the sexes should not be overstated.

Clearly, care is needed not to make generalisations, from the research which focuses on intelligence and intellectual performance at a pre-university age, to the sorts of issues that concern performance in higher education. It is not clear how much the cognitive differences between the sexes, where they exist, change over the years. Change, in this respect, does not seem constant across the age range. The different sexes have different maturing speeds at a younger age, and it is contestable how, and

---

[59] Gipps and Murphy, A Fair Test?, pp. 28-29.

[60] This is a view gaining increasing support. Jerome M. Sattler makes this point eloquently and succinctly when he writes:

> Theories of intelligence are beginning to show a coalescing of views, stressing the importance of both innate and developmental influences. Intelligence is viewed as being a central, "fluid" kind of genetically determined basic ability which is modified by experience. However, the ways in which people use their intelligence are determined by the unique learning history of the individual.

Assessment of Children's Intelligence, Revised Reprint, W. B. Saunders Company, Philadelphia, 1974, p.15.

to what extent, this affects the intellectual ability of the sexes in general, and individuals in particular. It is also the case that adults, children and adolescents are involved in different sorts of intellectual activities.[61] For these reasons, it would be unwise to attempt to make generalisations to adulthood based on research that has concentrated on the cognitive performance differences of males and females at a pre-university age. It should be recognised, anyway, that very little concrete evidence exists to make generalisations from.

Although at a glance little seems to have been achieved from taking a look at the research literature on this broad topic, an appreciation of the complexities of the subject area has been gained. Not many answers have come to light in this section; nevertheless, some very important issues have been surveyed. For example, this section has highlighted the types of explanations that are commonly used to understand the differences in intelligence between the sexes. It has indicated the sorts of ability areas that have been studied and which are assumed to characterise human intelligence. It has shown the problems and limitations with the studies in this subject area, and the difficulty with inferring any concrete conclusions from a mass of research which itself produces conflicting findings. These are all issues which need to be borne to mind when considering the evidence in the rest of this Chapter which specifically focuses on research into the performance differences between men and women at degree level.

---

[61] For an illustration of this from an educational, non-psychological perspective see Truscot, Red Brick University, pp. 141-143.

Bradley and sex bias

One of the most important contributions to the debate in higher education as to why women students gain fewer firsts and thirds compared to their male counterparts comes from Clare Bradley. Bradley's findings are among the most referred to by organisations and institutions concerned with the issue of bias in marking and the practice of anonymous marking. Her study has stimulated considerable debate. For this reason, Bradley's findings need to be considered carefully.

Clare Bradley, in her article 'Sex bias in the evaluation of students', argues that the existence of sex bias in marking is an issue that should be taken seriously and not ignored. She makes a case which runs counter to the argument that the differences in academic achievement are attributable to inherent differences in the abilities of male and female students.[62] She states in her discussion of data collected and analysed from five separate university departments:

> *The implications of sex bias in examining are disturbing. If sex bias occurs not only in the assessment of projects but also in the other examinations which contribute to the final degree classification, individual students may be seriously affected. Women students may be receiving lower second class degrees while men students with comparable abilities are awarded upper second degree classes. Weaker men students may be penalized more severely than their female counterparts.[63]*

---

[62] An argument put forward by scholars like R. Dale in 'University standards', Universities Quarterly, 13, 1959, pp. 186-195; R. J. L. Murphy in 'Sex differences in objective test performance', British Journal of Educational Psychology, 52, 1982, pp. 213-219; and more recently E. Rudd in 'A comparison between the results achieved by women and men studying for first degrees in British universities'.

[63] Bradley, C., 'Sex bias in the evaluation of students', p. 151.

How does Bradley come to this conclusion? Bradley conducts a study of the marks awarded to psychology projects of male and female students for which there was initial disagreement between the first and second markers over the class of mark to be awarded to the project. In four of the five departments Bradley analysed, all of the markers had access to the names of the students and hence also to the sex of the students. Bradley claimed to have found evidence of sex bias in the marking of the student projects on the basis that the results gathered showed that relative to the first marker, the second marker marked the projects of the male students more extremely, while a centralising tendency was shown when the second marker marked the projects of the female students. In line with the conclusions of K. Deaux and J. Taynor,[64] and N. T. Feather and J. G. Simon,[65] who had found a similar pattern emerging in the evaluation of the sexes in non-educational sectors (women receiving worse or better evaluations compared to men when both sexes performed well/had good qualifications or performed badly/had poor qualifications, respectively), Bradley saw this pattern of results as evidence that the second markers showed sex bias relative to the first markers. The validity of Bradley's sample is not entirely clear; however, Clarke, who is somewhat sympathetic to Bradley's findings, described her research base as 'limited'.[66]

---

[64] Deaux, K., and Taynor, J., 'Evaluation of male and female ability: Bias works two ways', Psychological Reports, 32, 1973, pp. 261-262.

[65] Feather, N. T. and Simon, J. G., 'Reactions to male and female success and failure in sex-linked occupations: Impressions of personality, causal attributions and perceived likelihood of different consequences', Journal of Personality and Social Psychology, 31, 1975, pp. 20-31.

[66] Clarke, 'Another look at the degree results', p. 324.

It is important to understand that in order to reach the conclusions Bradley does, she makes a number of bold assumptions. First, she goes along with the writings of G. T. Pheterson et al[67]; J. R. Terborg and D. R. Ilgen[68]; and F. S. Hall and D. T. Hall[69] who all argue that the degree of bias in evaluation depends on, among other important factors, the level of inference involved in the evaluation process. In this way, the less information given about the student to be evaluated and the more ambiguous the evaluation criteria, the more likely there is to be bias.

For the purposes of her study, Bradley translates this to mean that the first marker, the supervisor of a student, will be unlikely to be influenced in his or her judgement by any sex role expectation, since he or she will know the student as an individual and will have been involved in the planning of the student's project. By contrast, the second marker is expected to be far more susceptible to sex bias because of the limited information on which the second marker has to base an evaluation through having less contact with the student and his or her work, and less knowledge of the area of study under evaluation. Thus Bradley assumes there will be a greater tendency for sex role expectations to influence the marking of the second markers. Based on this assumption the marking patterns already described are claimed by Bradley to be evidence of the existence of sex bias in the second marker.

---

[67] Pheterson, G. T., Kiesler, S. B. and Goldberg, P. A., 'Evaluation of the performance of women as a function of their sex, achievement and personal history', Journal of Personality and Social Psychology, 19, 1971, pp. 114-118.

[68] Terborg, J. R. and Ilgen, D. R., 'A theoretical approach to sex discrimination in traditionally masculine occupations', Organisational Behaviour and Human Performance, 13, 1975, pp. 352-376.

[69] Hall, F. S. and Hall, D. T., 'Effects of job incumbents' race and sex on evaluations of managerial performance', Academy of Management Journal, 19, 1976, pp. 476-481.

It is easy to get swept along by the polemical nature of Bradley's study, the data seeming to fit exactly the desired conclusions, but it is important to exercise caution and to test the validity of the assumption on which Bradley's argument rests, for if the assumption is questionable then the study's findings are vulnerable. Is it right to assume that the supervisor of a student's project is unlikely to be biased, whether consciously or unconsciously, in favour or against some students, whether along lines of gender or not? Has Bradley done enough to prove this? Who is to say that the first marker (the supervisor) did not show sex bias in this study? Could not these marking patterns have been reproduced without the second marker, or either marker for that matter, showing *sex* bias? All these questions point to potential areas of weakness; answers to these questions will now be considered.

Bradley attempts to prove that the use of the assumption for the purposes of her study is valid by attempting to justify what she refers to as the "less likely possibility that the first marker might also be demonstrating sex bias" by analysing a fifth department.[70] In this fifth department the second markers were unaware of the sex of the students, while the first markers (the project supervisors) were. Bradley predicted that in contrast to the other four departments, the fifth department would be likely to produce marking patterns in which the first marker, relative to the second marker, would mark the projects of the male students more extremely while displaying a centralising tendency with the projects of female students. However, no difference between the two markers was found, thus leading Bradley to conclude that in the fifth department no sex bias was apparent from either marker. Therefore, contrary to what was outwardly expected -

---

[70] Bradley, C., 'Sex bias in the evaluation of students', p. 148.

and conveniently for Bradley's argument - the first marker showed no sex bias relative to the second marker according to Bradley's reading of the data. This Bradley used to confirm the assumption that the first marker's greater knowledge and familiarity with the student and the subject area, made him or her less susceptible to sex bias.

It is difficult to accept, just because two independent markers agree in their marking patterns, as was the case in the fifth department, that this necessarily means that neither of them is biased, nor does it rule out the possibility of other influences over and above gender affecting their evaluation. Similarly, just because two markers show different marking patterns, as was the case in the first four departments studied, it may not mean that one of them must be biased in the way Bradley described. The same marking patterns could have been produced as a result of one or both markers marking on the basis of the legibility of the handwriting or on the extent to which the views of the project were in agreement with the ideology of the marker. Equally, these results might simply be the product of a marker who is tired when marking certain scripts. None of these possible infuences would have been eliminated by not knowing the names of the students. Not enough information is known about the whole situation to be able to say with confidence that one thing or another is making the distribution of marks appear in the way they are. One need only look as far as the classic works on assessment; works like P. Hartog and E. C. Rhodes An Examination of Examinations[71]; R. Cox's article 'Examinations and higher education: A review of the literature'[72]; and the

---

[71] Hartog, P. and Rhodes, E. C., An Examination of Examinations, Macmillan, London, 1935.

[72] Cox, R., 'Examinations and higher education: A review of the literature', Universities Quarterly, 21, 1967, pp. 292-340.

Robbins Report[73], to see that the whole process of marking is highly unreliable. In fact it is open to question whether something like the existence or absence of sex bias can ever be proved beyond doubt, however much information is made available.

There are a number of reasons which suggest it is possible that the first marker could display sex bias in the departments Bradley analysed and in any other department. For example, it is feasible that the first marker does show sex bias, but the influence of other factors serve to counteract this bias such that it doesn't show itself in the marking patterns described by Bradley. Equally, the effect of physical attractiveness is more likely to influence the first marker than the second marker, since it is the first marker who has regular face-to-face contact with the student. The influence physical attractiveness has on marking is far from clear; D. Landy and H. Sigall[74] claimed to have found evidence that 'attractive' females were marked more highly by student markers; R. Bull and J. Stevens[75] in a similar study found no evidence of this; whilst J. Swim *et al.*[76], alert us to evidence that in some situations females who are attractive are deemed to be less able. In spite of the inconclusiveness of this evidence this debate serves to illustrate that supervisors, just as much as second markers who mark blind, are vulnerable to forms of sex bias, an

---

[73] Robbins, Lord C. B., Report of the Committee on Higher Education, Cmnd 2154 (Robbins Report), London: HMSO, 1963.

[74] Landy, D. and Sigall, H., 'Beauty is talent: Task evaluation as a function of the performer's physical attractiveness', Journal of Personality and Social Psychology, 29, 1974, pp. 299-304.

[75] Bull, R. and Stevens, J., ' The effects of attractiveness of writer and penmanship on essay grades', Journal of Occupational Psychology, 52, 1979, pp. 53-59.

[76] Swim, J., Borgida, E., Marayuma, G. and Myers, D. G., 'Joan McKay versus John McKay: Do gender stereotypes bias evaluations?', Psychological Bulletin, 105, 1989, pp. 409-429.

admission that Bradley seems reluctant to make because such an admission would jeopardise her conclusions.

Dr Ernest Rudd, in a two page follow up article to Bradley's article, points to a number of reasons why on occasions the supervisor could be more biased than the second marker, whether positively or negatively. From his own experience he observed the bias that might result from the 'teacher-disciple' relationship that was likely to form in the writing of student projects; the acceptance of the supervisor's "ideas, approach and prejudices" may be likely to result in the supervisor giving too high a mark for the project. Rudd believes that this sort of relationship is most likely to form between male staff and female students. However, he provides no evidence for this other than his own observations. It is difficult to accept that such a relationship would be confined to male staff and female students.

On the other side of the coin, negative bias, on the part of the supervisor, was thought to be a possibilty in situations where students failed to turn up to project meetings and appointments to discuss their projects. Rudd argues this is more likely to affect male students rather than female students, because women as a group are thought to be more conformist than men. Rudd argues this point on the basis that a higher proportion of women are religious and far more women respond to things like postal surveys, thus he concludes women are more likely to be conformist. However, this is at best a tenuous argument which Clarke turns to his own advantage by using it as an example of the sort of differential expectations which help form the stereotypes which contribute to the differing social pressures put on men and women that might affect their results. It is precisely because of the awareness of a supervisor to the susceptibility of

his or her feelings towards his or her disciples that the external examiner system exists.

Rudd makes a similar point about Bradley's over reliance on assumptions, stating in strong terms that "results that follow inexorably from their being fed in as assumptions are worthless".[77] In direct debate with Bradley, Rudd states that most of Bradley's results "follow from her having built into her study as an assumption what she is setting out to prove, namely that failing to award the same pattern of marks to women as to men is the result of gender discrimination". Rudd makes no secret of his views, describing Bradley's conclusions as "no more than a reiteration of her unproved assumptions".[78] Clearly, therefore, it is important to recognise the limitations and potential weaknesses, as well as the positive contribution this study has made in stimulating further research and debate in this area, particularly since advocates of anonymous marking commonly refer to this study, and similar studies, to make their case.[79]

Newstead and Dennis

In a similar study to that of Bradley, Stephen Newstead and Ian Dennis failed to find the sort of bias Bradley claimed. Collecting data from the psychology department in their own polytechnic (Polytechnic South West, formerly Plymouth Polytechnic, now Plymouth University) from three separate years, Newstead and Dennis analysed the marks of psychology

---

[77] Rudd, E., 'A reply to Bradley', p. 96.

[78] Rudd, E., 'A reply to Bradley', p. 95.

[79] For example, the Students' Unions of Warwick, Durham and Edinburgh Universities, as well as the National Union of Students and the Association of University Teachers, refer to the research findings of Bradley's study as evidence that sex bias has been proven to exist. They fail, however, to recognise the potential flaws and weaknesses of these findings.

projects where there was disagreement over the class boundary. They found that, in contrast to Bradley, the results indicated that it was the males rather than the females who were more likely to be marked less extremely. The figures are not statistically significant which means that the above effect is insignificant, making the marking patterns of both markers much the same. As a result, Newstead and Dennis concluded that there was little to indicate that the sort of sex bias Bradley claimed existed in marking in their own study.

Newstead and Dennis also made a comparison between their own psychology department, which did not use blind marking, and a department at another polytechnic which did, finding that the standard deviations of the marks were higher for females than for males which goes against the findings and expectations of Bradley. The data also indicated that whether marking blind or not, made no significant difference to the marks awarded to the sexes. As with Bradley's research, Newstead's and Dennis' study ignored the possible influence of other kinds of bias like, for example, the favouritism which might exist in non-blind marking. The same can be said for various other factors which might have affected the marking process in their study. It is also worth noting that a psychology department, as distinct from many other subject departments, may be more aware of the issue of bias. Thus, the data in this study should not necessarily be treated as representative of other subjects.

Interestingly, the reason for Newstead and Dennis becoming involved in this area of research was due to their higher education institution asking them to "provide evidence that biases existed which might be overcome by blind marking". They made no secret of the fact that they "hoped to

find evidence to support the introduction of blind marking", pointing out that one of them had co-authored a document specifying the recommended policy of the British Psychological Society in favour of blind marking.[80] Having not found evidence of sex bias in their own study, and having considered the evidence in this area as a whole, they write:

> *In the light of such inconclusive evidence, any decision to introduce blind marking will be more a political one than one based on firm evidence.*[81]

Clearly, this is a telling statement from Newstead and Dennis bearing in mind the conclusions they hoped to draw from their investigation. This is not to say that sex bias in marking does not exist, merely that at present there is, in the words of Newstead and Dennis, "not enough empirical evidence to decide either way."[82] In their 1993 article they go on to refer to what they describe as the "powerful" arguments that can be levelled against blind marking; extra administration and the increased likelihood of errors due to examination scripts being nameless (two practical problems that shall be explored in the next Chapter with reference to the experience of the University of Durham).

Despite Newstead and Dennis concluding that they had found no evidence of sex bias in their study, a study modelled very much on Bradley's own study, Bradley herself interpreted Newstead's and Dennis' data very differently. Where Newstead and Dennis found no difference in the marking patterns of the first and second markers and took this to mean

---

[80] The document concerned is, British Psychological Society, The, Guidelines for External Examiners on Undergraduate Psychology Degrees, Leicester: The British Psychological Society, 1989.

[81] Newstead, S. and Dennis, I, 'Bias in student assessment', The Psychologist, Oct. 1993, pp. 451-2.

[82] Newstead, S. and Dennis, I, 'Blind marking and sex bias in student assessment', Assessment and Evaluation in Higher Education, 15, 1990, p. 138.

that both markers were equally unbiased, Bradley counter argued that this was evidence that both markers were equally *biased*. In a highly detailed article, 'Sex bias in student assessment overlooked?', Bradley explains how this may be so. Newstead and Dennis suggest that in their study of data from their own polytechnic department, the supervisor may have had less contact with the students, compared to the closer supervision in the university departments Bradley analysed, due to the difference in teacher work loads between polytechnics and universities. As a result, Newstead and Dennis believe that a polytechnic supervisor would be less prone to personal bias. However, Bradley refutes this point, arguing that the first marker in polytechnic departments would be more, not less, prone to bias because the greatest potential for sex bias, according to Bradley, exists in large departments, where there are a greater number of students and less staff-student contact; accordingly, the marking patterns of the first and second markers may agree, as Newstead and Dennis found. Where both markers are "equally susceptible to this kind of bias and share the same cultural stereotypes" they would be expected to "have fewer disagreements than would be likely if one of the markers was protected from bias".[83] It should be noted that the teacher work load distinction, between what were former polytechnics and universites, is becoming less and less evident; therefore, the extent to which Bradley's line of argument works out in practice is questionable.

An alternative explanation of the conflicting findings of Bradley on the one hand, and Newstead and Dennis on the other, was suggested by Professor John Archer who argued that both studies "probably involved the same markers on several (or many) occasions". As a result, "a few

---

[83] Bradley, C., 'Sex bias in student assessment overlooked?', <u>Assessment and Evaluation in Higher Education</u>, 18, 1, 1993, p. 6.

biased individuals could give the impression of a general bias, which was in fact spurious" and, equally, "a few unbiased individuals used for another study would give a general impression of lack of bias".[84] Thus both studies suffer from not being able to separate each marker as a source of data. In this way, both studies could be misleading in the conclusions they draw.

Hartley findings

In a 1992 article entitled, 'Sex bias, blind marking and assessing students', James Hartley makes a number of interesting points which are worth bearing in mind in the bias debate. In his own study, the possibility of sex bias in the marking of student projects and examinations was examined in a psychology department in a British university. Whilst the details of this study are not unimportant, it is with Hartley's concluding remarks that particular interest lies. The details of this study (sample size, study method, etc.) can be obtained from the article itself.[85]

Hartley concludes his article by pledging support for blind marking (anonymous marking). He does so in spite of the fact that his study has found "no real evidence to support the idea that sex bias occurs in the marking of student projects or examination papers in this particular department".[86] Further than this, Hartley highlights the degree of success that examiners have in identifying the gender of a student through their handwriting, demonstrating that in some cases over 75% of examination

---

[84] Archer, J., 'Sex bias in evaluations at college and work', The Psychologist, 5, 1992, p. 202.

[85] Hartley, J., 'Sex bias, blind marking and assessing students', Psychology Teaching Review, 1, 2, 1992, pp. 66-73.

[86] Hartley, 'Sex bias', p. 73.

candidates can be identified as male or female. Even accounting for the influence on gender identification that the content of certain examination scripts might have, as well as the fact that individual examiners will differ, Hartley states that "despite these qualifications, the results make it clear that blind marking is hardly likely to remove sex bias from examination marking completely".[87] So why adopt blind marking? One of the reasons Hartley gives is because blind marking helps reduce the potential for personal bias, even if it isn't totally effective in eliminating gender bias. Here again, though, anonymous marking is not much more effective than other methods of marking which are not anonymous, because the first marker (the course tutor) will in many cases recognise an examination script as being that of an individual student, particularly if the class size is small. Thus, personal bias, whether it be to the student's advantage or disadvantage, is still possible with blind marking. Where blind marking will reduce the effect of possible personal bias is in preventing a student being prejudiced against because of departmental gossip, which might have conjured up an image of a student in the eyes of a marker who has no direct knowledge of the student. Here, however, external examiners and independent second markers provide a safeguard. These are safeguards which exist whether marking is anonymous or not.

The contribution of Dr Rudd

Bias, and more particularly sex bias, is not the only possible explanation for these figures. Rudd's articles are a good illustration of this. Referring

---

[87] Hartley, 'Sex bias', p. 73. Hartley cites the following studies as examples of those which examine gender and handwriting: Awramoff, D., 1903, cited by Young, P. T., 'Sex differences in handwriting', Journal of Applied Psychology, 15, 1931, pp. 486-498; McCullough, M. L., 'Blind marking and gender identity', Bulletin of the British Psychological Society, 40, 1987, p. 103; Hartley, J., 'Sex differences in handwriting: A comment on Spear', British Educational Research Journal, 17, 2, 1991, pp. 141-145.

to the University Grants Committee's <u>Statistics of Education</u>[88], Rudd concluded that the prevalence of sex bias as a factor in the explanation of the differing degree results of male and female students was not likely, rather the degrees of women are bound to be inferior because of the lower percentage of women compared to men with extremely high levels of measured intelligence. To quote the abstract of Rudd's 1984 article:

> *The only explanation that seems to fit all the facts is that this difference is linked to differences in the distribution of ability as measured by the scores gained in intelligence tests.*[89]

For credence for this view, Rudd refers to the work of A. Heim which identifies a difference in intelligence test scores between men and women.[90] A smaller percentage of women achieve intelligence test scores at the top and at the bottom of the mark range - "fewer geniuses and fewer dunces", as Rudd summises. This difference is held by Rudd to be the most plausible reason for why women gain fewer firsts and thirds at degree level, and a comparatively small number of women achieve the highest grades at 'A' level (the possibility of examiner bias at 'A' level was dismissed on the basis that the examiners will not know the students, however, Bradley would counter argue that this makes an examiner <u>more</u>, not less susceptible to bias).

The argument of Rudd fails to appreciate that intelligence tests are different from degree examinations. The two forms of assessment are not measuring the same sorts of things; therefore, performance in one is not necessarily a good indicator for performance in the other. This is a point

---

[88] University Grants Committee, <u>Statistics of Education</u>, vol. 6, London: HMSO, annually 1967-1979.

[89] Rudd, E., 'A comparison between the results achieved by men and women studying for first degrees in British universities', p. 47.

[90] Heim, A., <u>Intelligence and Personality</u>, Harmondsworth, Penguin, 1970.

clearly made by Tol Bedford, the Director of Research of Recruitment and Assessment Services (RAS), in relation to the cognitive tests used in the Civil Service Administrative Fast Stream selection process. Bedford states that "educational attainment, though related to IQ, is by no means synonymous with it".[91] The ability to think critically and to be able to develop a reasoned argument are good examples of skills which are required for degree performance, but not for intelligence tests.

Interestingly, Bedford points to research which supports the sorts of findings described by Heim above: first, the *average* IQ of males and females is nearly identical; second, male IQ has a broader distribution than female IQ; and third, this broader distribution means that there are more men than women at either extreme of the IQ distribution.[92] In relation to civil service recruitment this poses a problem for women since the fast stream Qualifying Test is selecting people right at the top of the IQ distribution (efforts to redress the perceived disadvantage for women have resulted in RAS assessing biodata through a supplementary application form).[93] However, in relation to the performance of women compared to men at degree level the situation will not necessarily follow the same pattern because of the very different skills demanded. Furthermore, the use of intelligence/IQ tests to illlustrate sex differences is a pointless exercise because the tests themselves are designed in such a

[91] Correspondence with Tol Bedford, Director of Research at Recruitment and Assessment Services, 6 Dec. 1994.

[92] For a good summary of the evidence see Sattler, J., Assessment of Children's Intelligence and Other Special Abilities, 2nd edition, Allyn and Bacon, Boston, 1988.

[93] See Chapman, Richard A., 'Civil Service Recruitment: Fairness or preferential advantage?', Public Policy and Administration, vol. 8, 2, 1993, pp. 68-73; Carter, Adrian, 'Reply to Richard A. Chapman article "Civil Service Recruitment: Fairness of preferential advantage?"(volume 8, no. 2)', Public Policy and Administration, vol. 8, 3, 1993, pp. 46-48; Harrison, Paul, 'The CSSB Supplementary Application Form: A candidate's reply to Adrian Carter', Public Policy and Administration, vol. 9, 1, 1994, pp. 65-67.

way that they do not favour either sex. Therefore, as J. Ryan explains, "items showing large or consistent sex differences are excluded. This fact vitiates all attempts to show sex differences in ability by use of intelligence tests, as is sometimes done".[94] This would seem to cast doubt on Rudd's hypothesis.

In the course of advancing his argument Rudd explores four possible reasons for the sex differences in degree results, some environmental and some biological. These are prejudice in marking, social pressures, medical or psychological differences, and the differences in measured intelligence, as discussed above.

Rudd dismisses male prejudice as a cause of women gaining a lower proportion of firsts on the basis that there is no conclusive evidence for this claim. He argues that the fact that women perform relatively well in traditionally male dominated subject areas like, for example, in engineering where bias is thought to be most likely, and not particularly well in subject areas like sociology and the Arts in general which don't have the same stigma attached, and in which there is a greater awareness of the possibility of bias and, therefore, a greater effort to eliminate any effect, indicates that bias is not prevalent. However, Clarke makes the point that virtually all subject areas at university are male dominated, quoting statistics from the Universities Statistical Record.[95] Hence, one

---

[94]Ryan, J., 'IQ - the illusion of objectivity', in Richardson, K. and Spears, D. (eds), Race, Culture and Intelligence, Penguin, London, 1972, p. 49.

[95] Clarke, S., 'Another look at the degree results', p. 324. Clarke quotes the following statistics as evidence of the male dominance of universities: percentage of university staff who are male-
      Sciences: 90 %
      Social Studies: 85 %
      Arts: 80 % +
Taken from University Grants Committee, Statistics of Education, vol. 6, annual to 1979, HMSO, London.

could argue that women are performing universally worse than they might otherwise have done, if marking bias is apparent. An important point to make here, is that female members of staff, as well as their male colleagues, are susceptible to the same forms of stereotypical bias, thus, the discussion should not assume that all prejudice is male prejudice. Rudd does make the point in a reply article to Bradley that he himself has not proved that gender discrimination does not exist, rather that he finds it an unlikely explanation of the differences that have been observed for the degree results of men and women.[96]

Clearly, the existence or absence of bias cannot be proved by simple reference to data like that given above. A whole host of other factors could explain the pattern of results referred to here, not least the fact that the Sciences (engineering always being a good example) award significantly more firsts than the Arts (see pp. 63-64), thus, statistically women in the Sciences have a better chance of achieving high grades. Equally, the motivation, drive and character of women entering what are perceived to be 'male domains' could have a bearing on the results.

The biological effect of menstruation or stress on the degree results of women was another possible explanation explored by Rudd. Evidence produced by K. Dalton is cited which shows that before and during menstruation the academic performance of girls at 'O' and 'A' level is lower than average, and that the stress of examinations increased the number of girls menstruating, thus accentuating the problem.[97] This biological explanation was thought not to be the most plausible because it

---

[96] Rudd, E., 'A reply to Bradley', p. 95.

[97] Dalton, K., 'Menstruation and examinations', Lancet, 28 December, 1968, pp. 1386-1388.

couldn't explain the shortage only of firsts, as all examination performance would be affected. Having said this, Clarke suggests that were it not for such biological conditions women might do better across the board.[98]

The last factor which Rudd explores is the possibility that women are less driven than men. Again, however, this psychological difference is dismissed on the basis that it would cause lower results for women right across the ability range, it would not prevent women from attaining firsts only. Here, though, no account is taken of the differences in motivation between men and women entering different subjects, such that a woman choosing to do a professional subject which has always been considered a male preserve like, for example, engineering or business management (two subjects which Rudd picks up as being subjects in which women have perfomed relatively well), will be more highly motivated to succeed than in other subject areas.

In a response to Rudd, Clarke dismisses Rudd's hypothesis as a plausible explanation for the sex differences in the degree results, instead concluding that 'discriminatory social and institutional pressures' (i.e. 'environmental' factors) lie behind this phenomonen. In particular, sex stereotyping, of which marking bias can be a feature, is thought to be the problem, such that women in all areas of university life are being required to conform to the expectations of a culture that is male-dominated:

> *These expectations structure the form and content of teaching, they underlie the informal processes of academic sponsorship and of pastoral care, and they are expressed in the stereotypes in the back of examiners' minds discussed by Bradley ... The problem is not that women conform, but that*

---

[98] Clarke, 'Another look at the degree results', p. 325.

they conform to a set of expectations that do not bring them success.[99]

This is the reason why Clarke believes women under achieve at degree level. The gap in Clarke's argument is that he does not actually show how the social and institutional pressures he sees as being at the root of the problem cause the pattern in differential degree results for men and women.[100] He simply advocates the need for further research.

Subjects and results

Up to this point in the discussion the differences between the degree performance of men and women has focused on the pattern of degree results as a whole; however, this approach can be misleading as Clarke illustrates in his article, 'Another look at the degree results of men and women'.[101] By looking at individual subjects one can see that there are noticeable differences in the performances of men and women which are not reflected in the pattern of degree results as a whole. Thus, the results at the subject level could be contributing to a distorted picture of the overall pattern of results. Very few scholars in this debate recognise this fact, and as a result they draw conclusions which perhaps misrepresent the sex differences in results.

The following are examples designed to show how the pattern of results at subject level does not always correspond with the pattern of results for all degrees combined. Clarke's article highlights that in nine out of the thirty-five subjects covered in the UGC's, <u>Statistics of Education</u>, women

---

[99] Clarke, 'Another look at the degree results', pp. 324-325.

[100] This is a point recognised by Ernest Rudd in his article, 'Reply to Clarke', p. 333.

[101] Clarke, 'Another look at the degree results', pp. 315-331.

got at least as high a proportion of firsts as men, although in no subject did women get more firsts than men. Also, in ten of the fourteen subjects where men got significantly more firsts than women, they also got more good honours degrees compared to women. In other words, the men simply performed better than the women in these subjects, rather than them performing at either end of the mark range as the overall pattern of degree results suggests they should. Again, counter to the overall trend in results, women did particularly badly in the Arts, Mathematics and Physical Science subjects and were awarded significantly more thirds than men in Mathematics and Music.

It is also the case that certain subjects like, for example, the Science subjects, award many more firsts than subjects in the Arts or Social Sciences. In 1994, the total number of firsts awarded in the Faculty of Science at Durham University was 107 out of a total of 663 passes, compared to 45 firsts in the Faculty of Social Science out of 687 total passes.[102] This is a large difference which is significant in that men predominate in the Sciences. Clarke confirms this pattern by referring to the data for all universities. For example, 48% of male students and only 23% of female students studied subjects in Science and Engineering, yet 60% of the total number of firsts, and 56% of the thirds, were awarded in this subject area. To add to this, women in this subject area were mainly in the Biological Sciences which awarded fewest firsts and thirds.[103] The inevitable outcome of this is that numerically, at least, men are bound to attain more firsts and thirds than women. It might even be the case that

[102] University of Durham, 'First degree result analysis by course and Faculty', Planning Section, June 1994.

[103] Clarke has extracted this data from the University Grants Committee (UGC), Statistics of Education, vol. 6, HMSO, London, 1976-79.

the pattern of degree results overall and/or from subject to subject differs from one university to another. From published statistics this is very difficult to check. Therefore, it is clear that greater attention needs to be paid to researching the results, and the reasons for these results, in individual subjects rather than for degrees as a whole. The analysis of individual subjects is all too often ignored to the detriment of this debate.

The University of Wales College of Cardiff (UWCC)

Research into the influence of gender on the pattern of degree results has been a noticeable feature of three scholars at the UWCC. Chris Weedon, lecturer in the Department of German, studied the results of the Faculty of Arts between 1977-1981 and her analysis formed part of a discussion of the role of gender stereotypes in education and Britain at large. Weedon found that there was a much higher percentage of male students getting firsts and upper seconds compared to female students, and that the majority of female students were gaining lower seconds even in departments where women far outnumbered men.[104]

Catherine Belsey, now Professor of English and one time member of the Women's Committee of the AUT, compared the degree results before and after anonymous marking in her own department, between the years 1977-1981 and 1985-1988. Belsey's findings are described below, as are the findings of a Report produced by Nina Parry-Langdon who examined data at the same institution but over a longer period (1977-1989).

---

[104] Weedon, C., 'Engendering stereotypes', Journal of Literature, Teaching and Politics, 1, 1982, pp. 37-49.

Belsey

Figures quoted from Catherine Belsey's examination of the degree results of her own department, the Department of English at the University of Wales College of Cardiff, are often used as evidence of the desirability of anonymous marking, or "marking by numbers" as it is referred to in this department. The table below sums up her findings.

	ENGLISH DEPARTMENT (UWCC)	
	% OF MEN STUDENTS GAINING 1ST OR 2.1	% OF WOMEN STUDENTS GAINING 1ST OR 2.1
1977-81 (BEFORE MARKING BY NO.'S)	45	27
1985	54	47
1985-88 (AFTER MARKING BY NO.'S)	61	50

Belsey interpreted her statistical findings as evidence that discrimination against female students had existed and concluded that the introduction of 'marking by numbers' had made a "startling difference to the examination performance of women students" while not disadvantaging men students, who also did better.[105] Interpreting these figures, Belsey assumes that gender discrimination is the sole explanation. This may be so, but it could equally be argued that in each of the years studied the male students were more able than the female students, or perhaps the minority of men who chose to do English were more highly motivated in a subject which was studied predominently by women at UWCC. Both of these factors could

---

[105] Belsey, Catherine, 'Marking by numbers', <u>AUT Woman</u>, 15, Autumn 1988.

explain the figures presented above, without any reference to gender discrimination. Furthermore, no account of the changing proportions of the male and female student populations studying English in the two periods of analysis has been taken. This is something that Nina Parry-Langdon is concerned with in her own study of UWCC figures which reports different findings. It is also important to recognise the difficulties involved in measuring different student years because one is not measuring like with like. The abilities of the students, the courses, the markers and the assessment procedures themselves will change from year to year. Therefore, a conclusive picture will never be gained with a comparative analysis of degree results of this sort. This point needs to be explicitly stated in any interpretation of data of this sort. This is something Belsey fails to do.

The significance of the figures Belsey presents are themselves doubtful. The sample of students used in the English Department is unclear from her article reporting her findings, but what is clear is that it contained only a very small number of men (the figure mentioned by Belsey was less than 20 men a year between 1985-88, and in the earlier period 1977-81 it is known that approximately 80% of English students were female).[106] Questions over the statistical validity of Belsey's study need therefore to be raised.

It is interesting to note that Belsey was a member of the Women's Committee of the AUT which raised the issue of anonymous marking; the issue has since gained the full support of the AUT, becoming its national policy in December 1994. The policy statement itself is open to the

---

[106] Belsey, 'Marking by numbers'.

criticism that it *assumes* the existence of gender discrimination (see Appendix II for the full AUT policy statement supporting anonymous marking).

Parry-Langdon Report

Nina Parry-Langdon examined data in the same institution, but this time for students in the Faculty of Arts rather than just the Department of English. Parry-Langdon's report drew a very different picture. This report was significant in that it examined data over a period of thirteen years (1977-1989). For the first eight years the examination system was based on the practice of marking the students' final examination scripts with names on. The subsequent five years witnessed the operation of a system of 'marking by numbers', otherwise known as anonymous marking. A comparison of before and after anonymous marking was made. In particular, the results gained by men and women were studied. The following conclusions were drawn from the data that was analysed:

> *Marking by numbers has not increased either the percentage of women getting good degrees or the percentage of good degrees going to females by a statistically significant figure.*

> *In the last 5 years, the Faculty has awarded a greater proportion of 'good' degrees compared to the previous 8 years: 51% compared to 38%. This increase is more striking than any observation on gender differences discussed in this report.*

> *However, marking by numbers is believed by staff and students to be more equitable, because it removes the possibilities of gender bias. The disadvantages are all minor administrative matters which are relatively easy to alter.*

The statistical evidence does not substantiate the contention that either sex is advantaged or disadvantaged by marking by numbers. Therefore it can be introduced without fear of damaging the life-chances of either sex.[107]

A number of interesting points can be drawn from the findings of this report. First, counter to the claim of Catherine Belsey about the English Department, gender differences in the marks received by students across the Faculty of Arts do not seem significant. Second, anonymous marking does not seem to have made a great deal of difference to either sex, although the proportion of firsts and upper seconds awarded appears to have increased generally. This seems to suggest that neither sex was being prejudiced against before the introduction of anonymous marking. Despite the fact that Parry-Langdon and Belsey draw very different results from their analyses, each use their results to argue in support of blind marking. This is, perhaps, more surprising in Parry-Langdon's case as the conclusions she draws (outlined above) suggest there is no problem in terms of gender bias in marking and that the effect of introducing a system of anonymous marking is minimal. One could argue that these are very good arguments against, rather than for the need to introduce anonymous marking.

The Report also stresses the importance marking by numbers has in relation to the public assurance it gives. This is something which comes out of the case study of the experience of Durham. Two further points of interest arise. First, Parry-Langdon states that the only real problems with

---

[107] Parry-Langdon, N., '"Marking By Numbers": Evaluation of the marking of final degree examinations in the Faculty of Humanities and Social Studies', report prepared for the Deans Committee, University of Cardiff Social Research Unit, May 1990, p. 1.

the marking scheme are administrative and that these are relatively easy to overcome. This is highly contentious. Of the thirteen higher education institutions replying to the survey conducted by the City University in April 1989, less than half (five) said that their system of blind marking did not cause extra administration while eight institutions said it did.[108] The administrative problems encountered by the University of Durham, many of which were also present in the Faculty of Arts at the UWCC, will be discussed in the next Chapter.

The second point concerns Parry-Langdon's treatment of sex bias. From the text quoted above, it is assumed that a system of marking by numbers automatically reduces the possibilities of gender bias in marking. However, the gender of a student can in many cases be recognised from a student's handwriting, as has been highlighted earlier in the Chapter. Therefore, the possibility of gender bias is not removed by marking anonymously. Similarly, no account seems to have been given of the groups of students a system of marking by numbers might prejudice against, as well as the other forms of bias that anonymous marking does not prevent - e.g. ideological bias. Whilst neither sex may be disadvantaged by blind marking, foreign or dyslexic students may, for the reasons that have already been discussed. The 'life-chances' of these students, as Parry-Langdon puts it, may be damaged. These are issues which need addressing.

---

[108] See Appendix I for the survey conducted by the City University, April 1989.

Association of University Teachers

Problems arise from the bias debate as a result of some influential organisations assuming that the claims of scholars, and the conclusions they have drawn from data they have collected, have been proved beyond doubt. Little recognition is made of the fact that the studies that have so far claimed to have found the existence of sex bias are highly contentious. This is not to say that prejudice along lines of gender never influences the marking process, rather that no conclusive evidence exists which categorically proves it does or it doesn't. The issue is still very much an open one.

This point is not recognised by organisations like the Association of University Teachers (AUT). In a recently endorsed policy statement on anonymous marking (*see Appendix II*) the AUT states:

> *Bias in the marking of student's written work has been extensively researched, particularly in relation to sex bias. Comparisons made before and after the introduction of anonymous marking provide **strong** [my emphasis] evidence that bias in marking is at least part of the reason for otherwise unexplained differences in performance between male and female students.*

Is there 'strong' evidence to show this? The AUT is making a very big assumption considering the available research on the matter which has been shown to be open to question. The paper continues to assume the existence and influence of bias without ever entertaining the fact that at present there is little evidence to show that bias is an *actual* problem rather than a *potential* problem. One might think this is relatively unimportant. However, the AUT's document was prepared as a paper to be considered, discussed and voted upon by AUT Council. It has, since

December, been endorsed and is now widely accessible. For many people it will be accepted as a reliable statement and the only source of information upon which decisions are made. For these reasons, it is paramount that the AUT's paper should be fully informative, particularly since the opening section is entitled, 'The case for and against anonymous marking' which in itself leads readers to assume that the arguments have been presented carefully and objectively. By making the assumption that bias in marking exists without offering any justification for this claim, and by pressing all institutions to support the effective implementation of anonymous marking on this basis, the document makes a very one sided case which is campaigning and dogmatic in its style.[109]

Racial bias

This Chapter has focused primarily on bias as a possible explanation for why women get fewer firsts than men in higher education. It has concentrated mainly on gender bias because that is where the bulk of research lies. It is more difficult to look at the possibility of something like racial bias because (a) the sample size at many higher education institutions, Durham being just one example of this, is too small to enable adequate research and (b) no statistics are available which link the ethnicity of a student to his or her degree examination results. Therefore, it is extremely difficult to conduct research into the possibility of racial discrimination in marking.

---

[109] The research sources the AUT bases its policy statement on are given in Appendix IV. Belsey, among others, are cited. Virtually all the sources have been considered for this thesis and none have produced conclusive evidence of the existence of bias in marking.

Two examples regularly cited by Student Unions in their literature advocating anonymous marking claim suspected racial bias. They refer to the experience of the University of East London and the University of Glasgow's Dental School. At the University of East London, research showed that black undergraduates received marks that were on average 4.2 % lower than white undergraduates between 1987-1989. One must be clear that this statistic does not necessarily mean that racial bias was the reason for this pattern. Any number of factors could lie behind this difference in results, therefore one must be careful before jumping to conclusions which might form the basis of important policy decisions. For example, there may be a genuine difference in the ability of students during these particular years. The figures alone do not exclude this possibilty. At Glasgow, anonymous marking was introduced because in 1990 there was controversy over the fact that out of the student population who failed, 80 % were Asian, whilst only 20 % of the total student population at the Dental School were Asian. Once again, this figure on its own does not conclusively prove the existence of racial bias in assessment; equally, it does not conclusively prove that assessment practice at this institution is bias free. For concrete evidence either way, in depth scholarly research is required.

Concluding remarks

It is clear from this Chapter that the issues of bias and anonymous marking are closely interrelated. There has been much debate over the reasons for men gaining a larger proportion of first and third class degrees compared to women, and for the degree results of women tending toward the centre of the mark scale. Bias in marking has been put forward as one explanation for this phenomenon. Having examined the literature, there is

little evidence to conclusively prove that sex bias in marking is a problem. This is not to say that bias, whether sex bias or other forms of bias, do not influence the marking process, but that research has not shown them to do so. There is little firm evidence either way. Equally, assessing the impact that the introduction of anonymous marking has had in various cases, has proved far from easy. Problems in analysis, sampling and research methodology have limited the conclusions that can be drawn. Quite clearly more research into departments, in particular, is required because the pattern of results in departments, with or without anonymous marking, do not display the same pattern for degree results as a whole. Building upon this and earlier Chapters, Chapter Three will focus on the practical implications of introducing a university-wide system of anonymous marking.

CHAPTER THREE: A STUDY OF THE PRACTICAL EXPERIENCE OF THE UNIVERSITY OF DURHAM IN 1994.

This Chapter will examine the introduction of a system of anonymous marking at the University of Durham. The material in this case study is intended to highlight the sorts of issues which are raised by anonymous marking at a practical level, as well as presenting an example of the sort of activity which is inherent in a policy decision and policy change. It is important to be clear that the Durham experience is a single, isolated case which has its own unique features; its own culture; interests and actors. The views and information presented should be understood in the Durham context, and against the methodological restrictions which arise from the limited time span with which the system of anonymous marking has been running. The experience of other higher education institutions undergoing similar change may be very different. They would be worth studying in their own right.

A plotted history/chronology

18 Nov., 1985   A joint meeting of Senate and Durham Students' Union recommended that an examination system using numbers instead of names should be looked into. On December 3, Senate agreed to "ask the Registrar to look at the practice in other universities and to report back".
[Senate Minute (S.M.) 304]

11 March, 1986   Senate agreed with the Joint Committee of Senate and Durham Students' Union that the matter be referred to the Boards of Studies who should refer their views to the Boards of Faculty for further

consideration. The Registrar was also to draw up a note outlining the consequences of a system of anonymous examinations.
[S.M. 541]

13 May, 1986   Senate agreed with the Registrar and the Boards of Studies that there should be "no change" to the present examination scheme which was not anonymous.
[S.M. 688]

19 June, 1990   Senate endorsed the recommendations of the Regulations and Admissions Committee which had considered a paper from Durham Students' Union pressing for anonymous examination papers. In view of the "caution" expressed by many Boards of Studies Senate agreed that there was "insufficient support at present to make a major change to the general University marking procedures". Some departments already ran their own schemes of anonymous marking. They could continue to do so.
[S.M. 762-763]

9 March, 1993   Senate noted that the Regulations and Admissions Committee had "agreed, with four votes against, that in all University examinations, except for higher degrees, the scripts of candidates taking formal, timed, written examinations should be marked on an anonymous basis". A Working Party consisting of the Examinations Officer and the Faculty Deputy Deans, was set up to prepare a scheme of anonymous marking to be introduced from October 1993.
[S.M. 305-308]

11 May, 1993   Senate agreed to approve the scheme of anonymous marking recommended by the Regulations and Admissions Committee which was based upon a Report from the Working Party. It was agreed that

the scheme would be reviewed after three years of operation "unless in the meantime there were found to be major difficulties".
[S.M. 400-402]

16 August, 1994   Following a number of criticisms of the anonymous marking scheme as operated in 1993/94, the Senate Teaching and Learning Committee (TLC) agreed that the Advisory Group on Examinations and Assessment (AGEA) should review the system. The Group was to make recommendations for changes to improve procedures for 1994/95. The principle of anonymous marking was not under review.

29 Nov., 1994   Senate approved the revised arrangements for the anonymous marking scheme to be operated in 1994/95 which were recommended to the TLC from a Report by the AGEA.

Background

It is clear from the chronology above that university-wide anonymous marking for undergraduate written examinations took a long time to manifest in practice from being a proposal from Durham Students' Union (DSU), almost ten years in fact. The principle was eventually endorsed by Senate on DSU's third attempt in 1993. Up until 1993, the general consensus was that those who favoured a scheme of anonymous marking were in a small minority, thus no change was deemed necessary. Before discussing the examination systems that existed before anonymity was implemented, and the scheme that is currently in operation, it might be useful to examine the reasons why anonymous marking was not introduced before 1993.

The 1986 attempt

In 1986, Senate agreed that there should be no change to the present system of examinations which were not anonymous. This was based on the recommendations of the three Boards of Faculties which had considered the views of the Joint Committee of Senate and Durham Students' Union, the Registrar and the Boards of Studies. The following reservations about a system of anonymous examinations were reported in Senate Minutes. The Board of the Faculty of Arts noted their concern about the practical difficulties involved in anonymous marking. It was felt that anonymity could never be achieved for all students, particularly at a university like Durham because, presumably, the class sizes were relatively small. It was also contended that anonymous marking would result in a likely delay for examination procedures. Fairness was already deemed to be protected by the existing marking procedures which involved double marking all examination scripts and external marking a selection of these.

The Board of the Faculty of Science made a number of points against introducing anonymity in assessment. First, "students in some departments ... were not enthusiastic for anonymity in examination scripts and did not like to feel anonymous".[110] Second, a number of practical problems were anticipated. These ranged from concern about the possible risk of error that could occur when transcribing the marks of students, to questions over when in the assessment process medical evidence should be considered. It was also felt that the system would encounter problems where internal markers recognised the handwriting of a significant

---

[110] University of Durham, Senate Minutes, 13.5.86, Minute 694 (b).

proportion of students. The amalgamation of results was thought to present difficulties particularly when the marks of assessed work, which were not anonymous, were combined with the marks of assessments which were examined anonymously. Concern was expressed over whether course tutors would be allowed to invigilate their own examinations so that any student queries could be addressed there and then without delay. Questions were also raised about how the university would evaluate such a scheme. One comment made by the Board was that for such a scheme to be effective "bias would have to be almost universal amongst the Board of Examiners" which is an unlikely situation. Once again, there were fears that the scheme would extend the time needed for marking the examinations; estimations ranged from an additional two days being incurred to up to a week in this respect.

The Board of the Faculty of Social Sciences preliminarily reported that it had reservations and other methods might be considered instead.

The 1990 attempt

In 1990, Durham Students' Union raised the issue of anonymous examinations again, by presenting a paper to the University's Regulations and Admissions Committee. At that time, Senate decided that there was insufficient support from the Boards of Studies, who voiced their caution through the Boards of Faculties, for a major change to anonymous marking. It was decided by Senate that those Boards of Examiners that wanted to run their own pilot scheme of anonymous examinations could do so, and the results would be reviewed by the University at a later stage. Senate endorsed the fact that error and bias should be safeguarded

against by the use of double and external marking as standard university practice.

Among the opposition expressed, there was some fear that the marking process would become much slower and less efficient (an objection raised previously), jeopardising Durham's traditional practice of holding all degree ceremonies within a week of the end of the Easter term (a practice which the University prides itself on and which compares very favourably with practice in almost all other universities). If the duration of the terms had to be changed to accommodate the extra time it took to complete the examination process and the awarding of degrees, then this would have a direct effect on the vacation earnings of the University. There was also concern that anonymous marking would affect the implied relationship of trust between students, teachers and examiners which was thought by many academics to be the cornerstone of the existing system. Some individuals believed that the morale of the academic community would be seriously affected in an adverse way (something not in the best interests of the university).

The Sciences were particularly concerned with the increased complexity anonymous marking would add to assessment practice, especially in light of the fact that much of the work carried out in this Faculty was project or practical work and involved presentations that were continuously assessed. There was a general feeling that there was less of a need for anonymous marking in the science subjects because the work was more factual and, hence, the marking of assessments was less subjective than it might be in the Arts or Social Sciences, for example.

The 1993 attempt which resulted in the planned introduction of a scheme of anonymous marking

There are four possible reasons why anonymous marking was accepted in 1993 after the issue had been rejected twice previously years before. Perhaps the most important, in terms of creating the right conditions for the Boards of Studies to give it a chance, was the fact that anonymous marking was to apply only to written examinations and these examinations were to be centrally time-tabled. This certainly would have gone a long way to alleviating the concerns of the Faculty of Science which voiced opposition to every form of assessment being run anonymously. Second, the climate of opinion in higher education and society at large contributed to conditions that made a policy like anonymous marking more acceptable. Political correctness and equal opportunities were influential in this respect. Third, according to Dr Charles Shaw, who was sitting on Senate at the time in question, much discussion was focused on who would be responsible for coping with the new system. The impression given was that the bulk of the work of an anonymous scheme would be on the Examinations Department rather than the departments themselves. The reality of the situation has been the other way around with the departments bearing the weight of the work in 1994, much to the annoyance of some of them. It seemed that while Senate discussed the pros and cons of anonymous marking, the exact manner of the scheme was not decided upon. One could argue that, as a result, Senate did not get caught up in the practical details and implications of how a particular system would operate (an issue with which the Boards of Faculties expressed their repeated concern in 1986 and 1990). Instead, a Working Party decided upon the particular system and Senate, in effect, went along with the chosen system once they had agreed on the principle

of anonymous examinations. This was a possible weakness of Senate. Shaw suggested that Senate had very few members who were Chairmen or Secretaries of Departmental Boards of Examiners, and therefore it had an incomplete understanding of the day-to-day practices involved in the examination process. Consequently, the majority of people sitting on Senate would have had a very limited grasp of the extra levels of complexity that anonymous marking would add on a practical level. This is something the Working Party should have appreciated. Last, there was a change of Vice-Chancellor at the University which may or may not have had a bearing on the willingness of Senate to pursue and subsequently approve the matter.

Before anonymous marking

Before a scheme of anonymous marking was introduced across all departments of Durham University, each department adopted its own procedures for assessment. Some departments like Geological Sciences, Geography, Psychology, History and Sociology and Social Policy already marked anonymously, though their schemes differed from one another. Other departments did not mark anonymously, but continued to double mark with the names of students on the front of examination scripts. All departments, whether marking anonymously or not, practised double marking and external marking (for a selection of examination scripts) for final honours students.

The scheme of anonymous examinations operated by the University of Durham

It is clear from the literature provided by the National Union of Students that there are many different ways in which a system of anonymous marking can be operated. Some schemes require the examination candidate to write his or her name on the examination script and then to conceal it, other schemes avoid using names altogether, prefering to use only codes as an identifier.[111]

The scheme operated in 1993/94: The following scheme of anonymous examinations was approved by Senate on 11 May 1993:

"(i) A system of anonymous marking for undergraduate written examinations should be introduced, whereby the candidate is instructed to:-

> *(A) Write his or her name in the top right-hand corner of the examination script.*

> *(B) Turn down that corner of the form as a flap to conceal the name, and then to secure the flap with the sticky label supplied.*

> *(C) Not to write his or her name on any other part of the examination script or supplementary answer-book.*

(ii) Each main answer-book would bear a pre-printed number, consisting of a letter and four digits (e.g. A1234); the candidate would be instructed to enter this number on any supplementary answer-book, in a blank box in the same position as the number on the main answer-book.

---

[111] For a more detailed description of the different types of anonymous marking that have been operated see, 'Anonymous Marking', NUS Scotland briefing, date unknown.

(iii) At the end of each examination candidates (sic) *to remain in their seats until the examination scripts had been collected from each desk.*

(iv) Examiners would mark scripts without breaking the seal over the candidates' names; pairs of examiners would confer and would draw up their lists of agreed marks using only the numbers on the answer-books. It was important that they did not at this point break the seals on answer-books and discuss marks with reference to candidates by name.

(v) Examiners would communicate their agreed marks, still referring only to numbers and not to candidates' names, to a designated member of their Board of Examiners, who would have the responsibility of transcribing them to a marksheet bearing candidates' names. This would obviously involve the breaking of the seals on examination scripts. Because of the burden of work which this would bring, Boards of Examiners should consider the nomination of different individuals to take responsibility for different year groups.

If it happened that scripts were to be sent by post to the External Examiner, it would be necessary to break the seal in order to record the marks given by the Internal Examiners. In such cases the scripts could then be resealed.

(vi) Marksheets bearing marks against candidates' names (which might contain marks provided by other forms of assessment) should remain confidential to the individual responsible for it until the last set of marks had been entered.

(vii) When marksheets were completed, Boards of Examiners could implement their procedures for the monitoring of borderline cases and the consideration of special cases. It was accepted that meetings of Boards of Examiners might be conducted with the identity of students known.

(viii) Candidates who sat examinations in special rooms or who arrived late for examinations should be provided with answer-books of the same colour and design as other candidates; special arrangements should be

*notified to the Chairman or Chairwoman of the Board of Examiners
concerned.*

*(ix) Instructions to Examiners should be revised to take account of these
new arrangements. Boards of Examiners should modify their statements
of procedures and should submit them for approval (by the appropriate
area/sub-committee of the new Teaching and Learning Committee)."*[112]

The scheme operated in 1994/95: The following revised arrangements
for anonymous marking were approved by Senate on 29 November 1994:

*"Senate has agreed that the University should retain an anonymous
marking scheme for undergraduate examinations and that a common
central University scheme should continue in a modified form as
indicated below.*

*The scheme applies to Undergraduate written examination papers only.
The consideration of a policy on anonymous marking for other forms of
examination assessment will be considered after the operation of a
revised scheme for a year.*

*(a) A single personal 'examination candidate code' will be used for
each academic year. The code will be allocated by the Examinations
Officer and will compromise four digits which will be the unique
identifier for each candidate in the range of 1000 to 9999, together with
two digits indicating the examination year for the academic year
1994/95, candidates will be sent their personal 'examination candidate
code' at the beginning of the Easter term.*

*(b) Candidates should write their allotted code and description of the
examination on their main answer book and on all supplementary
answer books.*

*(c) Candidates should write their name and code on their attendance
card and should sign it.*

---

[112] University of Durham, Senate Minutes, 11.5.93, Minutes 400-403.

(d) Examination code reports will be produced showing the candidates'
(sic), the degree course and code, the reports will be sent to a
designated member of staff of each School/Department.

(e) In the event of a student mislaying his/her code:

>   *(i) **before the examinations:** he/she should apply in person to*
>   *the Examinations Section where, on production of suitable*
>   *identity he/she will be informed of his/her code.*

>   *(ii) **at the time of the examinations:** staff from the Examinations*
>   *Section who normally collect 'absentee cards' will carry a*
>   *full report of 'examination candidate codes'. Invigilators will*
>   *therefore be able to obtain the student code on behalf of any*
>   *student who has mislaid their code.*

(f) No candidate should be permitted to leave the examination room for
the 15 minutes prior to the end of the examination. This is necessary in
order to help invigilators collect and check examination scripts.

(g) At the end of each examination candidates must remain in their
seats until the examination scripts have been collected from each desk.
Invigilators must check that all answer books have a code written on
them, and also that the number of books collected from each desk
corresponds to the number written on the main book.

(h) Attendance cards will be retained in the Examinations Section but
will be made available upon request to the Chairman/Chairwoman of
the Boards of Examiners.

(i) Examiners should communicate their agreed marks to a designated
member of their Board of Examiners, who will have the responsibility
of transcribing them to a mark sheet bearing the examination candidate
codes and candidates' names. Because of the burden of work which this
will bring, Boards of Examiners should consider the nomination of
different individuals to take responsibility for mark sheets for different
year groups.

(j) Mark sheets bearing marks against candidates' names (which might contain marks provided by other forms of assessment) should remain confidential to the individual responsible for it until the last set of marks had been entered.

(k) When mark sheets are completed, Boards of Examiners should implement their procedures for the monitoring of borderline cases and the consideration of special cases. It is accepted that meetings of Boards of Examiners may be conducted with the identity of students known.

(l) Candidates who sit examinations in special rooms or who arrived late for examinations should be provided with answer books of the same colour and design as other candidates; special arrangements will be notified to the Chairman or Chairwoman of the Board of Examiners concerned. "

Three additional points were made by the Advisory Group on Examinations and Assessment which are worth mentioning because they indicate one or two implications involved in the modification of the 1993/94 anonymous marking system:

"(xiii) instructions to Examiners should be revised to take account of these new arrangements. Boards of Examiners should modify their statements of procedures and should submit them for approval by the appropriate sub-committee of the Teaching and Learning Committee.

(xiv) if a modified scheme was approved (this scheme was approved by Senate on 29 November 1994 as shown above), *resources should be provided urgently by the IT Services to ensure that an acceptable coding system was in operation for 1995. They would be required to liaise with the Examinations Officer in producing acceptable codes, and reports for informing students of their own personal 'student code', and in providing suitable examination mark sheets for departments.*

(xv) special arrangements should apply for the examinations which are to be held in January/March/April 1995, unless the modified anonymous marking scheme can be finalised in time."[113]

The differences between the schemes of anonymous marking in 1993/94 and 1994/95

The main difference between the anonymous marking systems operated at the University of Durham in 1993/94 and 1994/95 is that in the former year the candidates' names were incorporated on the examination answer booklets (sticky labels were used to hide the names), whilst the modified system for 1994/95 used only codes and no names. In 1993/94 the seals of the labels needed to be broken before any scripts were sent to the external examiner as a safety measure, so that a record could be kept of the marks awarded. Attendance cards were not used for this system. Each student would end up with a different anonymous number (pre-printed on the answer booklet) for each course of study. The candidates' numbers would be completely random and would depend on which answer booklet a candidate happened to use.

By contrast, in 1994/95 the sticky label system was replaced by a system which used personal examination codes. Before each examination period, all students would be allocated a code for each year of their study. This code would apply to all the courses of the candidate in that year. As a safeguard, lists were kept at departments and at the Examinations Section of the names and codes of candidates. Candidates were also required to fill out attendance cards, with their names and codes as a further

---

[113] Taken from an 'Extract from the Report to the [Senate] Teaching and Learning Committee from the Advisory Group on Examinations and Assessment Meetings held on 21 and 28 October 1994', University of Durham.

precaution, which could be referred to in the event of problems occuring. Because the codes for 1994/95 were not completely random, in the sense that a record existed of every candidate's name against his or her code, mark sheets could be produced in name and code order and given to a designated member of each department.

Perhaps the only feature missing from both the systems described here is any mention of the implications that anonymous marking might have for the University's appeals system. Students may be likely to appeal if, like 'A' levels, there is no easy and quick means by which their teachers can be sure the marks are correctly attributed. For this reason, the nature of appeals, irrespective of whether the number of appeals increases or decreases, will change with anonymous marking. As a result it would be in the best interests of the University to produce an efficient appeals system which addresses the particular demands of anonymous marking.

*   *   *

Attitudes towards the adopted system of anonymous marking

The material in this section is based on formal and informal interviews, together with correspondence and the responses of academic staff to the requests of the Academic Registrar for comments contributing to a review of the system after its first year of operation. The material presented in this section was selected for the interesting and relevant points it makes about anonymous marking, as well as to give an indication of the richness of the debate at Durham on the issue.

Mr Alan Heesom, former Secretary to the Durham Association of University Teachers (DAUT): (For Heesom's views on the Durham system see the comments of the History Department below).

Anonymous marking was not a matter specifically considered by DAUT. However, when asked by Durham Students Union whether DAUT would lend their support to the issue when DSU took the matter to Senate, DAUT agreed. At this time the AUT had no national policy document in favour of anonymous marking; therefore, the decision was purely Durham based. It should be noted that DAUT is not represented in its own right on any of the university committees, as might be the case at other universities where some unions are represented as ex-efficio members (the reason for this is thought to be to do with Durham being a collegiate university). Committee members may be, and often are, members of a union like DAUT, but they are not elected on to the committee solely because of this.

Heesom favoured the adoption of a system of anonymous marking in which the names of the students were not declared until the class of degree had been fixed. The system that has been implemented across Durham University allowed anonymity to be lifted at the meeting of the Board of Examiners (a stage earlier than Heesom ideally would have liked). For reasons of practical politics DSU did not press for the more anonymous system because they wanted to increase the likelihood of Senate endorsing the principle. At the time, Heesom remembered DSU stating that they would press for the more anonymous system as a next step. This has not proved to be the case.

Heesom, in his capacity as a DAUT representative on AUT Council, was present at Council in December 1994 when the issue of anonymous marking was raised and a policy statement supporting the principle was approved. The research used in the discussion leading to the AUT's position, which favoured the widespread introduction of anonymous marking, already existed in the public domain.[114] Little research of its own was conducted on the issue; the AUT had sent questionaires to their local branches some one or two years previous to the AUT's policy commitment, asking them whether a system of anonymous marking was in operation.

Dr Rosemary Stevenson, former President of DAUT: Whilst Stevenson held office as President of DAUT, the Association of University Teachers were concerned generally with the issue of teacher workloads, but had not, to her knowledge, specifically addressed the subject of anonymous marking. Much criticism of anonymous marking has focused on the extra workload it has created in some cases.

Association of University Teachers (AUT): (For the full contents of the AUT's policy statement on anonymous marking see Appendix II. For a critical examination of the contents of the document see Chapter Two).

AUT Council endorsed a policy statement supporting the practice of anonymous marking in higher education in December 1994. The Women's Committee of the AUT initiated discussion on anonymous marking in 1991, or thereabouts, and a formal position was taken in favour of the

---

[114] For a list of the research used by the AUT see Appendix IV. Virtually all of these sources have been critically analysed in Chapter Two.

principle following AUT discussions with the National Union of Students in 1994. It is interesting to note that Catherine Belsey, now Professor of English at the University of Wales College of Cardiff, was a member of the Women's Committee of the AUT which raised the issue for discussion. As the last Chapter indicated, Belsey conducted her own study of marking bias at Cardiff concluding discrimination against women.

Paul Cottrell, AUT Assistant General Secretary, indicated that there was a general feeling within the AUT that anonymous marking was increasingly becoming a feature of higher education, and that the "growing acceptance of its desirability within the academic community" should be supported by the AUT.[115] This is a good example of the influence that the social and political climate of opinion can have on decisions affecting policy.

Durham Students' Union (DSU): The pressure for change in Durham came from DSU. On June 8, 1989 DSU made a policy commitment to press for the implementation of a system of anonymous examination papers, on the grounds that "the present system is incompatible with an equal opportunities policy". A paper entitled 'Anonymous Exam Papers: the arguments' was submitted to the Joint Committee of the University and DSU early in the academic year 1989/90.

It should be noted that certain senior figures in DSU were not in favour of a system of anonymous marking for Durham, and because of this a fair amount of political compromising behind the scenes was necessary in order to present a united front. The then President of DSU, Dan Redford, objected to anonymous marking on three counts (see below for more

---

[115] Correspondence with Paul Cottrell, Assistant General Secretary of the AUT, 3 April 1995.

detail), but Redford had his own political agenda he wanted to pursue. In particular, Redford was keen to reform the student union network in the North East. Without the support of his Education and Welfare Officer, Redford would have had little chance of being successful with this project. Lara Fromings, the then Education and Welfare Officer, was strongly in favour of pressing for a system of anonymous examinations and said she had done a lot of research on the topic. The result of these political pressures saw the President backing his Education and Welfare Officer in DSU's campaign to get anonymous marking implemented in Durham.

Mr Dan Redford, former President of DSU: Redford objected to anonymous marking on three counts: first, he believed the introduction of such a system implied that lecturers were "not on a level", as he put it. In other words, the implication was that examiners were not to be trusted. Second, to set up a system of anonymous marking implied that the present system was not fair. Third and last, Redford was against the bureaucracy that he felt would need to be created in order to run the system. However, in his capacity as President of DSU, Redford supported the implementation of anonymous marking for the reasons that have already been stated.

Ms Lara Fromings, former Education and Welfare Officer of DSU: Fromings was Education and Welfare Officer in 1993 when DSU successfully managed to get Senate to approve the introduction of a system of anonymous examinations for Durham University. She was a member of the Working Party whose job it was to establish which of the systems of anonymous marking should be adopted by Durham. Fromings

was also present at Senate, along with the President of DSU, when the principle of anonymity and the chosen system were endorsed.

Anonymous marking was strongly supported by the Education and Welfare Officer and the Academic Affairs Officer (Mariel Bennett) within DSU. The issue did not originate from these two people because from 1989 it was already DSU policy to press for anonymous marking as part of the university's equal opportunities policy. The issue was taken up by Fromings who had a personal, as well as a union interest in the matter; she had been a student of the Psychology Department which already ran its own system of anonymous marking, thus in her own words she felt "aware of the benefits and of the literature" surrounding anonymous marking. Fromings, therefore, personally supported anonymous marking.

In her capacity as the Education and Welfare Officer, Fromings wrote to all the Heads of Departments asking them for their comments on anonymous marking and whether they were for or against the principle. Receiving approximately twenty replies from the thirty or so departments in Durham (a figure she claims was greater than that of the University when they tried to do much the same thing), Fromings responded to any criticism of anonymous marking in a paper supporting the principle. The research she used was largely that used by the National Union of Students, or that which could be drawn from the experience of other universities like, for example, the University College of North Wales, Bangor. The only *academic* research used was that which had been cited in the campaigning literature of the NUS, the weaknesses of which have already been discussed in the last chapter.

The major problem, perceived by Fromings, with anonymous marking was an administrative one. DSU believed there was enough support in the university for the principle of anonymity in theory; however, the reservations focused on how much a system would cost in practice and how it would be set up and operated. For this reason, DSU aimed to get the endorsement of the University on a simple mechanism of anonymous examinations then, once the system was in operation, DSU would press for more comprehensive anonymity and a more effective system. The sticky label system used in 1993/94 was judged to be the system with the smallest administrative burden, hence this was the system DSU were prepared to secure as a first step. However, many departments complained about the administrative burden of the sticky label system which was not a good indicator for future practice.

Dr John Hogan, Academic Registrar: The Registrar was dissatisfied with the anonymous examination system, as operated by Durham University in 1993/94, on two counts. First, there was criticism from various university departments that the system was too complex to administer. Second, there was concern that in practice the system was not particularly anonymous. In an attempt to solve these problems Dr Hogan, the Academic Registrar, set up a review: departments were invited to share their own views on the scheme of anonymous examinations, and to suggest how improvements could be made for future practice. The principle of anonymous marking was not under review, since it was perceived by the Academic Registrar that the majority of those involved with the system agreed with its underlying aim. This is not to say, however, that opposition to the principle of anonymous marking did not exist amongst academics and students.

Mr Joe Halpin, University Examinations Officer: The introduction of anonymous marking in 1993/94 made no real difference to the work of the Examinations Office, for although the system was centrally run the burden of the new work lay with the academic departments rather than with the central administration. Accordingly, no major problems were reported with the day-to-day running of the examination scheme from the point of view of the Examinations Office (the experience of individual departments is somewhat different and will be examined in the next passage).

The changes under the modified scheme of anonymous marking introduced for 1994/95 has resulted in greater responsibility and extra costs for the Examinations Office. With the introduction of individual candidate codes to replace the much criticised sticky label system, the Examinations Office, in collaboration with the University's Information Technology Service, has had to develop a workable student coding system. Each student has been provided with his or her personal code; a process which involved printing out each code in a wage slip format, tearing each slip and then sending it to the appropriate college for student collection. In addition, every department required a list of the names, codes and degree courses of students. The Examinations Section had to compile and send these lists which were important because the lists acted as an extra safeguard to ensure that examination scripts could be easily identified, and so that the marks of students who were joint and combined honours could be passed without delay between departments. All this amounted to a substantial amount of extra work, time and cost for the Examinations Office in preparation for the University's 1994/95 examinations. Whether or not the increased burden was merely the result of setting up the new system will have to be seen. It is possible that once

the coded system of examinations is fully established it will be no more costly in terms of resources than previous examination schemes have been. However, it must also be understood that no account has been taken of factors like the time that might be taken by the Examinations Office answering queries from students who have lost their codes, bearing in mind that only certain people, even in the Examinations Office, have access to the relevant information.

Selection of Departmental Examinations Officers (chosen from a list of contacts responding to the Academic Registrar):[116]

English: Prof. J. R. Watson (Chairman of Examiners), Mr J. S. McKinnell (Convenor of Examiners); **Geological Sciences** (ran their own system of anonymous marking four years previously): Dr C. T. Scrutton (Chairman of Board of Examiners); **Engineering:** Dr M. J. Holgate (Chairman of Board of Examiners), Dr John Wilson (Examinations Secretary); **Geography** (ran their own system of anonymous marking even in 1993/94): Dr I. S. Evans (Chairman of Board of Examiners); **Law:** Jacqueline A. Priest (Chairwoman of Board of Studies); **History** (ran their own system of anonymous marking): Dr Howell J. Harris (Secretary to Board of Examiners), Don Ratcliffe (Chair of Examiners); **French:** Dr G. N. Bromiley (Chairman of Board of Studies in French and Board of Examiners in Modern Languages, Year 1); **Classics:** Prof. P. J. Rhodes (Chairman of Examiners); **Philosophy:** Christopher Long (Secretary to the Board of Studies); **Modern European Languages:** Chris Perriam (Secretary to the Board of the School of MEL and to the Final Board of Examiners in Modern Languages); **Mathematical Sciences:** D. H. Wilson; **Physics:** A. D. Martin (Chairman of Board of Examiners), Dr G. H. Cross (Secretary to Board of Examiners), Dr K. J. Orford (immediate past Chairman of Board of Examiners), Prof. David Bloor (Chairman of Board of Studies); **Psychology** (ran their own system of anonymous marking): Dr R. F. Drewett; **Archaeology:** Dr C. C.

---

[116] Chosen from written evidence which suggested that individuals had considered views on anonymous marking.

Haselgrove (Chairman of Board of Studies); **Centre for Middle Eastern and Islamic Studies:** Prof. T. C. Niblock (Director of CMEIS); **Politics:** Julia Stapleton (Secretary to Board of Examiners); **Sociology** (ran their own system of anon. marking): Prof. D. Chaney; **Faculty of Science:** Dr J. Anstee (Dean of Science).

Zoology Department-

Professor Ken Bowler: The Zoology Department, which no longer exists as a separate department (Zoology and Botany merged in 1988 to form the Department of Biological Sciences), operated its own system of anonymous marking in the early 1980s. Starting around 1983, the system ran for approximately two/three years. The reason behind the introduction of anonymous marking was a fear in the department that a number of female students were not receiving fair treatment from certain internal examiners. The details of this situation are not important, suffice to say that anonymous marking was introduced as an extra safeguard, on top of the practice of double marking blind, in an attempt to counteract any potential problem of the sort mentioned above.

The reason for the discontinuation of anonymous marking in the Zoology Department was because it was judged to have been ineffective in its objective to secure fairness. The department continued to double mark blind and only reintroduced marking anonymously alongside other departments in 1993/94 (by this time the Zoology Department had been incorporated into the Department of Biological Sciences). According to Bowler, there were two reasons for why the system had proved ineffective. First, there were not many students in the department (Bowler gave an approximate figure of twenty-nine or thirty students) which meant that most classes had on average twelve students. As a result, the

handwriting or style of the work of most of the students was recognisable to the examiners. The system was, therefore, not particularly anonymous. The second reason was that some students deliberately "sabotaged" the marking system by writing their name at the top of each page of their examination script, which again undermined the principle of anonymity. There was also concern within the department that the codes used to hide the names of students increased the likelihood of errors. Subsequently, it was felt that because teachers within Zoology did not hold strong views in favour of continuing to mark anonymously, the system was deemed to be unnecessary and was subsequently scrapped.

Bowler was not convinced that the system of anonymous examinations in 1993/94 made any significant difference to marking in his own department, since the number of students was now so big that most students were effectively anonymous anyway. From his own experience, Bowler was of the opinion that a tutor could not hope to know the majority of the students, their work or their handwriting in a class. He saw the introduction of university-wide anonymous marking as little more than an attempt by the university to be seen to be fair and to provide a visible safeguard for outside observers in this respect.

Dr John Horton, Secretary to the Board of Studies in Zoology when the department's own system of anonymous marking was introduced: Horton, like Bowler, referred to the experience of the Zoology Department in which there were suspicions of favouritism on the part of examiners, but no bias was proven. However, anonymous marking was introduced because the department had decided that it would be the fairest thing to do. The system was tried for a few years, found to be okay and just, but, according to Horton, it was deemed to be not absolutely

necessary. Questions concerning whether the system was worth the "hassle", as Horton put it, were especially pertinent in light of the fact that a high proportion of the handwriting of students was easily recognised in such a small department.

Horton described the university-wide system of anonymous marking as another example of the red tape that universities are now affected by. He accepted the system and was not averse to the ideals behind it, but felt uncomfortable with the impersonal barrier it imposed between the teacher and the student. Generally speaking, Horton was happy with the way the system worked in 1993/94, although he pointed out that under the 1994/95 system the handwriting of a student could still be recognised, thus a teacher could still favour a student if he or she wanted to.

Dr Charles Shaw, Senior Lecturer in the Department of Biological Sciences and a member of Senate when anonymous marking was approved in 1993: Shaw was one of four people sitting on Senate who voted against the introduction of anonymous marking. In a letter written to the Chairman of the Board of Examiners in Biological Sciences, Shaw gave three reasons why he was not in favour of anonymous examinations: first, such a system was time consuming and confusing; second, it was ineffective; and third, it was unfair. Drawing on the content of this letter and comments made during interview the observations of Shaw will be considered further.

Shaw drew on his own experience of the examinations in the Biological Sciences to highlight how anonymous marking was a burden on the time of students, examination invigilators and examiners. Students of Biology are required to fill in four separate examination answer booklets in order

that there will be one answer per booklet. This differs from most other subjects which allow more than one answer to each booklet. The implications for this, are that more time was needed for the students at the start of the examination to fill in the necessary details, whilst more time was needed for the invigilators to lay out four booklets, question papers and sticky labels. At the end of the examination, additional time was needed to collect and check the scripts of the students who had to remain seated until this task was completed.

The extra tasks and time incurred by the new system in the above respects do not seem issues of major concern, but Shaw also complained that the instructions to invigilators were poor. This confounded the problem because invigilators were not being instructed to turn up early enough to set up the examination and to lay out the four, rather than the usual one, answer booklets. Similarly, the instructions to students about how to fill in the booklets and the attendance cards, with student codes on the former, and codes and names on the latter (this included how to use the sticky label and supplementary examination books and the need for students to remain seated until everybody's script had been checked and collected), were vague. As a result, Shaw found himself having to be at the examination far ahead of time (one hour before each S examination) to ensure that everything got done correctly.

Shaw illustrated the extra time burden on examiners by estimating the time it took him to process the scripts for the Biological Sciences: "I had 2000 scripts to process, which alone took a whole week, because opening the flaps doubled or even trebled the time taken". The opening of the flaps here refers to the sticky labels that had to be slit in order to transfer the marks of the written examinations onto mark sheets, to combine them

with the marks for practicals and assessed essays. The process also involved resealing the scripts of final honours students for the benefit of the external examiner. The 1994/95 anonymous marking scheme overcame some of this burden by replacing the sticky label system, and by making the recommendation that one designated person should only be responsible for the scripts in any one academic year. This way, three or so people shared the work that one person had done prior to anonymous marking, thus reducing the burden on any one individual.

Anonymous marking was felt to be ineffective because in classes that are relatively small the handwriting of the students is recognisable by the course tutor and examiner. Equally, in classes that are large, most students will not be known to the examiner thus undermining the need for examinations that are anonymous. This was thought to be particularly so for first year courses which were large and which seriously limited the possibility of a teacher getting to know the new intake of students. Shaw considered it to be far fairer to mark openly without anonymity because this would ensure that all students were treated the same, rather than some being known to the examiner, for whatever reason, while others remained anonymous. The argument here being that, the whole point of introducing a practice like anonymous marking was so that students would be treated equally, rather than differentially which is the consequence outlined above.

The question as to how fair a system of anonymous marking is, was challenged further by Shaw's reference to the unequal treatment that minority groups like, for example, dyslexic or foreign students received through the marking process. Shaw made the point that no allowance was made for the unequal command of the English language that such groups

had compared to the majority of students. Examiners had no choice but to mark all examination scripts on the same basis, which inevitably meant that any disability, physical or otherwise, was likely to be a disadvantage at the marking stage, even though most universites make provision for extra time to be allowed for certain groups of students - e.g. dyslexic students. By the time the meeting of the Board of Examiners had taken place, the flexibility that existed when students could be identified and their special circumstances known, was no longer there, because by this stage the marks were effectively in "tablets of stone". In this way, Shaw believed that borderline candidates would suffer: "at the examiners meetings it is impossible for individual examiners to ascribe marks to a particular candidate, and suggest they might be more generous".[117] This is apparent because examiners at the Board meeting would not know which scripts they had marked related to which candidates.

Shaw saw no need for anonymous marking because of his conviction that his colleagues, and academics in general, were honourable people who were not seeking to do students down. He also felt that anonymity would serve no purpose in the Sciences where students were being assessed on their ability to grasp the facts rather than their ability to present a critical argument, as would be the case in the Arts or the Social Sciences. For this reason, the scope for personal prejudices in marking was far less of a concern in the Sciences because a student would be examined on things that were, to a large degree, either right or wrong.

Although Shaw was opposed to the introduction of anonymous marking he accepted that the system was here to stay. He has turned his attention

---

[117] Letter about anonymous marking written by Charles H. Shaw to Dr D. Hyde, Chairman, Board of Examiners, 3 March 1995.

to looking for ways in which assessment practice could be improved to ensure that the most workable system was attained. Shaw was of the view that the system adopted for 1993/94 was probably the least time consuming for examiners. From the viewpoint of his involvement with the Biological Sciences' Board of Examiners, Shaw believed that many of the practical difficulties of 1993/94 were due to the scheme's rushed implementation. He was not convinced that the modified system for 1994/95 would prove any easier for examiners or students. He was of the view that the 1994/95 code system carried its own operational problems in terms of organisation and responsibility. For example, it was not clear whether it would be better for the bulk of the responsibility for the operation of anonymous marking to lie with departments or the Examinations Office, and which should have responsibilty for what.

One recommendation Shaw made was that the examination booklets should be redesigned. Shaw argued that there was too much rubric on the front of the booklets, the large part of which was rarely read by nervous students. More importantly, the content of the rubric was misleading in 1993/94: as stated already, students in the Biological Sciences were required to write one examination answer per examination booklet (this made it easier to divide the examination scripts up for marking); however, the rubric on the 1993/94 booklets explicitly demanded that students should make sure that no pages were left blank, hence students received contradictory instructions. As a means of safeguarding against confusion in the examination room Shaw suggested that a designated member of staff should be in the examination room for the first twenty minutes to settle any problems swiftly. In addition to this, examination invigilators should be given better instructions about how to work the new system.

History Department-

Mr Alan Heesom: Heesom's views as a former Secretary to the DAUT were reported earlier. Here his attitude to the system operated by Durham University is considered in his capacity as a senior member of the History Department. The History Department had been running its own system of anonymous marking for approximately five years before the university-wide system was introduced.

Heesom felt that the system operated by the History Department was "infinitely better" than the university system, and was a little aggrieved that the university hadn't consulted far enough with the departments already running their own schemes. History used one code name per person for all examination papers (this was the practice employed by the university in 1994/95; for further details of the modified system refer to earlier text). The History Department differed from other departments in that they only had third year examinations, whereas most other departments ran second and third year examinations, both of which contributed to the final degree mark. Four letter codes (which were in fact Anglo Saxon words) were used as personal identifiers. It was felt that such codes were easier to remember than nonsense words or the codes used by the University for 1994/95.

Heesom's criticisms of the university system were that it was too complicated and too bureaucratic. Difficulties were found identifying and locating students when there were problems. This was due to the complicated and intricate process involved in matching the names of students to their examination paper numbers. It was also difficult to be sure that, as an examiner, one had the right number of scripts to mark.

The university system was thought to have increased the workload for the Chair of Examiners in History which was further complicated by the need to use a computer to store information, the operation of which the individual concerned was unfamiliar with.

As courses go modular and continually assessed essays become more commonplace at Durham, it was felt by Heesom that anonymous marking would create big problems by making it very difficult for a tutor to talk with students about their work and to give them decent feedback.

Geography Department: The view of the Board of Examiners in Geography was that anonymous marking should be run as a devolved system in which each department would devise its own anonymous codes. This would allow Geography, and other departments which already operated anonymous examinations, to continue to do so without creating extra work. Those departments which were new to anonymous marking should be encouraged to adopt one code per student, as opposed to each student having a separate code for each year of his or her study (the revised arrangement operated in 1994/95), so that the workload and potential for error would be kept to a minimum. The issue of whether the anonymity of the system would be undermined by having a single code per student, particularly in a small department where staff room discussions of a particular candidate's code and performance might influence the marking of his or her future scripts, was not addressed by the department.[118]

---

[118] Taken from a letter written by Dr I. S. Evans, Chairman of the Board of Examiners in Geography, which was written to Dr John Hogan, Academic Registrar, in response to a request for the Department of Geography's comments on the system of anonymous examinations operated in 1993/94. Letter dated 29 September 1994.

Music Department: It was estimated that less than 20 % of the work done in the Music Department could be marked anonymously. For example, singing or conducting (performance work) would be impossible to assess on an anonymous basis. Clearly, therefore, anonymous marking is irrelevant in some subject areas.

Mr Peter Bantock, Residential Caretaking Supervisor: The examination system affected the work of caretakers at Durham University in various ways: examination rooms needed to be prepared to ensure that desks had sufficient space between them; materials like, for example, examination booklets needed to be transported from the Examinations Office to the rooms; and each room's lighting and ventilation needed to be checked. However, the 1993/94 and 1994/95 anonymous examinations, as compared with the former examination system, made no real difference to the caretaker's job. The same duties as existed under the old system were carried out with the same resources. No extra responsibilities were incurred and no changes to usual caretaking practice were noticed. It was admitted that with anonymous examinations there was more mess in the examination rooms, particularly in 1993/94, due largely to the students discarding on the floor the backs of the sticky labels used to hide their names on their examination scripts (sticky labels were not used before anonymity was introduced). However, this mess was not sufficient to disrupt the usual over night cleaning duties (chewing gum was thought to be more of a problem). With the same staff and at the same expense, the rooms were cleared and the system operated with no undue problems.

Cleaners of Dunelm (a major examination room): The scheme of anonymous examinations for 1993/94 had the unanticipated consequence of causing problems for cleaning staff. Unlike other examination sites at

Durham which have lino floors, Dunelm House has carpeted rooms. This caused a problem in that the peel-off sticky labels, used to anonymise the examination scripts, could be swept easily from lino but not from carpet, thus a lot of extra work for the cleaning staff was created. The 1994/95 system avoided this problem because it no longer employed sticky labels.

Students: In 1993/94, half to two thirds of History students made sure their scripts were not anonymous. The students placed the sticky labels in such a way that their names could be read, thus consciously or unconciously undermining the system. It is highly unlikely that the History Department was the only department to experience this situation. One of the reasons for the withdrawal of the Zoology Department's own scheme of anonymous marking in the mid 1980s was because students deliberately sabotaged the system by writing their names on their examination scripts. Also, the Working Party set up to look into anonymous marking recorded that there was not unanimous support for anonymity amongst students. The views of students were heard through the Boards of Studies. It was recorded in Senate minutes that many students had expressed the view that they "did not like to feel anonymous".[119]

Colleges: As a result of student codes being sent to each College for student collection (the codes could not be sent to each individual student because the Examinations Office could not be sure that the private addresses they had for students were up to date and correct), more work was created for the Colleges.

*   *   *

---

[119] University of Durham, Senate Minute 694 (b), 13.5.86.

The difficulty in assessing the impact of anonymous marking on degree results

As the previous Chapter demonstrated, a number of studies have been conducted comparing the degree results before and after the introduction of anonymous marking at a particular institution or in a particular department. Little can be gained from a similar analysis of the results at the University of Durham because the anonymous system has only been running for two years across the whole university; therefore, it would be extremely difficult, and somewhat naive, to attempt to assess what, if any, impact anonymous marking had had on examination results. A sustained effect might only show itself over a period of a number of years. Similarly, a longer term view would enable the effect of other factors which might affect the pattern of results like, for example, the introduction of modular degrees to be isolated. It should be noted that only very limited conclusions can be inferred from comparative analyses of this kind anyway. The students, courses and assessment methods will change from year to year; hence, any analysis of the degree performance of men and women which tries to get at the effect of anonymous marking, needs to recognise that other factors will be present from year to year.

*   *   *

Analysis of the costs incurred by the system of anonymous examinations

This analysis is not a valid statistical study, rather it is a series of cost estimates designed to give some indication of the extra tasks and costs incurred by the new system of anonymous examinations operated by

Durham University from 1993/94. The following are, therefore, statistical examples; the figures, where given, are no more than that.

In 1993/94, the Secretaries of the Boards of Studies were given the responsibility of breaking the seal of the sticky labels used to hide the names of students on their examination papers, in order that a record could be made of the candidates' names against their examination results, before any examination scripts were sent to the external examiners. This was a task in 1993/94, but not in 1994/95 under the modified system of anonymous marking.

There are thirty-one departments in Durham and, accordingly, thirty-one Secretaries to the Boards of Studies. Each Secretary took an average of 1.5 hours to complete slitting all the sticky labels. One can estimate that each Secretary to the Board (on Academic-related scales) costs the University not less than £26,000 per year. Based on this information it is possible to make a calculation of the minimum cost incurred by this added task in 1993/94.

The calculation is as follows:

Earnings per hour of Secretaries (assuming a 40 hr. week)		*Total hours of slitting scripts (31 x 1.5)*		*Extra cost incurred*
£12-50	X	46.5	=	£581-25

It is difficult to estimate the time and cost spent on reviewing the anonymous system operated in 1993/94. However, there were

unanticipated tasks involved in reviewing the system after only one year instead of the agreed three years. For example, letters from the central administration were written and sent out to every department asking for their comments; Boards of Secretaries had to reply to these letters, ideally after consulting with their own Examination Boards; these replies had then to be read and processed and a report compiled by the Academic Registrar.

Following the review, a modified scheme of anonymous marking was recommended which involved using only anonymous codes and not students' names on examination scripts. To make this possible, the Information Technology Service was instructed to set up an acceptable coding system. It was estimated by the Academic Registrar that it had taken one computer analyst two week's work to devise a suitable programme to generate and store the personal anonymous codes which were used by the students on their examination scripts in place of their names in 1994/95. Based on the University's salary scale one can assume that the computer analyst was costing the University something in the region of £26,000 per annum. Therefore, the cost of setting up an effective code system was around £1,000. This was paid for out of the IT Service's budget. Assuming that the code system encountered no problems and continued to run smoothly, there should be no additional costs on top of those involved in setting up the system.

The code system, unlike previous assessment systems, carried with it additional tasks for the central administration. For example, the student codes, which were produced in a wage slip format, had to be printed, collated, folded and sent to students in their colleges (this effectively amounted to approximately 6,000 individual letters). In a similar fashion

to 1993/94, coded mark sheets were sent to all departments; however, in 1994/95 two different versions of the mark sheets were produced and sent in an attempt to make the process of transcribing the examination results easier and quicker - e.g. the codes were produced both in numerical and course order. A designated member of each department received a copy of each student's code against his or her name as a safeguard. The Examinations Office printed all this information in different versions - i.e. numerical code order, alphabetical name order and lists compiled in order of department. With the modified arrangements for 1994/95, there was also the unknown time taken to deal with those students who had not received their code or who had misplaced it. These were all tasks that involved additional time and energy on the part of the university's administrative staff.

Various other costs, most of which are very difficult to put a figure on, resulted from exercises like the total amount of University time that was spent in relation to Committee meetings, setting up the system and reviewing it, and the investigations of the Working Party. There were changes and expense involved in the re-design of the examination answer booklets to secure the anonymity of candidates. This change was inevitable, and to be expected, once the introduction of anonymous marking had been agreed upon; however, what was not to be expected was that the examination answer booklets would need re-designing again for 1994/95 following the review of the 1993/94 system. Equally, the Boards of Examiners were required to modify their statements of procedures for 1993/94 and 1994/95 and submit them for approval to a sub-committee of the Senate Teaching and Learning Committee.

It is also important to be aware of the sorts of costs that often get overlooked with a policy change such as the introduction of anonymous marking. One must account for hidden costs like those resulting from the lost revenue of the University through losing a week to the conference trade by extending its final term, in order that the examination period, which includes the time needed for teachers to mark and process the results of students, can be completed (the term may have had to be extended anyway so that the examination period could accomodate the increasing student numbers). The added time for administrative duties experienced by many members of academic staff had some members of staff complaining that costs were arising in terms of the lost opportunities they had for research.

<p align="center">*   *   *</p>

The criteria used in the evaluation and definition of a preferred scheme of anonymous marking

> *- It should be effective, 'foolproof ', easy to administer and it should not give rise to extra administrative burdens of any magnitude.*

> *- It should satisfy the need to see that justice is done and that bias, however unintentional, is eliminated as far as possible from the marking of scripts; it should ensure that when a script is marked, only what is on the paper counts.*

> *- It should allow for the proper monitoring of borderline cases and for the proper consideration of special cases.*

The criteria above were used in the Report of the Working Party of Deputy Deans and the Examinations Office to help choose which system

of anonymous marking would be most favourable to the University. The Regulations and Admissions Committee made recommendations for a preferred scheme based on this Report; these recommendations were subsequently approved by Senate.

The extent to which the scheme of anonymous marking, as operated in 1993/94, met the criteria agreed upon by the Working Party is highly debatable. The selection of attitudes already outlined in this Chapter raise a number of important points which question each criterion. Clearly, just because an individual questions the effectiveness of the scheme, one should not automatically assume that the criticism is reasonable or fair. To ensure that the following discussion is constructive, any criticism made will be substantiated with examples, where possible.

Plainly, the chosen scheme of anonymous examinations was intended to be effective. The ambiguities involved in 'effectiveness' as a concept have already been looked at in Chapter One. On a more practical level, the operation of anonymous marking was clearly not as effective as had been desired by the Working Party that made the original decision over which scheme to choose. The Academic Registrar made his own dissatisfaction, and that of others, very plain in a document produced as part of the review of the 1993/94 system in which he stated:

> *It is difficult to underestimate the amount of ill-feeling expressed by some Departments towards the current anonymous marking scheme. There seems to be a widespread view that the scheme adopted in 1993/94 was complex to administer and unacceptable. Part of the criticism stems from those Departments or individuals still opposed to the principle of anonymous marking but even allowing for this there seems*

to be a widespread view that there must be an easier way to make arrangements for anonymous marking.[120]

The passage clearly indicates that anonymous marking was found to be ineffective by many across the University. It may even be taken to show that the Academic Registrar might share that view. The particular wording of the passage was thought to be unhelpful by some members of the academic community to the extent that the criticisms of the scheme were described as a *view*. The point made by certain academics was that the operation of the anonymous marking system *showed* it to be complex to adminster and as such unacceptable. Seen in this way, it was not a 'view', rather it was fact with evidence to prove it. The passage indicates implicitly that the *principle* of anonymity was not up for discussion in the review of anonymous marking. Many of those who were opposed to the principle objected to this narrowing of the review on the basis that they remained unconvinced by the evidence of the need for anonymous marking in the first place.

For the scheme to have been effective one of its primary objectives was to secure the anonymity of the students under examination. However, a charge against anonymous marking in 1993/94 was that it failed in many cases to secure this anonymity. Leaving aside the fact that in many cases the handwriting or style of work (especially of final year students where the class sizes tended to be much smaller) was recognised by teachers, a number of departments complained that many students, whether by intention or by mistake, failed to hide their names on their examination scripts sufficiently well. The sticky label system adopted by the

---

[120] University of Durham document, 'Review of anonymous marking procedures', written by John Hogan, Academic Registrar, 4 October 1994.

University in 1993/94 required students to put their names on their examination scripts and then to mask their name with the aid of a sticky label. In the History Department it was estimated that the names of between half and two-thirds of all students doing final honours examinations were readable.[121] The President of DSU, who proposed anonymous marking in Senate is known to have ensured that his examination scripts in Politics were clearly known to be his.[122] Redford (former DSU President) had recently been elected Councillor in a London Borough and illustrated several of his answers in the British Government and Local Government examinations with relevant references to experiences in Harrow and in local government elections in 1994. The system introduced for 1994/95, which used only codes and not a combination of codes and names on examination scripts, went some way to eradicating this problem.

The total number of examination scripts which could be identified by the students' names, across all departments of the University in 1993/94, was unknown. This problem should not, therefore, be blown out of all proportion on the basis of the History Department's findings and a few examples of known cases. However, if the reason behind the system's lack of complete anonymity was the students themselves consciously sabotaging the system by writing their names on their examination scripts, then no scheme trying to secure anonymity will be effective. If this was the case, and students were voicing their opposition to the principle of anonymity by undermining it, then perhaps the issue of whether any

---

[121] This figure was quoted in correspondence between the Secretary to the Board of Examiners in the History Department (Dr Howell Harris) and the Academic Registrar, 7 July 1994.

[122] Interviews with Dan Redford, former President of DSU, and Professor Richard Chapman, Politics Department.

anonymous assessment system can be effective should be addressed, because one measure of an effective system of assessment is one which students and academic staff accept and with which they feel comfortable. For this to be so, the presentation of convincing evidence to justify why anonymity is necessary, that is that a problem exists and that anonymous marking can solve it, may be required.

The Working Party agreed that the system of anonymous examinations should be "easy to administer and it should not give rise to extra administrative burdens of any magnitude". Again, there were complaints from many departments and the Academic Registrar that the 1993/94 system was too complex to administer and that it resulted in an unacceptable increase in the workload. The Engineering Department, for example, estimated that anonymous marking had led to the equivalent of an extra one person's worth of work for a week and on this basis decided that the system as it stood could not be justified.[123] Much of the extra burden for departments was centred around the time it took to transcribe the numbers and marks of the students, as well as their names (this would be the responsibility of a nominated individual who would be the only person to have access to the names of students) to ensure security of marks for any scripts lost, for example in transit to the external examiner. This whole process involved slitting and resealing the sticky labels on the examination scripts in order to make a record of each student's name. The experience of one member of staff in the School of Biological Sciences estimated that processing the anonymous examination scripts doubled or even trebled the time it usually took. Similarly, the Chairwoman of the

---

[123] Taken from a letter written by the Chairman of the Board of Examiners in Engineering which was a reply to a request from the Academic Registrar for comments on anonymous marking procedures. Letter dated 14 September 1994.

Board of Studies in the Department of Law made the comment that it was her impression that "the work involved for the support staff in this Department this year exceeded the load generated when we were 'going it alone' and making all necessary arrangements ourselves".[124]

Clearly, the modified system in 1994/95 which used only codes and not sticky labels reduced some of the pressure on departments, although the transcribing process remained a lengthy and detailed process which required great care and accuracy. Furthermore, the increased possibilities for errors, caused by names no longer being on the examination scripts, had the potential to cause untold problems. Proposals to make the system for 1994/95 less of an administrative burden for departments included increasing the responsibilities of the central Examinations Office. In 1994/95, the central administration had control over the student codes; any queries about lost codes, for example, could be directed through them and departments. Also, departments were supplied with lists of the names, codes and degree courses of students, as well as name-less mark sheets with pre-printed codes, in numerical order (to make the process of transcribing marks as easy and as safe as possible), which helped shift some of the administrative burden. Attempts were made to distribute the work within departments more evenly across a number of individuals in order to spread the load. These steps went some way to addressing the dissatisfaction with the 1993/94 scheme. The worry remains that the sticky label system was originally chosen largely because it was believed to be the easiest system to administer; this did not bode well for any other

---

[124] Correspondence from Jacqueline A. Priest, Chairwoman of the Board of Studies in the Department of Law, to Dr John Hogan, Academic Registrar, which formed part of the review of the anonymous examination system as operated in 1993/94. Letter dated 7 September 1994.

anonymous marking scheme, including the one that was operated in 1994/95.

The point was made in the last Chapter that a system of anonymous marking does not eliminate the possibility of bias altogether. The criteria followed by the Working Party did not demand this; however, it did set as one of its aims that "bias, however unintentional, is eliminated as far as possible". It could be argued that in this respect the scheme adopted by Durham University met its objective. However, one might also take the wording of the criterion to assume that bias already existed across the university. No evidence of this is available, but it is worth noting that some staff have said the introduction of the anonymous marking system was a slur on their professional integrity. Perhaps the only way assessment could be made less prone to bias was if examinations were type-written, thus reducing the possible effect of gender identification, particularly, through a student's handwriting. Although the feasibility of something like this is not practical or desirable for Durham at the moment, and, in any case it would still be possible to identify scripts of students who wished to make their scripts identifiable, experiments at the Open University have been looking in to the use of electronicly submitted assignments. Thus in years to come this option may become viable.

Does any scheme of anonymous marking allow for the "proper monitoring of borderline cases" and the "proper consideration of special cases"? Clearly, one's interpretation and understanding of the word 'proper' will dictate the answer to these questions. It seems that anonymous marking does accomodate borderline and special circumstance candidates - borderlines can be identified once all the results of a candidate have been brought together against his/her code and special circumstances can be

considered once anonymity has been lifted - but that this happens only very late in the assessment process. This provision was found to be workable by most Boards of Examiners in 1993/94, but not ideal. For example, one Senior Lecturer who had served on Senate and who holds a position of responsibility on his Department's Board of Examiners, voiced concern that the examination system under anonymous marking was too inflexible to adequately deal with borderline cases. The meeting of his Board of Examiners was described as little more than a forum to "rubber stamp" cases. It was felt that because the identity of the students was not known until the final meeting of the examiners, it was difficult to identify potential borderline candidates early and impossible to ascribe marks to particular students in order to bump them over a class mark to a final degree classification the examiners felt would have been justified on the basis of their knowledge of the students' other work (mature students are sometimes quoted as examples of this).

In theory, there is still scope for discussing borderline and special cases at the final examiners meeting; however, in practice it is always much harder to make adjustments once the individual marks have been decided upon and circulated to members of the Board of Examiners who do not all know all the students. The Department of English Studies reported similar discomfort that special factors were, and would continue to be, considered only very late in the assessment process, stating that any changes in a candidate's grade would result in "some kind of artificial 'adjustment' of an agreed mark, to take into account previously unknown factors".[125] This criticism of the system was recognised by the Academic

---

[125] Correspondence from Professor J. R. Watson, Chairman of Examiners, and Mr J. S. McKinnell, Covenor of Examiners, in the Department of English Studies, to Dr John Hogan, Academic Registrar, which formed part of the review of the anonymous examination system as operated in 1993/94. Letter dated 20 September 1994.

Registrar in the review of its operation.[126] However, the early identification of borderline candidates was expected to be just as difficult in 1994/95 with the modified scheme because candidates had a separate code for each year of their study; therefore, it was only when the different sets of codes for the second and third year were matched to the candidate, which gave a full range of marks, that borderline cases became identifiable.

Overall, the scheme of anonymous marking implemented for 1993/94 failed to meet some of the more important criteria decided upon by the Working Party that chose the system. Whether the problems have been overcome with the modified arrangements is not clear. One problem which seems to remain, however, is that if academic examiners have not been convinced by the evidence that a problem exists, it is not surprising if many of them are unhappy about work-creating adminstrative arrangements that appear to be costly and inefficient.

This is not to say that in the case of Durham anonymous marking could or should be abandoned because it is less than perfect in meeting its intended criteria. The issue of anonymous marking is not as black and white as this. As was stressed at the beginning of this thesis, there is no such thing as a perfect assessment system. For this reason, not being able to attain the perfect system should not automatically be used as an argument against pursuing what it is possible to achieve. There are many at the University who support the good intentions behind anonymous marking but who demand considerable improvement in its operation.

---

[126] 'Review of anonymous marking procedures', Dr John Hogan, University of Durham Memorandum, 5 September 1994.

This Chapter has clearly illustrated the differences in opinion and argument that can exist in relation to anonymous marking. It has raised many important issues and has given an initial indication of the ways in which different interests act and interact to affect decisions of policy. At Durham there was a rich debate on the issue. It should be recognised that this level of argument and feeling may or may not exist at other higher education institutions.

CHAPTER FOUR: CONCLUSION

In this concluding Chapter it will be important to do two things. One must be clear about exactly what has been established in the three earlier Chapters. Here one will cover general issues relating to anonymous marking; specific issues relating to the University of Durham; and, in the concluding section, issues relating to the political activity involved in both, which in the case of Durham led ultimately to a policy change. Second, the lessons from the research need to be drawn together in order to understand their significance for future practice, to offer recommendations and, where necessary, to highlight areas which require further study. In setting about both of these tasks it is important to appreciate the relevance of political activity and political analysis.

Anonymous marking and the mission, goals and objectives of a university

It is clear from earlier Chapters that a number of highly significant issues have been surveyed. Chapter One illustrated the importance of understanding an organistion and its policy process in terms of its mission, goals and objectives. Being clear about what an organisation is all about from the different perspectives and interests that make it up provides an excellent framework by which a policy like anonymous marking can be evaluated. The two principal aims of most universities, Durham being no exception, are to attain the highest standards in teaching and research. It is to these goals that a university, seen as a political system, must utilise its resources efficiently and effectively. Integral to attaining these goals are quality considerations and the desire to ensure that its practices are equitable and fair to its academic staff and its

students. The question is, does anonymous marking fit in with the goals and objectives of a university, and more than this, does a policy of anonymous marking do what it sets out to do? If the answer to either of these questions is no, then there may be lessons to be learned about the University as a political system.

To address the first question, does a policy of anonymous marking fit comfortably with the goals and objectives of a university? There is a case to be argued for both sides here. Clearly, anonymous marking is an explicit statement of intent with respect to equal opportunities. Many student unions argue that it is a guarantee of fairness in marking and in this way it could be seen as contributing to furthering the highest standards in university teaching. One could also argue, based on evidence given in Appendix I, that in some higher education institutions the introduction of anonymous marking has not resulted in an increasing administrative burden after the scheme has been set up.[127] Therefore, in terms of university efficiency the policy is arguably justifiable. However, for teaching to be of the highest quality, the morale and voice of teachers themselves need to be taken into account, as also has the aspirations of students, to benefit as much as possible from their experience.

There is widespread doubt that standards in teaching and examining are raised by anonymous marking. The figures and research examined in Chapter Two bear little conclusive evidence that anonymous marking affects the pattern of degree results. Some academics believe the role of

---

[127] See also 'A summary report', Sunderland Students' Union, 1992. The report summises that any extra adminstration caused by anonymous marking is perceived by many universities to be a relatively minor problem. However, the report also states that the issue is of major concern to the universities planning to implement anonymous marking. No further details are given on either of these counts.

assessment in helping a student develop academically is handicapped by anonymous marking, because the scheme can prevent students from getting sufficient face to face feedback on some assessed pieces of work. Furthermore, it has been argued that it is by no means clear that anonymous marking is a guarantee of fairness in terms of providing a level playing field for all students. Some argue that certain groups of students are actually disadvantaged by the system, while others contend that the sorts of biases anonymous marking is designed to protect against (if they exist) are still possible. For example, the examination scripts of dyslexic and foreign students are marked against the scripts of students who have no such disabilities. Similarly, the influence of gender in the marking process is not eliminated by anonymous marking because the sex of students can often be identified through their handwriting. This would tend to suggest that, as long as handwriting is involved, procedures for anonymous marking can never achieve what they set out to do. This does not mean that attempts to make assessment more equitable should not be pursued and that efforts to take steps that fall short of a perfect assessment system are not worthwhile.

The issue of bias in relation to marking is considered elsewhere; however, one aspect worth mentioning here is the impact of the claim that examiner bias exists or might exist, and that it is adversely affecting the degree performance of women particularly, on the morale of some members of the academic community. Such a claim, made by many student unions and a number of scholars who have shown some interest in the subject, has caused much consternation to those who feel strongly that there has been little evidence of a scholarly nature to show that a problem of bias actually exists. Consequently, these claims, which have formed part of the campaigns advocating the widespread introduction of anonymous

marking, have been received by some university examiners as a slur on their professional integrity. This, alongside the fact that in many instances the introduction of a scheme of anonymous marking has resulted in a increase in the administrative workload of individual examiners, has apparently contributed to a loss of morale of some members of the academic community who are already feeling the pressure of increasing workloads and student:staff ratios, and the modularisation of degree courses. It goes without saying that the high morale of the teaching staff is a key to a successful university and to its attaining the highest standards in its teaching. Policies which undermine this morale are plainly not in the best interests of the university, and it is part of the job of university managers to ensure the stability of its institution and the welfare of its personnel, as well as facilitating the university's goals. It should be noted that there are also many teachers who are in favour of the practice of anonymous marking. Thus the overall impact of anonymous marking on staff morale is difficult to judge. What does seem evident is that whilst virtually all teachers agree with the ideals behind anonymous marking, the problem seems to be that in practice the system creates an additional, and arguably, an unnecessary workload, while not delivering all it is supposed to. This has been the experience of many departments at the University of Durham.

When one considers whether anonymous marking fits comfortably with the aims and objectives of a university, it is also important to evaluate whether introducing anonymous marking adds anything to the university's assessment practices beyond what already existed. The answer to this question will be highly subjective. Nevertheless, it is plain from Chapter One that the practice of double marking and external marking have always been regarded as guarantees of fairness and protection against

possible bias - e.g. favouritism, etc. As was said in the introduction, no method of assessment or marking is perfect, or can ever hope to be. However, each must be judged in the context of what its own purpose is, and how far it contributes to securing the objectives of the university and the criteria by which any system of assessment is evaluated. On both of these counts anonymous marking has serious shortfalls.

The main strength of anonymous marking lies in the ideals it represents. In her own study of anonymous marking at the University of Wales College of Cardiff, Nina Parry-Langdon finds an importance attached to the public value the system communicates. She states in her Report, "the introduction of 'marking by numbers' [the name given to the particular type of anonymous marking she is studying] is supported throughout the Faculty and is viewed as both a practical demonstration by the University of its Equal Opportunities Policy and extremely good for the image of the University as a whole".[128] For many people across the University of Durham the introduction of anonymous marking has been understood as part of the University's requirement to be seen to be fair. It is a move that is instantly recognisable as a statement of quality assurance and equal opportunities assurance. The value of anonymous marking perhaps lies more in its face validity than in its practical value, and in its intentions and its ideals, for there is a strong case to argue that if the community of a university is happy with itself then this is justification enough for the measures it adopts.

On the other hand, anonymous marking should also be assessed examining other aspects of fairness like, for example, the implications of

---

[128] Parry-Langdon, "'Marking by Numbers'", p. 12.

introducing costly procedures that have superficial validity, to rectify what are, as yet, unproven problems. As has been stated, some members of staff at Durham University have argued that the resources could have been better spent on other activities to advance the aims and objectives of the university.

There does seem to be a public/private dichotomy with anonymous marking. Certainly, in the case of the Durham University experience most staff and students now accept anonymous marking, are not averse to the ideals behind it and recognise that it would now be difficult to revert to previous practices. Publicly, the university has to be seen to be fair, but privately there is a great deal of dissatisfaction among people who are affected by the system, with the way it operates in practice, in terms of the administration it creates and the affect it has on staff-student contact and trust. There was also some discomfort felt with the impersonal barrier anonymous marking imposes between the teacher and the student. Thus, anonymous marking is a clear and visible statement of a university's intention to provide opportunities that are equal and open to all, but how far it goes beyond these good intentions is debatable.

The benefit to the management of the university is clear in terms of the safeguard anonymous marking provides in the area of litigation, such that a student who alleges bias in the marking of his or her examination script will find it almost impossible to prove the case. Whether this was one underlying factor which motivated a sea change in favour of proposals for anonymous marking at Durham, albeit under the guise of DSU/NUS pressures is not known.[129] However, the feelings of the English

---

[129] An indication of the University's current anxiety concerning possible litigation was exemplified in a memorandum about providing references for students, and present and former employees. The

Department encapsulate a view which may be widespread, and brings out the concern over litigation throughout the higher education sector that has been hinted at:

> *Not all members of our board were convinced of the need for anonymous marking, and we certainly do not accept that there was any improper personal bias in the way the old system operated in this department. However, it is certainly true that anonymous marking has become an important defence for examiners, in these litigious times, against unfounded allegations of bias or victimisation; and if the present system* [the system of anonymous examinations operated in 1993/94] *can be improved, it certainly should be.* [130]

It is important to appreciate that a student can still appeal for reasons other than bias. For example, a system of anonymous marking does lay itself open to an appeal that an examination result was entered against a wrong code, which would result in a mark being given to the wrong student. Hence, it is far from certain that anonymous marking would have the effect of reducing the number of appeal cases across the university because it might only shift the nature of the appeals. The criteria for students appealing against their results may simply change rather than be eliminated altogether. It is even possible that anonymous marking could

---

memorandum sent from the University's Director of Personnel to all academic and related staff was reacting to a recent House of Lords ruling in stating that "references must be fair and accurate to avoid claims for negligence (from either the employee or prospective employer)". Further than this, and somewhat surprisingly, the Director of Personnel advised that "unless the reference is likely to be satisfactory in all respects, it is sensible only to provide either the blandest of references or one which goes little further than merely confirming the period of employment and the nature of their job while at Durham". One could argue that the sort of reference the Personnel Office is advising University staff to write, in order that any grounds for litigation will be avoided, will be of little use to anyone; the prospective employer, the employee or the referee. See Appendix V for the full memorandum, dated 15 November, 1994.

[130] Correspondence from Professor J. R. Watson, Chairman of Examiners, and Mr J. S. McKinnell, Convenor of Examiners, in the Department of English Studies, to Dr John Hogan, Academic Registrar, which formed part of the review of the system of anonymous examinations operated in 1993/94. Letter dated 20 September 1994.

lead to an increase in the number of appeals because more students may be likely to appeal if (like 'A' levels) there is no easy and quick means by which their teachers can be sure the marks are correctly attributed to the right candidate. Before anonymous marking, the assessment system had the built-in safeguard of the informal check provided by the teacher knowing his or her students and the sort of examination mark that might be expected from their other work. Also, teachers well regarded by their students could be apprehended after examination results were published and this could resolve anxieties among students who might otherwise have appealed. Therefore, a prerequisite of anonymous marking may turn out to be the necessity for a university to produce an efficient appeals system. This is something that will need to be adressed by universities.[131]

Anonymous marking, degree results and bias

From Chapter Two it has been established that bias, in the context of this discussion, is a preference or prejudice imparted consciously or unconsciously by university examiners through the marking process on the students. All too often in this debate the use of the term bias has not been clarified. This can lead to confusion because there are different senses in which the term is used, as well as there being many different types of bias, a number of which have been identified in the course of this discussion - e.g. gender, racial, ideological, personal bias and bias in assessment. It has also become evident that anonymous marking has a limited effect in eliminating the possible influence of bias generally on the marking process. Even gender bias, which is considered by many as the

---

[131] It might be useful for universities introducing anonymous marking to monitor the number and nature of student appeals to see whether any marked difference is apparent, and to see how the appeals system can be developed to work most efficiently.

principal problem, potential or otherwise, could still affect the value judgements of examiners in spite of anonymous marking because there is evidence to suggest that handwriting often gives the gender of a student away. Clearly, anonymous marking reduces the possibility of gender and certain other types of bias, but by how much is questionable. As one senior academic commented, "if one really wanted to be biased then there are ways and means of doing so whatever procedures are in place." Whether or not anonymous marking does enough in this respect is a matter of debate. Whether it does enough, offset against the extra administration and cost that can be involved is a matter for each institution to evaluate. In the case of the University of Durham experience there is a case to argue that the introduction of anonymous examinations was not sufficiently justified, and it has not been evaluated against all costs and against its impact on the aims and objectives of the University.

Having reviewed the literature on sex bias in marking in Chapter Two it is clear that its influence on the results of degree examinations is as yet unproved. The problem has been that where research has claimed or assumed the existence of sex bias, others have accepted the validity of such claims, using them as evidence of the need to adopt the practice of anonymous marking. The Association of University Teachers and the National Union of Students are two influential organisations which have made policy commitments supporting and pressing for the widespread implementation of anonymous marking across higher education. Both of these organisations have drawn from the research findings of Catherine Belsey and Clare Bradley, to name but two scholars who have concluded that sex bias in marking is a factor in explaining the differing degree results of men and women. Chapter Two has demonstrated that there are enough questions and limitations which call into doubt the conclusions

drawn from these studies. Again, this is not to say that bias in marking does not exist, but only that the research conducted so far has not shown it to exist. The concern is that the conclusions of these studies are not being called into question in the circles that matter - i.e. in the institutions considering adopting anonymous marking - and because of this, those people who have neither the time nor the inclination to delve beyond these claims, will receive a wholly unbalanced picture on which future policy decisions will be made. Often, however, this is what emerges from studies of policy decisions.[132] This thesis is an attempt to provide some sort of balance to the overall picture.

It is also not clear whether the issue of bias has been sufficiently grasped by those groups advocating the adoption of anonymous marking. There seems to be some confusion over why anonymous marking should be introduced in universities. Is it, as some suggest, because there *is* a problem with bias in marking which anonymous marking is meant to counteract? If so, where is the evidence that such a problem exists? Again, reference can be made to the survey of research conducted in Chapter Two which found no conclusive proof of the existence of marking bias. Alternatively, is it believed that anonymous marking should be introduced because it would prevent the *possibility* of examiner bias in marking, regardless of whether such a problem currently exists or not? This is, on the one hand, a safer line of argument in that it does not carry with it the offensive overtones which call into question the academic integrity and professionalism of university teachers in a system which already seeks to minimise the influence of bias through double and

---

[132] See, for example, Willson, F. M. G., Administrators in Action, vol. I, Allen and Unwin, London, 1961; Rhodes, Gerald, Administrators in Action, vol. II, Allen and Unwin, London, 1965; Chapman, Richard A., Decision Making, Routledge and Kegan Paul, London, 1968.

external marking. Such an argument can also be presented as being in the best interests of all parties concerned as it is in everyone's interests to be as fair as possible.

On the other hand there has to be good evidence that the costs (including opportunity costs) of anonymous marking are amply justified and seen to be worth it, because the introduction of a new assessment system to address an issue such as bias, which may or may not be a problem, could be seen as a questionable activity which is using valuable university time and resources. Resouces which a number of senior academics at the University of Durham argue could be put to better use in the organisation's pursuit of its mission - 'the highest standards in teaching and research'.

Alternatively, the introduction of anonymous marking might be seen as part of a much wider problem of future university management, one that could conceivably lead to the abolition of external examiners. This well established safeguard is already coming under pressures of various sorts, and it is thought by some university teachers that it will become completely unworkable, in its present form, with increasing modularisation and the increasing use of continuous assessment, due to the resulting workload which will be spread throughout the academic year. Evidence to this affect is already noticeable at universities like Nottingham Trent where external examiners are only employed for final year courses under the University's modular system.[133] Could it be, therefore, that universities are introducing anonymous marking as an additional safeguard in light of the developments affecting the external

---

[133] Information from Professor Christine Bellamy of Nottingham Trent University.

examiner system? Whether or not this is the case, it is important to recognise that anonymous marking is no <u>substitute</u> for the function performed by external examiners. Whilst anonymous marking is designed to guarantee fairness in assessment, it cannot offer any guarantee that standards within, and across, universities are comparable. Further than this, anonymous marking has no consultative role; it cannot offer advice or feedback on teaching or courses. The two systems of marking do not, therefore, present an either-or option for assessment practice in universities.

Much of the debate surrounding anonymous marking and whether sex bias can be shown to exist arises out of the attempts of scholars to seek an explanation for the pattern of degree results of men and women. Men are gaining a higher proportion of first and third class degrees compared to women. Whether this is because of the environmental influence of sex stereotyping; the different expectations to which men and women are expected to conform; the nature of assessment and degree courses; the biological differences which have been thought by some to affect the cognitive ability of the sexes; sex bias in marking; or a combination of some or all of these, is open to debate. What is clear is that a closer analysis of the performance of men and women in particular subjects is required, because the performance differences are greater in some subject areas than others. Grouped together, as they so often are, the subjects can combine to produce a distorted picture of the overall pattern of degree results.

Policy and practice

Having examined anonymous marking from a philosophical and a socio-psychological perspective, the case study in Chapter Three provided a convenient and valuable opportunity to study, first, the practical operation of a system of anonymous marking and, second, the processes involved in the run up to and after the policy decision in the closed political system of the university. As a result, the case study will be useful to other higher education institutions, including Durham itself, because of the lessons that can be drawn from the unique Durham experience. Some of the points which will be made in this section will relate specifically to Durham University, whilst others will be intended to have a more general application.

Among the points that should be addressed are the timing of introducing a policy of anonymous marking; the operational implications of the system for the university, its staff and students; and, the attitudes and the interaction of the group and individual interests affected by anonymous marking.

(I) The timing of the introduction of anonymous marking

The sorts of pressures affecting mass higher education have been referred to many times already during the course of this thesis. It is important to understand that the introduction of a policy change does not occur in isolation from other pressures affecting the policy of an organisation.[134]

---

[134] Writing about the relationship between the civil service and society, Richard A. Chapman and J. R. Greenaway state:

There may never be an ideal time for a major change in policy; however, there clearly are better and worse times.

The timing of the implementation of anonymous marking in Durham has to be understood in the context of the transition that the University is undergoing with the introduction of modular degrees because the two developments have a significant bearing on each other. As with all universities, Durham is being required to go modular because it is thought that this will contribute to a more flexible and accessible higher education system. The implication for assessment practice at Durham is that it will become more complex. The reasons for this may have relevance outside of Durham also. Many departments across the University are being required to revise radically their teaching programmes to accomodate the new modular system. For example, Faculties which currently operate a four course structure (the practice of departments in the Social Sciences) are being required to fall into line with the six course structure demanded by modularisation. The impact on assessment is tremendous, especially as it is being introduced at a time when other changes are also being introduced.

In addition to the reorganisation that would anyway be necessary with modularisation and anonymous marking, each course module in Durham has the option of being assessed by final examination or by continually assessed essays or by a mixture of these procedures. This has repercussions for the whole assessment system. Forgetting anonymity for a moment, problems arise over the time-tabling of assessed pieces of

---

Public administration is not an activity which takes place in a vacuum. It exists in all structured societies and within all forms of government; it is part of the complex system of interrelationships between people and public institutions in the modern world.
The Dynamics of Administrative Reform, p. 224.

work because what often happens is that a student places all his or her time and effort into meeting deadlines for assessed work to the detriment of work required for modules assessed by different means or at other times. Inevitably, the courses being assessed by final examination are likely to suffer throughout the year. Students may choose options on the criterion of their assessment methods; and this, in turn, may attract certain sorts of students to particular courses with the potential to skew examination results.

The increased flexibility intended with the introduction of modular degrees will create additional problems where, in some cases, second year final honours students will be able to take first year courses. This will lead to the need for each Board of Examiners to identify and separate out the examination scripts of second year students from that of the first years. This may become a time consuming process since the introduction of anonymous marking will mean that students will have to be identified through their anonymous codes, unless the year of study is one piece of information displayed on the front of all examination scripts - and this, in turn, has potential for undermining anonymity in certain cases. The reason this must be done is so that the scripts that count towards a student's final degree classification (all scripts of second and third year students count in this way) are able to be double marked anonymously with some being sent to the external examiner, while those that do not count towards the final degree, can be marked pass or fail by one marker only (though it is generally accepted that scripts that are judged to have failed are always double marked). All this adds to the complexity of the assessment system.

With universities in general, and Durham University in particular, going modular and increasingly incorporating practices like continuous

assessment, the issue of anonymous marking is becoming more acute. The particular complication that anonymous marking adds in the case of Durham, is that only the courses that are assessed by written terminal examination will be marked anonymously; as yet, no provision has been made to apply anonymity to continuously assessed essays. Hence, a patchwork system of asssessment will be created with anonymity only in those modules which opt for an end of course written examination, which will effectively decrease anonymity. This patchwork system will cause all sorts of problems with amalgamating and administering the results of students, some of which will be anonymous and others not, and will result in university staff needing two different sets of rules and procedures to ensure that fairness and accuracy prevails. This has potential for chaos, or if not chaos, the consumption of considerable resources in terms of time and energy. Further than this, a situation could arise where every module offered by the University was continuously assessed; this is already known to be happening in some other universities. Hence, it is possible that anonymous marking would cease to be practised, in spite of the University endorsing a policy supporting the principle of anonymity for written examinations. This predicament, admittedly extreme though it is, would mean that anonymous marking since its introduction in 1993/94 would have been a complete waste of time and resources.

One solution to avoid this extreme situation would be for the University to decide to go totally anonymous in their assessment practices adopting a simple code system for continuously assessed essays and written examinations, like they do at the University of Sussex, or to abandon anonymous marking altogether. The perceived problems with marking continually assessed essays anonymously have been referred to in the interviews conducted for the case study. For example, difficulties have

been envisaged at Durham over providing the student with feedback on his or her work; in making sure that individual students have handed their work in on time, and if they haven't, that they receive the appropriate penalty for not doing so; and in cases where students seek advice from their course tutor on project or essay work which, in turn, undermines the system's anonymity. The quandary faced is that quality bodies and experts within the field of education and assessment are encouraging higher education institutions to make use of a variety of forms of assessment; however, the compatibility of this with the principle of anonymous marking is doubtful.

Clearly, therefore, timing is important. One lesson that can be drawn from this is that, for a university with no long standing record of anonymous marking, it is probably better to have modular courses, and how they are going to be assessed, in place and fully established before introducing anonymous marking. Similarly, anonymous marking, when introduced, should be made to apply as far as sensibly possible to avoid creating the potentially chaotic and complex system which could face Durham.

(II) The practical implications: administration, error, costs and anonymity

Drawing on the evidence of the experience of the University of Durham the practical implications of anonymous marking can be far reaching. The fact that the system has had to be substantially modified, following a review, after only one year of operation speaks for itself. The principal issue is the amount of work anonymous marking creates. The sorts of administrative difficulties that have been faced have been highlighted in Chapter Three by looking at the experiences of a selection of

departments. Among the additional tasks are the added time it takes to process and mark the examination scripts which, along with the increase in the number of students being examined in 1994/95, resulted in Durham having to extend its third term by one week; there was also difficulty in recording the full range of marks for each candidate because of the care that needed to be taken to correctly match the different codes to names, and difficulty in amalgamating the results of different types of assessments, some anonymous; some not. This is not to mention the administrative implications of the sticky label system used by Durham in 1993/94. Durham is not alone in finding that a system of anonymous assessment causes extra administration. Eight out of thirteen higher education institutions replying to a survey conducted by City University reported an increase in administrative duties.[135]

A number of practical recommendations have come out of the Durham experience. First, some departments felt that the year and degree subject should be left on the front of the examination answer booklets, even though there was a fear that this would make the system less anonymous.[136] The benefit to departments in doing this would be to help them identify early the examination scripts that needed to be passed onto other departments for marking. It would also save a lot of time over separating the final honours examination scripts, which must be double marked, from the preliminary honours scripts, which are only double marked in cases where a script is adjudged to have failed. There will be increasing scope for courses to have a mixture of preliminary and final

---

[135] See the City University Survey on blind marking (1989) in Appendix I.

[136] The Psychology, Law and Geography Departments, to name three, believed that the benefits of having the year and degree subject on the examination scripts outweighed the threat to anonymity that this caused.

honours students with the introduction of modularisation; therefore, an easy means of separating scripts will be all the more necessary to avoid unwanted administrative duties involving checking each candidate's course and year of study via their codes.

The Durham experience has also shown that the design of the examination booklets is important. The rubric on the booklets must be accurate for the course of study (the Biological Sciences experienced a conflict between the instructions on the booklets and the way they wanted to operate their examinations). With all systems of anonymous marking there needs to be a clear set of instructions for examination invigilators. Provisions for checking or advising students who are unsure of their anonymous codes need to be established. A practice which worked well in Durham involved a designated member of staff of the Examinations Office visiting each examination with a list of the names and codes of students, from which any candidate, already identified by the examination invigilators, could be told their personal code.

A potential problem with any anonymous marking system based on codes is that it is open to error which cannot easily be detected because of the anonymity of the system. A number of institutions in the University of Sunderland Students' Union survey[137] expressed concern over the mistakes that could occur when the anonymous codes are matched up with the names of candidates. Other potentially vulnerable areas in the marking process are when the marks of a candidate are transcribed onto the mark sheet, which is a task requiring great accuracy, to make sure the course marks get assigned to the correct codes and hence the right

---

[137] Sunderland Students' Union, 'A summary report', 1992.

candidates and, second, ensuring that a candidate has in fact used the correct code in the first place. The problem faced is that there is no longer the informal check of an examiner who, broadly speaking, knows what marks he or she has awarded to particular candidates. An element of trust is involved in assuming that the students have correctly remembered their codes. A sensible practice employed by the University of Durham for 1994/95 was to use attendance cards on which students would be required to write their names and codes; the assumption being that even if a student used a wrong code on his or her examination answer booklet, he or she would use the same code on the attendance card. Thus, the attendance cards could provide a check on identity if necessary. To minimise the risk of transcribing marks to the wrong anonymous code, the University of Durham produced mark lists in name and code order, for which a designated member of each department had responsibility. This list, held in each department and at the Examinations Office, was also used as a source from which students could find out their personal code in the event of misplacing it.

It has been suggested that a system of anonymous assessment using random numbering for the anonymous codes is more prone to mistakes because it is more difficult to cross check. To avoid using random numbers, some universities utilise the registration numbers or UCCA numbers of students for their anonymous codes. Even with systems that use codes which have a logic to them, and which use all the measures mentioned above to safeguard against possible mistakes, it is still possible for human error to occur. It is impossible to know how many errors, if any at all, go undetected with anonymous marking; nevertheless, such a system demands an efficient and effective appeals system, as mentioned previously, to deal with any suspicion of error.

The end of Chapter Three considered the sorts of costs involved in establishing a system of anonymous marking. Some of these costs will disappear once the system is up and running. The question is, do the benefits of the system justify the costs it incurs, and is the money and time spent being invested with a view to furthering the mission, goals and objectives of the university? Clearly, resources within higher education are at a premium and spending priorities need to be established in order to strive for teaching and research of the highest standards. Difficult funding issues will need to be faced within many areas of university life, like the libraries, for example, which have suffered as a direct result of the efficiency savings which have affected higher education. There are no easy solutions to the budget balancing questions posed above. Here, though, an understanding of the sorts of costs involved in introducing anonymity have been raised, even if only in a somewhat imprecise and illustrative way, to indicate how institutions might establish their own answers to these questions and accordingly their own priorities.

Earlier Chapters have shown that it is very difficult for any system of assessment to be completely anonymous for a variety of reasons. Smaller departments have problems ensuring total anonymity because the handwriting and style of many of the students are recognisable. The nature of certain degree subjects do not lend themselves particularly well to anonymity. Music, for example, bases a significant proportion of its work on performance related assessment. There was also evidence at Durham of some students not wanting to be anonymous and acting to undermine the system. It has also been argued that anonymous marking is less necessary in some areas than others. The Sciences are often used as an example to make this argument because Science subjects are more factually based, compared to essay writing subjects, and accordingly they

are more objectively marked. These are all implications that need to be thought through and addressed as part of the package of a system of anonymous marking.

(III) Interests and attitudes

The views considered in the case study of the Durham experience brought out a number of issues which are worth reiterating briefly for the lessons that can be learnt. Simple though it is, a point worth making is that widespread consultation with those staff and students who will be affected by anonymous marking is essential to establishing, first, whether anonymous marking is wanted, and, if it is, to make sure that the best system for the institution is chosen. In order that the best system is chosen, the consultation process should extend beyond the university in question, so that the experiences of other institutions, like Durham, can be learnt from, and where possible, 'best practice' can be identified and subsequently used. Consultation is important to maintaining the morale of university staff by involving all parties in the decision making process which should help to minimise the possibilities for unforeseen problems. It was felt by some at Durham that many of the problems experienced in the first year of university-wide anonymous marking could have been avoided with greater consideration and consultation. As W. K. C. Guthrie stated in his foreword of F. M. Cornford's satirical explanation of University politics in Microcosmographia Academica, "whereas Cornford's enemy was inertia ('There is only one argument for doing something; the rest are arguments for doing nothing'), we may reasonably hold today that the greatest peril to the things which we value (and he

valued) lies in too rapid change."[138] In other words, the implementation of a new system of anonymous marking should not be rushed.

During the course of this thesis, one has heard concern expressed over the impersonal nature of anonymous marking, and the problems it has been deemed to create for teachers in terms of student feedback, administration and the assessment of course essays. One has also heard about the ideals anonymous marking represents, the safeguards of fairness and equality the system aims to establish, and the confidence it imparts for some people. As was stressed at the outset, no system of assessment is perfect. Ultimately, however, anonymous marking can only be judged in terms of the aims and objectives of a university. Let it not be forgotten what a university is all about - its principal reason for being, is to achieve the highest possible standards in teaching and research. Everything a university does should seek to advance this mission. How far anonymous marking contributes to the advancement of this mission and the aims and objectives of a university is a question that has been open to debate in this thesis.

The relevance of political concepts and analysis

As well as giving the opportunity to focus on issues directly relevant to the operation of anonymous marking, the experience of the University of Durham could provide a practical example for the scholarly study of a system of 'closed politics',[139] in which a significant reform could be viewed through political activity in a higher education institution.

---

[138] Cornford, F. M., Microcosmographia Academica, ninth impression, foreword written by Guthrie, W. K. C., Bowes & Bowes, London, 1973.

The university system with all its sub-systems forms a complex, and often closed, political world in which political activity is an integral part. Senate and the various committees of the university are a system of government. University politics, as with all forms of politics, is characterised by the interaction of different interests in the perpetual and active process of agreement-seeking. Conflict is resolved through a system of authority or government which seeks stability for the University to achieve its aims and objectives. This situation is perfectly illustrated by the various definitons of the term 'political activity' which have been suggested by, among others, Bernard Crick[140], J. D. B. Miller[141], and Graeme Moodie and Rowland Eustace. A sense of this can be best gained by referring to the works of one of the above. Moodie's and Eustace's book, <u>Power and Authority in British Universities</u>, has been chosen because of its direct relevance to politics in a university environment:

> *To secure the agreement indispensable for action is not*
> *always easy since it may involve the exercise of various*
> *kinds of persuasion, argument, or pressure, including,*
> *even, the threat of force. This process of agreement-seeking,*
> *of trying to induce others to act together in a particular*

---

[139] The term 'closed politics' refers to the non-partisan, policial activity that occurs between interests within any organisation or closed system. Reference to the term, in the context of the closed politics of Whitehall, can be found in Chapman, Richard A., <u>Ethics in the British Civil Service</u>, Routledge, London, 1988, p. 311; and, Snow, C. P., <u>Corridors of Power</u>, Penguin Edition, Harmondsworth, 1966, p. 40. See also Davies, M. R. and Lewis, V. A., <u>Models of Political Systems</u>, Macmillan, London, 1971; Easton, D., <u>The Political System</u>, reprinted edition, Alfred A. Knopf, New York, 1959; and Wiseman, H. V., <u>Political Systems</u>, Routledge and Kegan Paul, London, 1966, for a discussion of the concept of a political system.

[140] Bernard Crick defines 'political activity' in the following way:
> *Politics, then, can be simply defined as the activity by which differing interests within a*
> *given unit of rule are conciliated by giving them a share in power in proportion to their*
> *importance to the welfare and the survival of the whole community. And, to complete*
> *the formal definition, a political system is that type of government where politics proves*
> *successful in ensuring reasonable stability and order.*
<u>In Defence of Politics</u>, 2nd edition, Penguin, Harmondsworth, 1982, p. 21.

[141]   *Political activity, then, arises out of disagreement, and it is concerned with the use of*
> *government to resolve conflict in the direction of change or in the prevention of change.*
> *It is about policy and position.*
Miller, J. D. B., <u>The Nature of Politics</u>, Gerald Duckworth & Co. Ltd., London, 1962, p. 16.

fashion, is what we regard as the activity of politics.[142]

From this passage, the existence and the importance of different interests and pressures emerges as the basic elements in University politics. A later passage in Moodie's and Eustace's book reinforces this point. Although the text is specifically talking about university rules, penalties and sanctions, exactly the same can be applied to the process of decision making that leads to the formulation of a new university policy.

> *Universities are political in the sense that in making, changing, and applying the rules, penalties, and sanctions they activate or mobilize important and conflicting perspectives and attitudes. The consequent disputes must be resolved, in the sense that decisions must be taken on the issues in contention without undermining or disrupting the coherence of the university itself. Moreover, one finds that at least some of the differences and disagreements in perspective and attitude are related, if not necessarily closely, to certain important underlying interests. In particular, at least some of the differences in perspective and attitude will be related to an individual's status, as a member of the academic staff, as a student, or as an administrator.*[143]

Clearly, interest and group analysis is an invaluable methodological approach by which any policy decision can be understood. The path breaking work of Arthur F. Bentley was one of the first to have appreciated this in his attempt to forge a group theory of politics.[144] Bentley's work is significant for the method of analysis he adopted. As Bentley himself stated with the sub-title to The Process of Government, his intention was to "attempt to fashion a tool"; the work is important for

---

[142] Moodie, Graeme C. and Eustace, Rowland, Power and Authority in British Universities, George Allen and Unwin, London, 1974, p. 15.

[143] Moodie and Eustace, Power and Authority, p. 18.

[144] Bentley, Arthur F., The Process of Government, (ed.) Peter H. Odegard, The John Harvard Library: The Belknap Press of Harvard University Press, Cambridge, Massachusetts, 1967.

its methodology as much as its content.[145] His understanding of the political process was gained through identifying and studying the behaviour and relationships of groups, because it was by analysing the interests involved that decisions affecting policy could best be understood.

Bentley understood a 'group' to be a 'mass of activity'. Every group has an interest. The claim of group theorists is that it is the group that gives meaning to the political behaviour of the individual, and it is the characteristic interactions between the individuals within the group that gives it its definition.[146] Here, 'interests' are considered to be the 'activity directions' or 'policy attitudes' of groups.[147] David Truman, another much referred to writer on the governmental process, points out the shared attitudes of groups can be potential as well as existing:

> *Although no group that makes claims upon other groups in the society will be without an interest or interests, it is possible to examine interests that are not at a particular point in time the basis of interactions among individuals, but that may become such.*[148]

---

[145] In his editorial of Bentley's The Process of Government, Peter H. Odegard remarks that the book is "not a substantive study of any particular social or political problem, and certainly not a comprehensive theory or system of political science". Further, it is "not designed to answer but to raise questions, and to indicate how we may go about finding answers which in any case will be tentative at best". p. XXIX.

[146] Geoffrey K. Roberts characterises group theory as an approach that considers "the group, rather than the individual or the state, is the basic unit of political activity, since individuals in politics act in group contexts and their behaviour is affected-some would say even determined-by group structures, norms, goals, etc." See the entry under 'Group basis of politics' in A Dictionary of Political Analysis, Longman, London, 1971.

[147] See 'Group basis of politics' in Roberts, A Dictionary of Political Analysis.

[148] Truman, David B., The Governmental Process, Alfred A. Knopf, New York, 1953, p. 34.

This group and interest approach can provide an invaluable insight into the interactions and relationships of those individuals involved in the policy decision which led to the introduction of anonymous marking at the University of Durham. One must appreciate however, that group theory has two serious limitations. First, in a practical setting it is not always a straightforward task identifying groups, potential or existing. In a similar vein, the theoretical definition of the group concept can easily be made so general as to be vacuous. The second limitation is that group theory leaves little or no room for the political behaviour of individuals.

One clear advantage of identifying and understanding the different groups or interests influencing the policy process is that it is possible to get behind the restrictions that arise if only the formal nature of the organisation, or the merits of the issue under discussion, are examined. The interaction within or between groups or individuals gives a picture of whose interests are actually being served by a decision, for the policy positions of interested parties are not always what they seem. A very good example of this is the united front that Durham Students' Union presented to the academic governing body of the University (Senate) in support of the issue of anonymous marking. However, behind the scenes there were major policy differences between the President of DSU who was opposed to the introduction of the principle of anonymity, and his Education and Welfare Office who backed the issue wholeheartedly. In spite of these differences agreement was achieved because the President had his own agenda he wanted to pursue and, therefore, political compromises could be struck. This is also an excellent illustration of how in any political system decisions can be made on information or criteria that are not always relevant to the matter in hand. Hidden agendas are not always easy to identify.

It is important to be aware of the motivations of certain groups which can help explain why their interests are as they are. DSU, for example, is a body formally organised, to which highly motivated and, more often than not, politically charged students obtain office through annual elections. The term of office tends to be for one year only. During this short time the members of the executive will want to feel that they have made their mark on university affairs. This inevitably will encourage individuals within DSU to take on campaigns which have a chance of succeeding in the short term (their election manifestos may have been built around such campaign pledges). An issue like anonymous marking fits nicely with such requirements, particularly as guidance booklets advising how best to mount a campaign pressing for anonymous marking are available on request from NUS headquarters. Thus, there may be more to the policy position of DSU than first appears, especially since it is known that some student opposition to anonymity has existed at Durham, as well as opposition within DSU over recent years. Hence, questions can be raised over whose interests are being served here.

The interplay of interests within an organisation like the AUT has been equally revealing. The influence of sub-groups has shown itself through the relationships between the women's section of the AUT, and individuals like Catherine Belsey, which have had what looks like a significant impact on the national policy position of the AUT.[149] It is also worth remembering the position of DAUT, as described in the case study, which leant its support to DSU on the issue of anonymous marking, thus taking a policy stance which at the time was in advance of the policy later adopted by the AUT.

---

[149] See pp. 65-71, 70-71, 89-90 and Appendix IV.

In this study of anonymous marking it has become apparent that various groups and individuals have produced policy documents and drafted papers based on, as yet, unproved assumptions. The most common assumption in this context has been that significant bias in marking has been shown to exist as a problem that needs addressing. As has been demonstrated in Chapter Two this is far from clear. One implication of this is that the drafting of papers and information relating to this issue has become a far from straightforward exercise. For example, some of the information reaching the public arena in the form of policy documents, or internal papers within higher education institutions, is in itself biased because of its own failure to recognise the flaws of the assumptions it is making. The AUT, NUS and DSU are all culprits of this by relying on research that has been shown to be inconclusive. The problem is that this is the information upon which policy decisions are being made. It may be sobering to reflect that this is so, and that pressure from these groups has been so effective in an academic institution that has as one of its aims excellence in research.

Another example of how the policy process can become a focus for political activity was evident in the conduct of the review of the anonymous system at the University of Durham. The Academic Registrar, who was responsible for the review, set out a definite policy position by not incorporating the *principle* of anonymity (as opposed to its detailed implementation) in the review. Clear indication was given that the principle of anonymous examinations was not under review because the arguments had already been had and won. This was stated during an interview with the author and can also be illustrated from the tone of review documents (see quote on pp. 114-115).[150] A firm position was

taken which was not widely recognised as such. There are many possible reasons for this position. The most likely is that it would not have been in the interests of pragmatic management to change back to the assessment practices which had operated before anonymous marking; certainly the stability and confidence of the university would have been affected. It might also be suggested that anonymous marking could give more control - within the University as a political system in the sense being used here - to one group rather than to another. More power could accrue to the central administration in the running of examinations and university affairs, as opposed to, for example, academics or departments.

It is clear, therefore, that the University of Durham presents a web of interacting interests, each of which exerts pressure on the policy process of the political system of the University. The combination of these pressures manifested in the policy decision to introduce anonymous marking. The interests, as has been highlighted, are not always as they seem. The internal politics and hidden agendas within groups make it very difficult to decipher whose interests are being pursued or served. In the case of Durham University the pressure for change was exerted by DSU over a ten year period. It has been seen how DSU was closely linked with the NUS and DAUT, and how discussions on anonymous marking had taken place between the AUT and NUS. The closed politics of DSU and the Union's motivation to want to make its mark have also shown themselves as influences. Equally, organisations like the AUT have their own internal interests which have been significant in establishing a policy stance on anonymous marking: the influence of the Women's Committee and individuals like Catherine Belsey and Paul Cottrell are evidence of

---

[150] Appendix VI, which presents a summary table of departmental views on the system of anonymous examinations produced as part of the University's review, demonstrates that the principle of anonymity was never questioned as an issue.

this. The interests of the administration of the University are to keep the institution on an even keel allowing it to pursue its aims and objectives.

Perhaps foresight of the effect that modularisation and increasing continuous assessment might have on the workability of the external examiner system, made the introduction of anonymous marking necessary, in the eyes of management, to guarantee fairness, quality assurance and as a further protection against litigation. However, the one guarantee that anonymous marking could not make as a substitute for external examiners is to ensure that standards across higher education remain comparable. It was suggested that Senate (as part of the University's political system), at the time of the introduction of anonymous marking, largely consisted of individuals who were detached from the day-to-day operations of Examination Boards. Hence, one could argue that Senate had a limited appreciation of the practical implications of anonymous marking and, for this reason, were more willing to accept the recommendations put before them by the Working Party. Any committee structure has an interest in not overturning the recommendations of one of its sub-committees. The interplay of interests identified here and the positions, hidden or otherwise, of groups are the sorts of considerations that one needs to be aware of when looking at how any policy, in this case anonymous marking, comes to be introduced.

The social and political climate of opinion of the day - external to the University - is also a major factor affecting the policy process.[151] In this

---

[151] Richard A. Chapman and J. R. Greenaway describe how the nature and processes of public administration are "conditioned by the society within which it exists and which in turn it affects." They go on to say, "the numerous officials who are involved in public administration are themselves citizens; they react to and are influenced by the social and political environment to which they also contribute." The Dynamics of Administrative Reform, Croom Helm, London, 1980, p. 224.

context some account of the pressure of needing to be seen to be politically correct and to be actively endorsing an equal opportunities agenda will have had an influence. The environment within which universities are now required to work means that quality assurance has become a priority. A policy of anonymous marking might be viewed as a positive measure in this respect likely to appeal to outside observers. Equally, the anxiety within universities over possible litigation, an instance of which was referred to earlier in the Chapter (see pp. 128-129), might encourage the management of higher education institutions to adopt safeguards like anonymous marking.

One should not forget the general impact that other pressures affecting higher education would have on the university community. Increasing student:staff ratios, the introduction of modularisation, workload and funding pressures, might be a higher priority of university staff than a proposed change to assessment policy. Hence, one could argue that anonymous marking would have more chance of being implemented at a time when the energies of those who will be most directly affected by such a policy might be sharply focused on these other policy issues. There is also the fact that more and more universities are adopting the practice of anonymous marking. At least twenty-nine universities were known to have some sort of anonymous marking scheme in operation by 1992.[152] This trend might have the effect of making a university like Durham more willing to consider and eventually accept a policy that is already in operation elsewhere; in matters like this, Durham has, in the past, generally not wished to be seen as exceptional. Clearly, therefore, the

---

[152] Taken from 'A summary report into the practice of anonymous marking in British universities', University of Sunderland Students' Union, 1992. For the names of these twenty-nine universities see Appendix III.

political environment and climate of opinion are important factors to consider in the policy process.

In conclusion, it is important to remember that political activity - the activity of seeking agreement between different interests - characterises every policy decision. Thus, the study of policy making or policy change, in discovering how agreement or a decision has been achieved, must analyse the following factors: the different interests involved and the interaction within and between them; the motivations of the relevant groups and individuals; human nature in general, and the influence of particular humans; the pressures of other events and changes affecting the organisation; the social and political climate of opinion; and the mission, aims and objectives of the organisation. This thesis has attempted to provide materials for such analysis in the context of the introduction of a policy of anonymous marking at the University of Durham, which has as its mission the achievement and sustainment of the highest standards of excellence in teaching and research. It has been plain, using the example of Durham, that no single pressure can explain sufficiently the complexity of the policy and decision making process. Any attempts to understand policy change must look at it from many perspectives.

*   *   *

Areas for further study

This thesis has posed many questions and raised a host of important issues. It does not claim to have supplied all the answers, instead it is hoped that relevant questions have been raised - and often they are important questions that seem not to have been so far sufficiently

considered. This final section offers a few suggestions about where further research is needed to advance this debate.

Clearly, more research is required into the performance differences of men and women in different subjects, rather than the differences for all degree subjects taken together, which is the area in which scholars have tended to concentrate. Earlier Chapters have demonstrated that there are some striking differences between subjects which might prove better platforms for research into the reasons for these differences, where they exist.

Having examined the research which looks at the possibilty of bias in marking, it is difficult to see how bias can be isolated as a factor and identified as being at work through a controlled experiment. However, more research into this area is to be encouraged. Perhaps, in the future, there will be more scope to examine all sorts of other possible influences like race, beauty and favouritism, even though it seems virtually impossible to separate the different biases or other factors affecting the marking process. For now, though, not having the results of such research - or perhaps even an effective methodology - it is important that the safeguards of double and external marking continue to be used. Whether or not anonymous marking contributes to these safeguards to a sufficient extent to justify its adoption (bearing in mind the practical implications that have been discussed) is a matter for individual organisations to decide.

The validity of the information in this subject area should continue to be evaluated because evidence that is based on questionable assumptions and unproven conclusions can be extremely damaging if its flaws and

limitations are not recognised. It has been seen how some of the research from which policy decisions have been made is not as conclusive as it purports to be.

Finally, a longer term view of the role of anonymous assessment and how it fits into the direction in which mass higher education is heading is important. This is particularly so, in the light of developments like modular degrees, and the increasing use of multiple strategies for assessment which encourage a wider use of practices like continuous assessment, practical and oral work, and vocational qualifications, among other things. To ensure that anonymous marking is operated as efficiently and effectively as it can be, other universities should seek to conduct research into the operation of their own systems.

APPENDIX I

CITY UNIVERSITY SURVEY, APRIL 1989

Questions (57 questionnaires, 43 replies)

1. Is there a system of blind marking of examinations at your university/college?
2. Does this system apply to all departments?
3. Are the examinations which use such a system organised centrally?
4. Is anonymity removed before/at the Board of Examiners Meeting?
5. Does the system cause extra administration?

INSTITUTION	QUESTIONS				
	1	2	3	4	5
Aberdeen	No				
Bath	No				
Belfast	No				
Birkbeck, London	Yes	Yes	Yes	Yes	Yes
Birmingham	No				
Bradford	No				
Bristol	Yes	No	No	-	-
Brunel	Yes	No	No	Yes	Yes
Cambridge	Yes	No	Yes	*	No
Cardiff	Yes	No	Yes	*	Yes
Cranfield	No				
Dundee	No				
Durham	No [Since 1993/4 anonymous marking has been introduced for all dept's]				
East Anglia	Yes	No	No	Yes	No
Edinburgh	No				
Essex	Yes	Yes	Yes	Yes	Yes
Exeter	No				
Glasgow	No				
Heriot-Watt	No				
Hull	No				
Imperial, London	Yes	No	No	*	No
Keele	Yes	No	No	*	Yes
Kent	Yes	No	Yes	*	No
Lancaster	No				
Leeds	No				
Liverpool	Yes	No	*	Yes	Yes
Loughborough	No				
Manchester	No				
Newcastle	No				
Nottingham	No				
Open	No				
Oxford	No				
Reading	No				
Southampton	No				
St. Andrews	Yes	No	No	Yes	Yes
St. Davids	No				
Surrey	No				
Sussex	Yes	Yes	Yes	Yes	No
UC North Wales	No				
UC Wales	No				
Ulster	No				
UMIST	No				
York	Yes	No	*	Yes	Yes

* Indicates no definite affirmative/negative reply
- Indicates it does not apply in the Institution's case

APPENDIX II

Association of University Teachers (AUT) Policy Statement on Anonymous Marking as Endorsed at AUT Council in December 1994.

The case for and against anonymous marking

1   Anonymous marking is a system of assessment designed to ensure, as far as possible, that the identity of candidates is not known to the marker. The purpose of anonymous marking is to reduce conscious or unconscious bias for or against particular students. Bias in the marking of student's written work has been quite extensively researched, particularly in relation to sex bias. Comparisons made before and after the introduction of anonymous marking provide strong evidence that bias in marking is at least part of the reason for otherwise unexplained differences in performance between male and female students (for example, significant differences in proportions gaining firsts or upper seconds).

2   Anonymous marking cannot be applied in all circumstances. A course may involve forms of assessment which cannot be anonymous, for example, courses requiring student presentations or performance, or the observation of student practice during professional training; although, even in these cases, precautions can be taken to minimise bias. The aim should be to apply anonymous marking wherever possible.

3   Advocates of anonymous marking do not pretend that it can by itself be a complete answer to the problem of bias. It should be viewed as one element (though a very important one) of an equal opportunities programme covering all aspects of the life of a higher education institution. Introduced as part of such a programme, anonymous marking should increase the confidence of students in the impartiality of the examining system and also help to safeguard teaching staff and institutions from unfounded accusations of discrimination. At least one institution has recently introduced anonymous marking following allegations of racial discrimination.

4   The most common objection to anonymous marking from staff is that it is ineffective because teachers can identify their students from their handwriting or writing styles. This is likely to be true only in relation to a small proportion of student scripts (particularly since the increase in the size of teaching groups and the wider use of word processing) and should not be allowed to detract from the overall case in favour of anonymous marking. No system will be absolutely perfect, but ensuring anonymity at the initial and "double" marking stages, and in external examining, should provide a high degree of protection.

5   Anonymous marking does not mean that account cannot be taken of relevant special circumstances affecting a student's performance, such as illness. There are established procedures for dealing with this which usually come into play at examining board level. Obviously, it would be appropriate to lift anonymity in cases where such considerations become necessary.

6   Staff may also be understandably concerned about the potential administrative burden of anonymous marking, and about the reliability of the systems adopted. Various methods are used, the following being among the most common: candidates' papers identified solely by reference to a unique number; papers identified by name and number on a tear-off slip to be removed before marking; and removable blank labels stuck over the candidates' names. There is no doubt that setting up an efficient and reliable system of anonymous marking will cause extra administration. This is unavoidable because time must be taken both to plan the system properly and to ensure that staff and students are fully informed about its operation and rationale. Evidence from a survey of anonymous marking carried out by City University in 1989 [see Appendix I] indicates that once the system is running it involves little or no extra work and that identification errors are rare. It is essential that institutions support fully the introduction of anonymous marking by ensuring that the necessary staff resources, time and training are made available.

Conclusion: AUT policy

7   From recent surveys carried out by AUT and other organisations, it appears that about half of our universities and colleges have either adopted anonymous marking or are actively considering doing so. AUT strongly supports this development since it believes that anonymous marking, as an essential part of an equal opportunities policy, will assist in ensuring the non-discriminatory, fair treatment of all students within the assessment and examining process. AUT, nationally and locally, working in collaboration with NUS and student unions, will press all institutions to support the effective implementation of anonymous marking as comprehensively as possible.

ston University *

niversity of Bath *

niversity of Birmingham

niversity of Bradford

niversity of Brighton *

niversity of Bristol *

unel University *

niversity of Buckingham

niversity of Cambridge*

niversity of Dundee *

e Montfort University,
eicester*

urham University *

niversity of East Anglia

niversity of Edinburgh *

niversity of Essex *

niversity of Exeter *

niversity of Glasgow *

niversity of Greenwich *

eriot-Watt University

niversity of Huddersfield

niversity of Hull

niversity of Keele

niversity of Kent at Canterbury *

ncaster University*

niversity of Leeds

University of Leicester

University of Liverpool

Liverpool John Moores
University

London Guildhall University *

Loughborough University *

Manchester Metropolitan
University

University of Manchester *

Napier University

University of Newcastle upon
Tyne*

University of Northumbria at
Newcastle

University of Nottingham

University of Oxford

University of Paisley

University of Portsmouth

University of Reading

Robert Gordon University

Sheffield Hallam University

Staffordshire University

University of Southampton *

University of St. Andrews *

University of Strathclyde *

University of Surrey

University of Sussex *

University College of Swansea*

University of Teesside

University College of Wales,
Aberystwyth *

University College of Wales,
Cardiff *

University of the West of England,
Bristol*

University of Wolverhampton

University of York *

*** Operates some system
of anonymous marking**

Taken from 'A summary report into
the practice of anonymous marking
in British universities', University
of Sunderland Students' Union, 1992.

APPENDIX IV

21 May 1995

Jake Yeo
Department of Politics
University of Durham
48 Old Elvet
DURHAM
DH1 3LZ

Dear Jake,

Thank you for your letter and my apologies for the delay in getting an answer to you. I have prepared the bibliography, but most of the sources are internal and therefore unpublished, so I have been waiting to hear from the relevant institutions to ascertain that copies are available before I finalise the list.

Mindful of the fact that you might be trying to introduce policy at Durham and working to a deadline, I am enclosing a copy of the details I have to date – and sorry not to be more helpful. I will send the final version on to you when it is ready.

Best wishes

Emma Westcott

ANONYMOUS OR 'BLIND' MARKING – A LIST OF SOURCES

The association has produced the following bibliography in response to frequent enquiries about anonymous or 'blind' marking in university assessment.

AUT Woman

Catherine Belsey, *'Marking by numbers'*, Autumn 1988

Linda Fitzsimmons, *'Gender and exam results'*, Autumn 1989

Hardip Begol, UNC West Midlands

A guide to anonymous marking, DATE

British Journal of Social Psychology

Clare Bradley, *'Sex bias in the evaluation of students'*, 1984, 23, pp 147–53

City University

AUTHOR, TITLE, April 1989

Polytechnic of East London (now UEL)

AUTHOR, TITLE, 1989?

Glasgow Dental School

AUTHOR, TITLE, DATE

The Guardian

James Meikle, *'Unsigned degree papers mark the death of gender bias'*, 15.6.93

NUS

NUS Briefing: *Education Campaign – Anonymous Marking*, August 1993

NUS Scotland: *Mark My Words* campaign, DATE

New Society

Maryon Tysoe, *'Do colleges mark women down?'*, 9.12.82

Sheffield University

Clare Bradley, TITLE, DATE

Times Higher Educational Supplement

Sian Griffiths, *'Numbers that count'*, 3.8.90

University of Wales at Cardiff

Catherine Belsey, TITLE, YEAR

University of Warwick

Simon Clarke, TITLE, DATE

h:\...emma.anon

PERSONNEL OFFICE Memorandum

To: All Academic and Related Staff

From: Director of Personnel

Date: 15 November 1994

Subject: Providing References

Following a recent House of Lords ruling (*Spring v Guardian Royal Exchange (1994)*, employers may be under a general implied contractual duty to provide references for present and former employees (for the latter only if within a reasonable time after they have left employment). The House of Lords also ruled that there was a duty of care to the employee in the compiling of references. Some practical effects of these rulings are:

- References must be fair and accurate to avoid claims for negligence (from either the employee or prospective employer), but there is no breach of the duty of care if the statements in the reference are true;

- Disclaimers of liability are unlikely to have any practical legal effect;

- Those providing a reference need to be able to support their opinions, if need be, by documentary evidence, otherwise it is best to stick to facts (which have been documented);

- Even where a positive reference is provided, the prospective employer may have a right to claim against the referee if the employee fails to measure up to expectations brought about by the referee's report;

I shall be seeking further advice on this issue, particularly whether there are any general implications for the provision of references for students. In the meantime, my advice is that if you are requested to provide a reference for a current or former employee, unless the reference is likely to be satisfactory in all respects, it is sensible only to provide either the blandest of references or one which goes little further than merely confirming the period of employment and the nature of their job while at Durham.

JB

UNIVERSITY OF DURHAM

REVIEW OF ANONYMOUS MARKING

Departments	No change to current system	Adopt a single reference device	Retain label	Use attendance card to record name	Other special proposals
English		✓	✓		✓
French		✓	✓		
SMEL		✓	✓		
Classics		✓	✓		
Geological Sciences		✓		✓	✓
Engineering		✓		✓	
Mathematics					✓
Physics			✓		
Psychology		✓			
Archaeology		✓		✓	
CMEIS		✓			
Geography		✓			✓
History		✓			
Law					✓
Politics		✓	✓		
Sociology		✓		✓	
Dean of Science		✓	✓		
Dean of Social Sciences		✓			

JH/JS/RevTab
5 October 1994

BIBLIOGRAPHY

Adelman, C. (ed.) (1986) Assessment in Higher Education: Issues and Contexts, US Department of Education, Washington DC.

Allen, M. (1988) The Goals of Universities, The Society for Research into Higher Education and Open University Press, Milton Keynes.

Archer, J. (1992) 'Sex bias in evaluations at college and work', The Psychologist, 5, pp. 200-204.

Archer, J. and McCarthy, B. (1988) 'Personal biases in student assessment', Educational Research, 30, pp. 142-145.

Association of University Teachers (December 1994) 'Policy statement on anonymous marking'.

Astin, A. W. (1991) Assessment for Excellence: The philosophy and practice of assessment and evaluation in higher education, Macmillan, New York.

Atkins, M. J., Beattie, J. and Dockrell, W. B. (1993) Assessment Issues in Higher Education, Employment Department, Sheffield.

Balderston, F. E. (1974) Managing Today's University, Jossey-Bass, San Francisco.

Beard, R. M., Healey, F. G. and Holloway, P. J. (1974) Objectives in Higher Education, 2nd edition, Society for Research in Higher Education, London.

Begol, H. (October 1992) 'A guide to anonymous marking', UNC West Midlands.

Belsey, C. (Autumn 1988) 'Marking by numbers', AUT Woman, 15.

Benbow, C. P. (1988) 'Sex differences in mathematical reasoning ability in intellectually talented preadolescents: their nature, effects and

possible causes', <u>Behavioural and Brain Sciences</u>, 11, pp. 169-232.

Bentley, A. F. (1967) <u>The Process of Government</u>, editorial by Odegard, P. H., The John Harvard Library: The Belknap Press of Harvard University Press, Cambridge, Massachusetts.

Berridge, G. R. (26 August 1994) 'The trouble with upper classes', <u>The Times Higher Education Supplement</u>.

Black, B. J. (1968) 'University examinations', <u>Physics Education</u>, 3, pp. 93-99.

Bligh, D. (1990) <u>Higher Education</u>, Cassell, London.

Block, R. A., Arnott, D. P., Quigley, B. and Lynch, W. C. (1989) 'Unilateral nostril breathing influences lateralized cognitive performance', <u>Brain and Cognition</u>, 9, pp. 181-190.

Bogdanor, V. (ed.) (1991) <u>The Blackwell Encyclopedia of Political Science</u>, Blackwell, Oxford.

Booth, A. (1993) 'Learning History in university: student views on teaching and assessment', <u>Studies in Higher Education</u>, 18, 2, pp. 233-234.

Bradley, C. (1984) 'Sex bias in the evaluation of students', <u>British Journal of Social Psychology</u>, 23, pp. 147-153.

------- (1985) 'Sex bias in examining reconsidered: a rejoinder to Rudd', <u>Studies in Higher Education</u>, 10, 1, pp. 91-94.

------- (1993) 'Sex bias in student assessment overlooked?', <u>Assessment and Evaluation in Higher Education</u>, 18, 1, pp. 3-8.

------- (1993) 'Sex bias in evaluations at college and work: A comment on Archer's review', <u>The Psychologist: Bulletin of the British Psychological Society</u>, 6, pp. 152-154.

Briggs, D. (1970) 'The influence of handwriting on assessment',
    Educational Research, 13, pp. 50-55.

British Psychological Society (1989) Guidelines for External Examiners
    on Undergraduate Psychology Degrees, The British Psychological
    Society, Leicester.

Brockbank, P. (1968) 'Examining exams', Times Literary Supplement,
    no. 3465, pp. 781-782.

Bull, R. and Stevens, J. (1979) 'The effects of attractiveness of writer
    and penmanship on essay grades', Journal of Occupational
    Psychology, 52, pp. 53-59.

Carter, A. (1993) 'Reply to Richard A. Chapman article "Civil service
    recruitment: Fairness or preferential advantage?" (volume 8, no.
    2)', Public Policy and Administration, 8, 3, pp. 46-48.

Carter, C. F. (1980) Higher Education for the Future, Blackwell,
    London.

Chapman, Richard A. (1968) Decision Making, Routledge and Kegan
    Paul, London.

------- (1977) 'Efficiency and effectiveness in the civil service', Appendix
    25, Eleventh Report from the Expenditure Committee Session
    1976-7, vol. III, Appendices, HC 535-III, HMSO, London.

------- (1988) Ethics in the British Civil Service, Routledge, London.

------- (1993) 'Ethics in the public sector', Politeia, 12, 2,   pp. 28-42.

------- (1993) 'Civil service recruitment: Fariness or preferential
    advantage?', Public Policy and Administration, 8, 2, pp. 68-73.

------- (20 January 1995) 'Hostage of exam fortune', Times Higher
    Education Supplement.

Chapman, Richard A. and Greenaway, J. R. (1980) The Dynamics of Administrative Reform, Croom Helm, London.

Charlton, B. and Smith, K. (4 December 1992) 'Tripping on red tape', Times Higher Education Supplement.

Chase, C. (1968) 'The impact of some obvious variables on essay test scores', Journal of Educational Measurement, 5, pp. 315-318.

------- (1986) 'Essay test scoring: interaction of relevant variables', Journal of Educational Measurement, 23, pp. 38-41.

City University (April 1989) Anonymous Marking Survey.

Clarke, S. (1988) 'Another look at the degree results of men and women', Studies in Higher Education, 13, pp. 315-331.

Clift, J. C. and Imrie, B. W. Assessing Students, Appraising Teaching, Croom Helm, London.

Cockburn, B. and Ross, A. (no date) Inside Assessment, Teaching in Higher Education Series, No. 7, School of Education, University of Lancaster.

Committee of Enquiry into the Academic Validation of Degree Courses in Public Sector Higher Education (April 1985) 'Academic Validation in Public Sector Higher Education', chaired by Sir Norman Lindop, HMSO, Cmnd 9501.

Committee of Vice-Chancellors and Principals (April 1984) 'The external examiner system for first degree and taught master's courses'.

------- (July 1986) 'Academic Standards in Universities: Universities' methods and procedures for maintaining and monitoring academic standards in the content of their courses and in the quality of their teaching', chaired by Professor P. A. Reynolds.

Cornford, F. M. (1973) <u>Microcosmographia Academica</u>, ninth impression, foreword by Guthrie, W. K. C., Bowes and Bowes, London.

Cox, R. (1967) 'Examinations and higher education: A review of the literature', <u>Universities Quarterly</u>, 21, pp. 292-340.

------- (1969) 'Reliability and validity of examinations', <u>The World Year Book of Education</u>, Evans Bros, London.

Council for National Academic Awards (May 1980) 'Notes for guidance for external examiners', in CNAA Handbook (1984).

Crick, B. (1982) <u>In Defence of Politics</u>, 2nd edition, Penguin, Harmondsworth.

Dale, R. (1959) 'University standards', <u>Universities Quarterly</u>, 13, pp. 186-195.

Dalton, K. (1968) 'Menstruation and examinations', <u>Lancet</u>, 28 December, pp. 1386-1388.

Davies, M. R. and Lewis, V. A. (1971) <u>Models of Political Systems</u>, Macmillan, London.

Deaux, K. and Taynor, J. (1973) 'Evaluation of male and female ability: Bias works two ways', <u>Psychological Reports</u>, 32, pp. 261-262.

Durham Students' Union (1989) 'Anonymous Exam Papers: The arguments', paper submitted to the Joint Committee of the University and Durham Students' Union.

Eagly, A. H. and Mladinic, A. (1989) 'Gender stereotypes and attitudes toward men and women', <u>Personality and Social Psychology Bulletin</u>, 15, pp. 543-558.

Eames, K. and Loewenthal, K. (1990) 'Effects of handwriting and examiner's expertise on assessment of essays', The Journal of Social Psychology, 130, pp. 831-833.

Easton, D. (1959) The Political System: An Inquiry into the State of Political Science, reprinted edition, Alfred A. Knopf, New York.

Ebel, R. (1965) Measuring Educational Achievement, Prentice Hall, New York.

Edinburgh University Students' Association (February 1993) 'Recommendations for the adoption of anonymous marking', paper submitted to the Student Affairs Committee.

Entwistle, N. J. and Entwistle, D. (1970) 'The relationship between personality, study methods and academic performance', British Journal of Educational Psychology, 40, p. 132.

Entwistle, N. J. and Wilson, J. D. (1977) Degrees of Excellence: The academic achievement game, Hodder and Stoughton, London.

Fajardo, D. M. (1985) 'Author race, essay quality, and reverse discrimination', Journal of Applied Social Psychology, 15, pp. 255-265.

Farrell, M. J. and Gilbert, N. (1960) 'A type of bias in marking examination scripts', British Journal of Educational Psychology, 30, pp. 47-52.

Feather, N. T. (1975) 'Positive and negative reactions to male and female success and failure in relation to the perceived status and sex-typed appropriateness of occupations', Journal of Personality and Social Psychology, 31, pp. 536-548.

Feather, N. T. and Simon, J. G. (1975) 'Reactions to male and female success and failure in sex-linked occupations: Impressions of personality, causal attributions and perceived likelihood of

different consequences', <u>Journal of Personality and Social Psychology</u>, 31, pp. 20-31.

Fenske, R. H. 'Setting institutional goals and objectives', in Jedamus, P. *et al.* (1981) <u>Improving Academic Management</u>, Jossey-Bass, London.

Fitzgerald, M. (28 October 1994) 'Assessment and audit in disrepute', <u>Times Higher Education Supplement</u>.

Fitzsimmons, L. (Autumn 1989) 'Gender and exam results', <u>AUT Woman</u>.

Gipps, C. and Murphy, P. (1994) <u>A Fair Test? Assessment, achievement and equity</u>, Open University Press, Buckingham.

Goldberg, P. A. (1968) 'Are women prejudiced against women?', <u>Transaction</u>, 5, pp. 28-30.

Goodhart, C. B. (1988) 'Women's examination results', <u>The Cambridge Review</u>, 109, pp. 38-40.

------- (1988) 'Examination results in single sex and mixed colleges at Cambridge', <u>The Cambridge Review</u>, 109, pp. 139-141.

------- (1992) 'Sex and class in examinations', <u>The Cambridge Review</u>, 113, pp. 43-44.

Greenwood, J. and Wilson, D. (1989) <u>Public Administration in Britain Today</u>, 2nd edition, Unwin Hyman, London.

Griffiths, S. (3 August 1990) 'Numbers that count', <u>Times Higher Education Supplement</u>.

Gronlund, N. E. (1968) <u>Readings in Measurement and Evaluation</u>, The Macmillan Company, New York.

Hall, F. S. and Hall, D. T. (1976) 'Effects of job incumbents' race and sex on evaluations of managerial performance', <u>Academy of Management Journal</u>, 19, pp. 476-481.

Halpern, D. F. (1992) <u>Sex Differences in Cognitive Abilities</u>, NJ: Lawrence Erlbaum Associates, Hillsdale.

Hannabus, K. C. (1991) 'Mixed results', <u>Oxford Magazine</u>, 74, pp. 4-5.

------- (1992) 'Mixed results', <u>The Cambridge Review</u>, 113, pp. 40-42.

Harris, N. D. C. (1975) 'What is assessment?', <u>Assessment in Higher Education</u>, 1, pp. 5-12.

Harrison, P. (1994) 'The CSSB supplementary application form: A candidate's reply to Adrian Carter', <u>Public Policy and Administration</u>, 9, 1, pp. 65-67.

Hartle, T. W. 'The growing interest in measuring the educational achievement of college students' in Adelman, C. (ed.) (1986) <u>Assessment in Higher Education: Issues and Contexts</u>, US Department of Education, Washington DC.

Hartley, J. (1991) 'Sex differences in handwriting: A comment on Spear', <u>British Educational research Journal</u>, 17, 2, pp. 141-145.

------- (1992) 'Sex bias, blind marking and assessing students', <u>Psychology Teaching Review</u>, 1, 2, pp. 66-73.

Hartog, P. and Rhodes, E. C. (1935) <u>An Examination of Examinations</u>, Macmillan, London.

Hayward, C. (1979) <u>A Fair Assessment: Issues in evaluating coursework</u>, Central Council for Education and Training in Social Work.

Heim, A. (1970) <u>Intelligence and Personality</u>, Penguin, Harmondsworth.

Henderson, E. S. (1980) 'The essay in continuous assessment', <u>Studies in Higher Education</u>, 5, p. 197.

Heywood, J. (1977) <u>Assessment in Higher Education</u>, Wiley, London.

Heywood, J. (April 1994) 'Enterprise Learning and its Assessment in Higher Education', A Technical Report published by the Employment Department's Learning and Methods Branch, Research and Development Series, Report No. 20.

Highfield, R. (7 July 1995) 'Women hold on to middle ground in the battle of brains', <u>The Daily Telegraph</u>.

Hines, M. 'Gonadal hormones and human cognitive development', in Balthazart, J. (ed.) (1990) <u>Hormones, Brain and Behaviour in Verebrates</u>, Vol. 1, <u>Sexual Differentiation, Neuroanatomical Aspects, Neutro transmitters and Neuropeptides</u>, Basel: Karger.

House of Commons (1976 and 1977) <u>The Civil Service, Eleventh Report from the Expenditure Committee (General Sub-Committee)</u>, Sessions 1975-76 and 1976-77 and Appendices, HC 535-I and HC 535-II, HMSO.

Hudson, B. (1973) <u>Assessment Techniques</u>, Methuen, London.

Hughes, D. C., Keeling, B. and Tuck, B. F. (1980) 'Essay marking and the context problem', <u>Educational Research</u>, 22, pp. 147-148.

Husbands, C. T. (1976) 'Ideological bias in the marking of examinations: a method of testing for its presence and its implications', <u>Research in Education</u>, 15, pp. 17-38.

Hyde, J. S., Fennema, E. and Lamon, S. J. (1990) 'Gender differences in mathematics performance: A meta analysis', <u>Psychological Bulletin</u>, 107, pp. 139-153.

Hyde, J. S. and Linn, M. C. (1988) 'Gender differences in verbal ability: A meta analysis', <u>Psychological Bulletin</u>, 104, pp. 53-69.

Inkster, I. (1971) 'Some criticisms of a coursemark system', <u>Higher Education Review</u>, 3, pp. 45-48.

James, H. (1927) 'The effect of handwriting on grading', <u>English Journal</u>, 16, pp. 180-205.

Jedamus, P. *et al.* (1981) <u>Improving Academic Management</u>, Jossey-Bass, London.

Jordan, T. E. (1989) <u>Measurement and Evaluation in Higher Education</u>, Falmer Press, London.

Kaplan, R. M. (1978) 'Is beauty talent? Sex interaction in the attractiveness halo effect', <u>Sex Roles</u>, 4, pp. 195-204.

Kiley, S. (Autumn 1988) 'Sexism shown in understy exams', <u>The Times, Student Special</u>.

Kirk, G. (6 January 1995) 'Practice with a clearer purpose', <u>Times Higher Education Supplement</u>.

Klug, B. (1976) 'To grade, or not to grade - a dramatic discussion in eleven parts', <u>Studies in Higher Education</u>, 1, 2, pp. 203-207.

------- (1977) <u>The Grading Game</u>, NUS Publications, London.

Landy, D. and Sigall, H. (1974) 'Beauty is talent: Task evaluation as a function of the performer's physical attractiveness', <u>Jounal of Personality and Social Psychology</u>, 29, pp. 299-304.

Layton, D. (ed.) (1968) <u>University Teaching in Transition</u>, Oliver and Boyd, Edinburgh.

Lenney, E., Mitchell, L. and Browning, C. (1983) 'The effect of clear criteria on sex bias in judgments of performance', <u>Psychology of Women Quarterly</u>, 7, pp. 313-328.

Lennox, B. (1974) <u>Hints on the Setting and Evaluation of Multiple Choice Questions of the One-from-Five Type</u>, Booklet No. 3, Association for the Study of Medical Education, Dundee.

Linn, M. C. and Peterson, A. C. 'A meta analysis of gender differences in spatial ability: Implications for mathematics and science achievement', in Hyde, J. S. and Linn, M. C. (eds) (1986) <u>The Psychology of Gender: Advances Through Meta-Analysis</u>, MD: John Hopkins Press, Baltimore.

Locksley, A., Borgida, E., Brekke, N. and Hepburn, C. (1980) 'Sex stereotypes and social judgements', <u>Journal of Personality and Social Psychology</u>, 39, pp. 821-831.

Lockwood, G. and Davies, J. (1985) <u>Universities: The Management Challenge</u>, The Society for Research into Higher Education and NFER-NELSON, Windsor.

Loewenthal, K. (13 March 1980) 'Handwriting and personality', <u>New Society</u>, pp. 544-546.

Lovell, K. 'Examinations and marking', in Layton, D. (ed.) (1968) <u>University Teaching in Transition</u>, Oliver and Boyd, Edinburgh.

McCrum, N. G. (1991) 'A fair admissions system', <u>Oxford Magazine</u>, 72, pp. 16-17.

------- (31 March 1995) 'The second sex', <u>Times Higher Education Supplement</u>.

------- (21 April 1995) 'Sixties peak of female performance', <u>Times Higher Education Supplement</u>.

McCullough, M. L. (1987) 'Blind marking and gender identity', <u>Bulletin of the British Psychological Society</u>, 40, p. 103.

McDavid, J. W. and Harari, H. (1973) 'Name, stereotype and teachers' expectations', Journal of Educational Psychology, 65, pp. 222-225.

McVey, J. (1975) 'The errors in marking examination scripts in electronic engineering', International Journal of Electrical Engineering Education, 12, p. 203.

MacIntosh, H. G. (1974) Techniques and Problems of Assessment, Edward Arnold, London.

Markham, L. R. (1976) 'Influence of handwriting quality on teacher evaluation of written work', American Educational Research Journal, 13, pp. 277-283.

Marshall, J. C. and Powers, J. M. (1969) 'Writing neatness, composition errors and essay grades', Journal of Educational Measurement, 6, pp. 97-101.

Marston, P. (21 June 1995) 'Oxbridge women losing battle of sexes', The Daily Telegraph.

Massey, A. (1983) 'The effects of handwriting and other incidental variables of GCE A level marks in English literature', Educational Review, 35, pp. 45-50.

Meikle, J. (15 June 1993) 'Unsigned degree papers mark the death of gender bias', The Guardian.

------- (15 December 1994) 'Anonymity "may end exam bias"', The Guardian.

Miller, A. G. (1970) 'Role of attractiveness in impression formation', Psychonomic Science, 19, pp. 241-243.

Miller, C. M. L. and Parlett, M. (1974) Up to the Mark: A study of the examination game, Society for Research into Higher Education, London.

Miller, J. D. B. (1962) <u>The Nature of Politics</u>, Gerald Duckworth and Co. Ltd., London.

Mischel, H. N. (1974) 'Sex bias in the evaluation of professional achievements', <u>Journal of Educational Psychology</u>, 2, pp. 157-166.

Moodie, G. C. and Eustace, R. (1974) <u>Power and Authority in British Universities</u>, George Allen and Unwin, London.

Murphy, R. J. L. (1979) 'Removing the marks from examination scripts before re-marking them: Does it make any difference?', <u>British Journal of Educational Psychology</u>, 49, pp. 73-78.

------- (1982) 'Sex differences in objective test performance', <u>British Journal of Educational Psychology</u>, 52, pp. 213-219.

National Association of Teachers in Further and Higher Education (no date) 'Assessment', a pamphlet produced by the Examinations Panel of the Association of Teachers in Technical Institutions.

National Union of Students (no date) 'Mark my words', NUS Scotland Campaign for Anonymous Marking.

------- (August 1993) 'Education campaign - anonymous marking', NUS briefing.

Newstead, S. E. (1992) 'The use of examination in the assessment of psychology students', <u>Psychology Teaching Review</u>, 1, 1, pp. 22-33.

Newstead, S. E. and Dennis, I. (1990) 'Blind marking and sex bias in student assessment', <u>Assessment and Evaluation in Higher Education</u>, 15, 2, pp. 132-139.

------- (October 1993) 'Bias in Student Assessment', <u>The Psychologist</u>, pp. 451-452.

Nieva, V. F. and Gutek, B. A. (1980) 'Sex effects of evaluation', Academy of Management Review, 5, pp. 267-276.

Norman, M. (6 May 1993) 'Why men get more firsts', The Times.

Nuffield Group for Research and Innovation in Higher Education (1975) A Question of Degree: Assorted papers on assessment, Nuffield Foundation, London.

Nuttall, D. L. and Willmott, A. S. (1972) British Examinations Techniques of Analysis, National Foundation for Educational Research in England and Wales, Windsor.

Parlour, J. (26 August 1994) 'Degrading end to study time', Times Higher Education Supplement.

Parry-Langdon, N. (May 1990) '"Marking by Numbers": Evaluation of the marking of final degree examinations in the Faculty of Humanities and Social Studies', Report prepared for the Deans Committee, University of Cardiff Social Research Unit.

Pheterson, G. T., Kiesler, S. B. and Goldberg, P. A. (1971) 'Evaluation of the performance of women as a function of their sex, achievement and personal history', Journal of Personality and Social Psychology, 19, pp. 114-118.

Pidgeon, D. A. and Yates, A. (1968) An Introduction to Educational Measurement, Routledge and Kegan Paul, London.

Race, P. (16 September 1994) 'The one quality that counts', Times Higher Education Supplement.

Rhodes, G. (1965) Administrators in Action, Vol. II, Allen and Unwin, London.

Rippere, V. L. (1974) 'On the validity of university examinations', Universities Quarterly, 28, pp. 209-218.

Robbins, C. B. (1963) <u>Report of the Committee on Higher Education,</u> cmnd 2154-XII (Robbins Report), HMSO, London.

Roberts, G. K. (1971) <u>A Dictionary of Political Analysis</u>, Longman, London.

Rowntree, D. (1977) <u>Assessing Students: How Shall We Know Them?,</u> Harper and Row, London.

Rudd, E. (1984) 'A comparison between the results achieved by women and men studying for first degrees in British universities', <u>Studies in Higher Education</u>, 9, pp. 47-57.

------- (1985) 'A reply to Bradley', <u>Studies in Higher Education</u>, 10, 1, pp. 95-96.

------- (1988) 'Reply to Clarke', <u>Studies in Higher Education</u>, 13, 3, pp. 333-336.

Ryan, J. 'IQ - the illusion of objectivity', in Richardson, K. and Spears, D. (eds) (1972) <u>Race, Culture and Intelligence</u>, Penguin, London.

Sanders, C. (23 September 1994) 'Degree doubts', <u>Times Higher Education Supplement</u>.

------- (4 November 1994) 'External Examiners given OK', <u>Times Higher Education Supplement</u>.

------- (23 June 1995) 'External examiner reform', <u>Times Higher Education Supplement</u>.

Sattler, J. M. (1974) <u>Assessment of Children's Intelligence</u>, Revised Reprint, W. B. Saunders Company, Philadelphia.

------- (1988) <u>Assessment of Children's Intelligence and Other Special Abilities</u>, 2nd edition, Allyn and Bacon, Boston.

Schiff, W. and Oldak, R. (1990) 'Accuracy of judging time to arrival: Effects of modality, trajectory and gender', <u>Journal of Experimental Psychology: Human Perception and Performance</u>, 16, pp. 303-316.

Sheppard, E. M. (1929) 'The effect of quality of penmanship on grades', <u>Journal of Educational Research</u>, 19, pp. 102-105.

Smith, G. A. and McPhee, K. A. (1987) 'Performance on a coincidence timing task correlates with intelligence', <u>Intelligence</u>, 11, pp. 161-167.

Snow, C. P. (1966) <u>Corridors of Power</u>, Penguin, Harmondsworth.

Society for Research into Higher Education (1980) <u>Higher Education at the Crossroads: Research into Higher Education 1965-1995</u>, (ed) Oxtoby, R., The Society, Guildford.

Spear, M. (1984) 'The biasing influence of pupil sex in a science marking exercise', <u>Research in Science and Technological Education</u>, 2, 1, pp. 55-60.

Stafford, R. E. (1961) 'Sex differences in spatial visualization as evidence of sex linked inheritance', <u>Perceptual Motor Skills</u>, 13, p. 428.

------- (1963) <u>An Investigation of Similarities in Parent-Child Test Scores for Evidence of Hereditary Components</u>, NJ: Education Testing Service, Princeton.

Stewart, D. (1988) 'Women and men', <u>Oxford Magazine</u>, 39, p.16.

Stones, I., Beckman, M. and Stephens, L. (1982) 'Sex differences in mathematical competencies of pre-calculus college students', <u>School Science and Mathematics</u>, 82, pp. 295-299.

Swim, J., Borgida, E., Marayuma, G. and Myers, D. G. (1989) 'Joan McKay versus John McKay: Do gender stereotypes bias evaluations?', <u>Psychological Bulletin</u>, 105, pp. 409-429.

Targett, S. (10 March 1995) 'Exam mark protest', <u>Times Higher Education Supplement</u>.

Terborg, J. R. and Ilgen, D. R. (1975) ' A theoretical approach to sex discrimination in traditionally masculine occupations', <u>Organisational Behaviour and Human Performance</u>, 13, pp. 352-376.

Thorndike, R. L. and Hagen, E. (1977) <u>Measurement and Evaluation in Psychology and Education</u>, 4th edition, Wiley, London.

Thyne, J. M. (1974) <u>Principles of Examining</u>, University of London Press, London.

Times Higher Education Supplement Editorial (31 March 1995) 'Of motes and beams', <u>Times Higher Education Supplement</u>.

Tittle, C. K. and Miller, K. M. (1976) <u>Assessing Attainment</u>, Independent Assessment and Research Centre, London.

Townsend, M. A. R., Lee Yong Kek and Tuck, B. F. (1989) 'The effect of mood on the reliability of essay assessment', <u>British Journal of Educational Psychology</u>, 59, pp. 232-240.

Truman, D. B. (1953) <u>The Governmental Process</u>, Alfred A. Knopf, New York.

Truscot, B. (1943) <u>Red Brick University</u>, Faber and Faber Ltd., London.

Tversky, A. and Kanneman, D. (1974) 'Judgements under uncertainty: heuristics and biases', <u>Science</u>, 185, pp. 1124-1131.

Tysoe, M. (9 December 1982) 'Do colleges mark women down?', <u>New Society</u>.

University of Durham (no date) 'Investing in Excellence', University of Durham Mission Statement.

------- (1980-1995) Senate Minutes, including supporting documents.

------- (June 1994) 'First degree result analysis by course and Faculty',
Planning Section.

------- (5 September 1994) 'Review of anonymous marking procedures',
University of Durham Memorandum, Academic Registrar.

------- (4 October 1994) 'Review of anonymous marking procedures',
Academic Registrar.

------- (1994) 'Extract from the Report to the [Senate] Teaching and
Learning Committee from the Advisory Group on Examinations
and Assessment Meetings held on 21 and 28 October 1994'.

------- (15 November 1994) University of Durham Memorandum on the
subject of providing references, Director of Personnel.

University of Essex (1993) 'Code of practice on equal opportunities for
students'.

University Grants Committee (1967-1979) Statistics of Education, Vol.
6, HMSO, London, annually 1967-79.

University of Sunderland Students' Union (1992) 'A summary report into
the practice of anonymous marking in British universities'.

University of Warwick Students' Union (March 1993) 'Should the Law
Department have anonymous marking for assessed work?'
Questionnaire prepared by Josh Wong.

Vernon, P. E. (1965) The Measurement of Human Abilities, University
of London Press, London.

Wallas, G. (1962) Human Nature in Politics, reprinted edition,
Constable and Company Ltd., London.

Weedon, C. (1982) 'Engendering stereotypes', <u>Journal of Literature, Teaching and Politics</u>, 1, pp. 37-49.

Wesman, A. G. (1952) 'Reliability and Confidence', <u>Test Service Bulletin</u>, No. 44, The Psychological Corporation, New York.

Wickham, M. (1958) 'The effects of the menstrual cycle on test performance', <u>British Journal of Psychology</u>, 49, pp. 34-41.

Willson, F. M. G. (1961) <u>Administrators in Action</u>, Vol. I, Allen and Unwin, London.

Williams, W. F. (1979) 'The role of the external examiner in first degrees', <u>Studies in Higher Education</u>, 4, p. 161.

Wiseman, H. V. (1966) <u>Political Systems</u>, Routledge and Kegan Paul, London.

Wiseman, S. (1961) <u>Examinations and English Education</u>, Manchester University Press, Manchester.

Wojtas, O. (16 April 1993) 'Anonymous marking at Edinburgh', <u>Times Higher Education Supplement</u>.

Young, P. T. (1931) 'Sex differences in handwriting', <u>Journal of Applied Psychology</u>, 15, pp. 486-498.