

An examination of the incidence of 'error variation' in the grading of law assessments.

**Jim Hanlon,
Sheffield Hallam University
Michael Jefferson, The University of Sheffield
Mike Molan, London South Bank University
Brian Mitchell, University of Wolverhampton**

Contents

1. Acknowledgements
2. Executive Summary
3. Introduction
4. Background: Assessment and Grading – an Overview
5. Methodologies/Description of research
6. Institutional Literature Review
7. Ethical parameters
8. The relevant QAA precepts
9. The Pilot Study
 - 9.1 Assessor Group 1: Exercise 1
 - 9.1.1 Assessor Group 1: Exercise 1(a) – course works
 - 9.1.2 Assessor Group 1: Exercise 1(b) – exam scripts
 - 9.2 Assessor Group 2: Exercise 2
 - 9.2.1 Assessor Group 2: Exercise 2(a) – course works
 - 9.2.2. Assessor Group 2: Exercise 2(b) – exam scripts
10. Broadening the Hypothesis
11. An exploration of intuitive marking:Exercise 3 – Assessor group 3 York ALT
 - 11.1 Exercise 3: Assessor Group 3
 - 11.1.1 Results of all Assessor Group 3 markers for Papers A & B
 - 11.1.2 Exercise 3: Results extracted for Criminal law markers only
 - 11.1.3 Exercise 3: Results extracted for non-criminal law markers only
 - 11.2 Exercise 3 - Summary of grades awarded by reference to marker HEI background
12. Intuitive and criterion-referenced marking
13. Exercise 4 – the Criminal Law Marking Workshop
 - 13.1 Exercise 4 - Unmoderated criterion referenced marking exercise
 - 13.2 Exercise 4 – Criterion referenced expert criminal law mock moderation meeting
14. Exercise 5: Parallel marking in Criminal Law
15. Conclusions
16. Findings
 - 16.1 Intuitive understanding of standards

16.2. Use and understanding of assessment criteria insufficient on its own.
16.3 Agreement as to student achievement dependant on rigorous moderation

17. Recommendations

17.1 Resourcing of assessment processes

17.2. Use of statistical tools to assist assessment

17.3 Further Research

17.4. Staff development

1 – Exercise 2 tables 2(c) and 2(d)

2 – Criminal law problem question

3 – Briefing for Exercise 3 ALT York 2004

4 - Exercise 3 – Complete data for ALT York 2004 intuitive marking exercise – all markers Papers A and B

5 - Complete data for ALT York 2004 intuitive marking exercise – Results extracted for Non-criminal law markers Papers A and B
6 – Marking Guidance for Criminal Law Marking Workshop

7 – Guidance on marking standards for Criminal Law Marking Workshop

8 – Criminal Law Marking Workshop individual marker record sheet Appendix

9 - Exercise 4 – Criminal law Marking Workshop – complete results data for Assessor Group 4

10 - Criminal Law Marking Workshop Moderation exercise record sheet

11 - Exercise 5 – Criminal Law Parallel Marking – complete results data for Assessor Group 5

1. Acknowledgements

We are grateful to the United Kingdom Centre for Legal Education (UKCLE) for the grant to conduct the research on which this report is based and particularly to Jo Sellick, of the University of Plymouth, for her encouragement and assistance. Our special thanks are extended to our colleagues across the sector who participated with such enthusiasm in the marking exercises that formed the basis of this research.

For further information contact:

<p>Jim Hanlon School of Social Science and Law Sheffield Hallam University Collegiate Crescent Campus Sheffield S10 2BP</p> <p>Tel: 0114 225 2543 Fax: 0114 225 2430 Email: J.Hanlon@shu.ac.uk</p>	<p>Michael Jefferson Department of Law University of Sheffield Crookesmoor Building Conduit Road Sheffield S10 1FL</p> <p>Tel: 0114 222 6771 Fax: 0114 222 6832 Email: M.Jefferson@sheffield.ac.uk</p>
<p>Mike Molan London South Bank University 103 Borough Road London SE1 0AA</p> <p>Tel: 020 7815 5733 Email: molanm@lsbu.ac.uk</p>	<p>Brian Mitchell School of Legal Studies University of Wolverhampton Molineux Street Wolverhampton WV1 6SB</p> <p>Tel: 01902 321511 Email: B.M.Mitchell2@wlv.ac.uk</p>

© Sheffield Hallam University, University of Sheffield, London South Bank University, University of Wolverhampton.

All rights reserved. No part of this report may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic or otherwise, without prior written permission of all four copyright holders.

2. Executive Summary

- The exercises conducted as part of this research project indicate that the marks awarded by legal academics, presented with identical pieces of student work, vary considerably – suggesting the potential for error variation in the grading of assessment.
- The major recommendation arising out of this research is the importance of ensuring that the process of assessment is adequately resourced. Markers must be given the support and time needed to carry out the process of assessment in a considered manner. This must involve ensuring that they have a full understanding of the assessment criteria to be applied and the outcomes students are required to show competence in. This ought to be supplemented by a moderation process in which student performance is subject to considered analysis and debate by members of the marking team.
- The work of the Quality Assurance Agency ('QAA') in devoting a section of the Code of Practice on Higher Education to assessment, and the work of the law benchmark panel in providing guidance as to the standards expected of students studying on undergraduate degrees in law provide, essential guidance for marking teams. The research raises questions as to whether this guidance goes far enough to ensure that error variation in marking of assessments is minimised.
- There are real benefits in terms of the development of common understanding across Higher Education Institution ('HEI') law schools through the continued funding by the UKCLE of a series of cross sector grading exercises.

3. Introduction

The idea for a comparative marking project was taken up in January 2000, when colleagues from four universities (Wolverhampton, London South Bank, Sheffield and Sheffield Hallam) decided to test and challenge some of the assumptions generally held by the academic community about marking.

The conventional wisdom is that A level marking, coupled with the moderation process involved, represents a strictly controlled "high water mark" of assessment. Generally an A level student knows that he or she will be judged on a national standard so that it does not matter whether the student sits an exam in Birmingham, Huddersfield or Southampton. A grade 'C' is the same standard throughout the country.

However, the marking processes are normally somewhat different when that pupil moves to a HEI with degree awarding powers. Each department in each university marks coursework, exams and other forms of summative assessment according to its own standards, albeit in the context of subject benchmarking and (for years 2 and 3 of the full-time undergraduate programme – typically referred to as levels 2 and 3) subject to the supervision of external examiners. This autonomy inevitably raises concerns that the processes within any HEI may not be sufficiently transparent, notwithstanding the QAA's Code of Practice on Assessment.

There were three main concerns that the research project wished to explore:

1. The widely held belief that HEI's operated on a shared understanding of marking standards, confirmed by the oversight of external examiners, and underpinned by the law subject benchmark statement.
2. The old adage of "I know a 2:1 when I see one" – i.e. that shared values within the HEI sector could be relied upon to justify an intuitive approach to assessment grading.
3. Whether there are any differences in assessment grading standards between the "old" and the "new" universities.

The research project also sought to address the question of whether there was any (or much) difference between the grades awarded by markers who marked intuitively (i.e. simply by reference to what they believed to be the 'right' mark) and those who were instructed to grade assessments with reference to some form of marking criteria or outline answer.

One of the reasons for undertaking this project was the recognition that assessment is of fundamental importance to the student experience. The consistent grading of student work is an essential component of the assessment process. It is important because it provides assurance of the equitable treatment of students. However, that in itself is not a sufficient measure of validity. It needs to be accompanied by evidence that the academic standard is "set and maintained at the appropriate level and that

student performance is properly judged against this" (QAA, General Principles: Code of Practice on Assessment). There is a need to ensure confidence in the assessment process so as to minimise inconsistencies between markers in the same department and markers of the same module.

We acknowledged that there would be variation in grades resulting from variability in student performance. What should not occur, however, is "error variation". This is due to "the imprecise nature of assessing students'...work" and can take a number of forms (Cuthbert, 2001:19), which are discussed in further detail below. Law schools have sought to minimise such error through the use of agreed assessment criteria and the process of double marking and moderation. These strategies are now recognised by the QAA (Precepts 7 and 8 Code of Practice on Assessment). The presence of an external examiner provides a "back-stop" position in the avoidance of error variation. However, what does not formally occur is the identification of variation between law schools. The research project also recognises that there is now a greater need for transparency and the provisions of the Data Protection legislation make it possible for student to see their marked work. This may lead to more complaints by students and the possibility that unresolved complaints may lead to litigation against a HEI.

4. Background: Assessment and Grading – an Overview

This section of the report sets out the background to the research project and describes the hypothesis the research project team sought to test. Assessment is arguably the single most important element of the student learning experience. An assessment task and accompanying submission deadline influences not just how a student perceives the module/unit to which they relate, but can impact on others areas of study and extra university fields of activity. As Gibbs (1992) has noted:

"... students will be powerfully influenced by the assessment system they are working within. ... (T)he criteria used in allocating marks can have a dramatic effect on the way students approach both assessed assignments and other learning activities."

For whatever purpose it is deployed, assessment generates large amounts of work for both students and those responsible for the setting and evaluating the multiple forms that it can take. In all its forms, whether used for diagnostic, formative or summative purposes it is of key importance to the process of higher education. The centrality of its importance has been well stated by the Quality Assurance Agency (2000, para 7, Section 6):

"Assessment plays a significant role in the learning experience of students. It determines their progression through their programmes and enables them to demonstrate that they have achieved the intended learning outcomes. It is assessment that provides the main basis for public recognition of achievement, through the awarding of qualifications and/or credit."

Assessment decisions label those to whom they are applied and can affect their ability to achieve career goals and secure progression through the educational process. Consequently the grading of assessed work is subject to intense scrutiny and can generate debate and dispute, which can end in litigation. It is, therefore, important that the integrity of the assessment process is beyond question.

Key to this is the consistency of assessment gradings. It is assumed that all HEI's operate quality assurance systems as a means of ensuring the integrity of their assessment processes. That there will be variation in student performance is accepted. What should not occur is variation in the application of grades to work of the same standard. Module/course teams have commonly sought to ensure the consistency of the assessment process through the use of assessment criteria and the process of moderation. The oversight provided by external examiners providing an additional tier in this process. These, along with other stages in the assessment process, are safeguards against what has been referred to as 'error variation'.

A question that is commonly asked is the extent to which there can be confidence in the comparability of assessment decisions made within different institutions. External examiners are commonly asked to report on the extent to which standards applied within an individual law school are equivalent to that applied within the HEI sector.

There is evidence that external examiners in law do share and utilise such an understanding. A report (1997) conducted on behalf of the Association of Law Teachers, Committee of Heads of University Law Schools and the Society of Public Teachers of Law, referred to research based on a moderation exercise, involving 20 experienced external examiners, that there was broad consensus as to grading standards. The same report also referred back to the Graduate Standards Report emanating from the Graduate Standards Project commissioned by the Higher Education Quality Council and the belief that:

“... the old processes by which internal examiners learnt the standards of their institution by osmosis are rapidly disappearing”

Extensive checks and balances, including the intervention of independent external examiners, operate within individual law schools as a means of trying to ensure the consistent application of agreed standards. However, the hypothesis this project seeks to evaluate is whether there can be confidence in the cross-institutional application of standards. Put simply, would the examiners of one institution agree with those of other institutions in respect of the grading of the same piece of student work? External examiners are commonly asked to formally record whether the standards applied in one law school are comparable with those applied across the sector. Their response is quite rightly taken as a key measure of the integrity of the standards operated by an individual law school. That there is evidence that external examiners do share common standards provides some reassurance, but to what extent is that shared by teachers in different institutions? Not all student work is subject to scrutiny by external examiners. It is, for instance, not uncommon for level 1

(i.e. the first year of the undergraduate programme) assessments to be subject to a less intense level of scrutiny by external examiners as compared to that they are required to devote to level 2 and 3 work.

That such work is subject to the same assessment and moderation practices applicable to student work subject to a higher level of external scrutiny provides tangential evidence of a shared understanding of standards across all levels. Furthermore, law in Higher Education is delivered in an environment where the student learning experience is increasingly defined by learning outcomes, external benchmarks, and programme specifications far advanced from the vague and loosely defined syllabus of previous practice. In addition most HEI's deploy assessment criteria drafted by academic teams and accessible by the students whose work will be tested against them. The development and deployment of these guides and measures should guarantee that there is consistency of practice on an inter and extra institutional basis.

They would seem to have a particular role in helping new entrants to the profession assimilate the standards that underpin the delivery of undergraduate legal education. As Ecclestone (2001: 312) has noted:

“... assessment guidelines and moderation processes appear, ... to play an important part in training assessors, particularly by helping novice assessors become part of an assessment community more quickly.”

Just as importantly as Ecclestone observes “... they provide an explicit reminder of the agreed public standards about fairness and rigour.” (ibid)

The importance of assessment criteria and the desirability of the social dimension of moderation practices are factors that this research bears out. They are essential tools in mediating the effects of what Ecclestone refers to as the 'mental model' of equality, which markers tend to apply irrespective of other sources of guidance.

The team that conducted the research that is the subject of this report undertook a study of the comparability of the mental mode of equality of law teachers drawn from a number of institutions. Of equal importance was the desire to generate a debate about such matters within HEIs involved in the teaching of law.

5. Methodology and description of the research

The research commenced with a series of grading exercises, including a pilot study, intended to stimulate debate about the issue of comparability of standards, and the extent to which there is, amongst law teachers, an intuitive understanding of the applicable standards. These exercises involved teachers drawn from a variety of institutions grading work both intuitively and by reference to assessment criteria. It concluded with a marking exercise, which had two fields of enquiry. The first focussed closely on the understanding and

use of assessment criteria and the social dimension of the moderation process. The second negated the latter and involved the grading by law teachers drawn from different institutions of a common set of answers. At different stages in the research the team sought to replicate or ameliorate the pressures that markers are commonly required to work under as part of the assessment process (i.e. control the amount of time allowed for various stages of the mock assessment processes).

The resultant data was subject to standard statistical analysis in particular that used to detect error variation amongst participating markers. As stated above, variation will always be found in the grades allocated to work submitted for assessment as a result of the differing levels of understanding and application demonstrated by the student body. Secondly, there will be what is referred to as 'error variation'. As Cuthbert (2001:17) has stated, error variation can occur in three ways:

1. When those responsible for grading differentiate consistently between assessments submitted by students, but work to different means.
2. While graders have a similar mean there is a difference in the range of grades each applies; and/or
3. When graders make errors and arrive at an incorrect or inappropriate grade.

Statistical analysis was used to determine the extent to which error variation of the first two types was to be found in the grading of participants in the grading exercises.

In addition, given the contemporary importance attached to assessment criteria and guides and measures a study was undertaken of the use of such tools in the institutions of the core research team. This in the main took the form of a review of the literature developed by these institutions to support the assessment process.

6. Institutional Literature Review

A review of the literature used to support the assessment process in the institutions of the core research team showed diversity of practice.

All but one law school used learning outcomes to define what was expected of both students and the teaching team. All utilised assessment criteria in grading student work but only two published generic criteria of the type found in the benchmark statement on law to students. Similarly, while all four utilised module specific assessment criteria, only two made these publicly available. Three of the institutions published module guides for use by students, but only two had, at the time of the research, published an assessment handbook, containing written guidance for teaching staff asked to set assessments. Two of the four required the production of a report on the performance of students by the unit leader reflecting a basic statistical analysis of grades. Significantly, none of the four law schools carried out statistical analysis to determine the existence or extent of error variation.

Given the limited range of HEI's in the sample it would be inappropriate to attach too great a significance to these findings. However, given the importance attached to the use of such documentation as a means of developing and supporting common understanding of academic standards further research is justified.

7. Ethical parameters

The project team reached agreement that the research should be undertaken in accordance with commonly agreed standards of good practice, such as those laid down in the Ethical Statements of our respective institutions. These fundamental and widely accepted principles may broadly be categorised as:

- beneficence
- non-maleficence
- confidentiality/anonymity

Based on the above, the guiding principles underpinning the conduct of the research were as follows:

- That the research should conform to all legal requirements. This included compliance with relevant data protection legislation.
- That the research should be scientifically sound and the purpose should be to contribute to the development of improved pedagogic practices.
- That the research should be undertaken by those who were appropriately qualified and experienced.
- That any details allowing individuals to be potentially identified should not be published, or made available, to anyone not directly involved in the research, and that all reasonable steps should be taken to ensure that confidential details are secure.

To implement these guiding principles we decided that good practice would encompass: maintaining professional standards, documenting results, questioning one's own findings, attributing honestly the contributions of others, ensuring the security and storing of primary data and compliance with relevant HEI ethical policies.

Although the names of the researchers and their respective institutions are in the public domain, all the data from those institutions has been made anonymous. In the marking sessions organised as part of the research, the names of colleagues who took part have been kept secret and their contributions anonymised. All recorded data is completely anonymous and the original work on which the data was based has been destroyed by shredding. It is not possible to identify the originator of any of the data. Ethical clearance was obtained from participating HEI's as appropriate.

8. The relevant QAA precepts

The very existence of national subject benchmarks in undergraduate law programmes (QAA, May 2000) assumes a reliable means of assessing whether or not the benchmark statements have been attained by a student. To this end the QAA has promulgated a number of precepts regarding the assessment process (2000, Section 4: External Examining; Section 5: Academic Appeals etc; Section 6: Assessment of students). In general terms these precepts expect academic institutions to be able to demonstrate effective procedures for ensuring

‘...the consistent implementation of rigorous assessment practices which ensure that the academic/professional standard for each award and award element is set and maintained at the appropriate level and that student performance is properly judged against this. (Section 6: Assessment of Students Precepts and Guidance: General Principles 1)

The precept goes on to provide that ‘The principles, procedures and processes of all assessment should be explicit, valid, and reliable.’ (Section 6: Assessment of Students Precepts and Guidance: General Principles 2)

Note the inclusion of the word ‘reliable’. Not only are institutions required to demonstrate how the reliability of assessment is ensured (for example, the consistent use of agreed marking and grading schemes, and moderation arrangements), they are also under an obligation to establish the robustness of arrangements to monitor, evaluate and demonstrate the fairness of assessments. The precept goes on to place the onus firmly on awarding institutions to ‘.....publish, and implement consistently, clear criteria for the marking and grading of assessments. (Marking and Grading: 7), and to ‘.....ensure that there are robust mechanisms for marking and for the moderation of marks.’ (Marking and Grading: 8)

In relation to marking criteria, Appendix 2 to the Code on assessment refers to the type of information that institutions should consider including in their published documentation on assessment, in particular ‘.....the criteria for assessment including, where appropriate, descriptors of expected standards of student attainment: what is expected in order to pass or to gain a particular grade or classification; (Section 6: Assessment of Students Appendix 2: A guidance note on published assessment information)

It is hard to see how a largely intuitive approach to the grading of assessed work would satisfy any of these demands. The precept points quite clearly to certain best practice aspects of the marking process including second marking ‘*where appropriate*’, ensuring the reliability of ‘.....methods used for the sampling of assessments from larger groups..’; the rules relating to internal moderation of marks; and the need to undertake ‘....an analysis of marking and marking trends to facilitate comparisons and provide evidence on standards.’ (Section 6: Assessment of Students Appendix 2: A guidance note on published assessment information).

The extent to which the role of the external examiner can provide a safeguard against capricious internal marking may depend on the role identified for the external examiner by the awarding institution. Section 4 of the Code of practice for the assurance of academic quality and standards in higher education, which deals with external examining, leaves it to institutions to determine the roles, powers and responsibilities of external examiners, including the extent of the external examiner's authority in respect of assessment boards. External examiners are to be given advice:

'...about how they are expected to undertake their responsibilities in evaluating the level of student achievement, for example, by approving examination papers, sampling students' work and generally moderating and validating marks.'

In practice, however, many external examiners are now confined to commenting on whether the overall level of marking, based on a sample of scripts, is satisfactory. The role does not extend to correcting individual marks. Clearly, if a sample of scripts seen by an external examiner contains evidence of aberrant grading, he or she can call for all the marks to be reviewed. If it does not, rogue grades can go uncorrected.

9. The Pilot Study

The starting point for the research team was the hypothesis that there might be significant variations between markers in old and new universities in the awarding of credit for the same piece of student work. The next stage was to consider how the hypothesis should and could be tested. The decision was made to concentrate on a subject taught at the same level in all participating HEI's, reducing the number of potential subjects to foundational modules, most of which were taught in the first year (i.e. level 1). After comparing the modules delivered it was decided to concentrate on what generally has the appellation of 'English Legal System' (ELS) but which goes under several different titles such as 'Legal Method', 'Legal Foundations', and 'Legal System and Institutions'. It was hoped that when the research widened, anyone teaching on an Anglo-Welsh law degree could join in the project. The content of such modules varied widely, a matter commented on elsewhere, and the modes of assessment also differed. To take the latter point, the unit might be wholly examined, half-examined and half assessed by coursework, and at one institution 20% of the marks were provided by multiple choice questions and 80% by a mini-dissertation.

9.1 Assessor Group 1: Exercise 1

The method of proceeding was to select topics within the modules/units delivering ELS, which were dealt with at all institutions. Five pieces of coursework and 12 exam scripts were selected. All internal marks and comments were deleted so that the assessors participating in exercise 1 had 'clean' scripts to work with. Each of the four assessors comprising Assessor Group 1 marked the same 5 pieces of coursework and the same 12 examination scripts. The essays had been submitted by Year 1 FT LLB ELS

students. The pass mark for the essay question and the exam questions was 40%. This was made clear to members of Assessor Group 1, and they were also supplied with module statements (including module/unit learning outcomes). Marking criteria were not used. The members of Assessor Group 1 were not advised of the mark actually awarded to the candidate by the internal marker until the completion of the marking exercise.

9.1.1 Assessor Group 1: Exercise 1(a) – course works

The results in respect of the 5 pieces of coursework are set out in **Table 1(a)**.

Table 1(a)							
Assessor Group 1							
	CW 1	CW 2	CW 3	CW 4	CW 5	Average marks for this marker	Range for this marker
Marker 1	35.0	80.0	57.0	20.0	58.0	50.0	60.0
Marker 2	46.0	68.0	58.0	35.0	73.0	56.0	38.0
Marker 3	35.0	80.0	56.0	42.0	60.0	54.6	40.0
Marker 4	40.0	72.0	58.0	30.0	58.0	51.6	42.0
Average	39.0	75.0	57.3	31.8	62.3	53.1	43.0
Range for this course work	11.0	12.0	2.0	22.0	15.0	6.0	
Original grade	45.0	75.0	69.0	35.0	58.0	56.4	40.0
Deviation between Assessor Group 1 average and internal mark	-6.0	0.0	-11.8	-3.3	4.3	-3.4	

The average of the marks awarded by each member of Assessor Group 1 shows a fairly tight grouping – suggesting a consensus amongst the group. Looking at the range of marks awarded by each marker for each course work, however, reveals some significant disparities – especially for CW 4 where the highest grade is 42 and the lowest is 20. CW 3 is marked as a high 2:2 by all four members of Assessor Group 1, compared to an internal marker grade of 69. Assessor Group 1 produced an average mark of 39 for CW 1 whereas the internal marker had awarded 45.

9.1.2 Assessor Group 1: Exercise 1(b) – exam scripts

For the exam script marking exercise markers were asked to consider scripts containing two answers, the aggregate mark for both answers producing the overall exam mark. Each question carried the same maximum mark. The results in respect of the marks awarded by Assessor Group 1 for the 12 examination scripts are set out in Table 1(b), along with the original mark

awarded by the internal marker and the deviation between the average of the marks awarded by Assessor Group 1 and the original mark.

Table 1(b)													
Assessor Group 1													
Exam Paper	A	B	C	D	E	F	G	H	I	J	K	L	Ave
Marker 1 Q1	54	38	35	5	56	49	54	58	66	60	59	50	48.7
Marker 1 Q2	40	38	25	35	65	60	65	63	54	53	58	49	50.5
Marker 1 Exam Total	47	38	30	20	60.5	54.5	59.5	60.5	60	56.5	58.5	49.5	49.5
Marker 2 Q1	53	46	34	10	13	47	60	55	58	52	57	58	45.3
Marker 2 Q2	55	45	24	54	66	58	63	47	58	55	56	43	52.8
Marker 2 Exam Total	54	46	29	32	39.5	52.5	61.5	51	58	53.5	56.5	50.5	48.5
Marker 3 Q1	50	48	30	15	40	38	57	60	58	56	55	53	46.7
Marker 3 Q2	48	44	31	0	57	72	69	61	57	49	58	46	49.6
Marker 3 Exam Total	49	46	30.5	7.5	48.5	55	63	60.5	57.5	52.5	56.5	49.5	47.9
Marker 4 Q1	52	40	45	0	38	52	56	54	54	52	58	50	45.9
Marker 4 Q2	56	60	50	50	65	70	68	54	66	55	64	50	59.8
Marker 4 Exam Total	54	50	47.5	25	51.5	61	62	54	60	53.5	61	50	52.7
Ave of Assessor Group 1 marks for Q1	52.3	43	36	7.5	36.8	46.5	56.8	56.8	59	55	57.3	52.3	46.6
Ave of Assessor Group 1	49.8	46.8	32.5	34.8	63.3	65	66.3	56.3	58.8	53	59	47	53.2

Table 1(b)														
Assessor Group 1														
Exam Paper	A	B	C	D	E	F	G	H	I	J	K	L	Ave	
marks for Q2														
Ave total exam mark for Assessor Group 1	51	44.9	34.3	21.1	50	55.8	61.5	56.5	58.9	54	58.1	49.9	49.6	
Internal original grade for Q1	48	40	34	10	31	36	54	61	61	66	72	53	47.1	
Internal original grade for Q2	46	40	34	60	53	58	58	57	53	60	64	49	53	
Internal original exam total	47	40	34	35	42	47	56	59	57	63	68	51	49.8	
Deviation Q1	4.3	3	2	-2.5	5.8	10.5	2.8	-4.3	-2	-11	-14.8	-0.3	-0.5	
Deviation Q2	3.8	6.8	-1.5	-25.3	10.3	7	8.3	-0.8	5.8	-7	-5	-2	0.2	
Deviation Total														
	4.0	4.9	0.3	-13.9	8.0	8.8	5.5	-2.5	1.9	-9.0	-9.9	-1.1	-0.2	

Candidates submitting papers B and C were awarded failing grades by the internal marker. The averages of the grades awarded by Assessor Group 1 shows the same two candidates failing, although Candidate D is a much clearer fail.

9.2 Assessor Group 2: Exercise 2

The outcomes of Exercise 1 convinced the team that the instance of error variation in assessment grading merited further investigation. For the next stage, reported as 'Exercise 2' below, a team of law lecturers with responsibility for teaching ELS were invited to attend a marking workshop. These academics (collectively referred to here as Assessor Group 2) were drawn from range of HEI's, some were novice markers, some experienced, and included Markers 1 to 4 from Assessor Group 1 (who had not been

reminded of their original grades), and the marker who had awarded the original grades for the course works and exam scripts used in Exercise 1. Markers in Assessor Group 2 were invited to mark four of the five course works that had been used in Exercise 1 and exactly the same examination scripts.

9.2.1 Assessor Group 2: Exercise 2(a) – course works

The results of the course work remarking exercise are set out in Table 2(a). Column headed 'Ave in 1(a)' shows the average of the marks awarded by each of Markers 1 to 4 (Assessor Group 1) for these course works when they marked them as part of Exercise 1. The table also shows the deviation between the average of the marks awarded in Exercise 1 and Exercise 2.

Table 2(a)					
Assessor Group 2					
	CW 1	CW 2	CW 3	CW 5	Ave
Marker					
Marker 1	48.0	68.0	66.0	56.0	59.5
Marker 2	35.0	70.0	58.0	53.0	54.0
Marker 3	42.0	70.0	55.0	67.0	58.5
Marker 4	40.0	74.0	37.0	55.0	51.5
Marker 5	45.0	65.0	42.0	46.0	49.5
Marker 6	42.0	60.0	54.0	53.0	52.3
Marker 7	40.0	65.0	50.0	52.0	51.8
Marker 8	42.0	75.0	48.0	53.0	54.5
Marker 9	48.0	70.0	59.0	60.0	59.3
Average	42.4	68.6	51.6	57.8	55.1
Original internal grade	45.0	75.0	69.0	58.0	61.8
Deviation of average from internal	-2.6	-6.4	-17.4	-0.3	-2.6

Again CW 3 shows a significant variation between the original internal mark and the average mark awarded by Assessor Group 2. Note that Marker 1 has increased the average marks awarded for these course works by over 10% between Exercise 1 and Exercise 2.

9.2.2. Assessor Group 2: Exercise 2(b) – exam scripts

Assessor Group 2 members were supplied with the exam scripts used in Exercise 1(b). As with the Assessor Group 1 exercise, assessors were briefed as to the pass mark and supplied with module statements (including outcomes). The markers were not advised of the marks awarded by the internal marker until the completion of the marking exercise. The results are set out in Table 2(b). Note that the exercise sought to replicate the time pressures of marking during an examination period, with the result that some of the less experienced markers in Assessor Group 2 were unable to complete the marking of all scripts – where this is the case the designation 'N/A' is used and the results have been weighted to reflect the missing marks.

For each marker the table indicates the marks awarded for each question on the exam script and the total marks awarded for each script.

Table 2(b)													
Assessor Group 2													
Paper	A	B	C	D	E	F	G	H	I	J	K	L	Ave
Marker 1 Q1	46	35	30	5	50	40	55	68	70	57	60	50	47
Marker 1 Q2	48	49	32	30	67	55	70	50	54	55	65	56	52
Marker 1 Total	47	42	31	18	59	48	63	59	62	56	63	53	50
Marker 2 Q1	50	44	37	7	20	45	45	57	57	62	54	48	44
Marker 2 Q2	50	47	30	46	55	52	55	45	57	58	48	48	49
Marker 2 Total	50	46	34	27	38	49	50	51	57	60	51	48	46
Marker 3 Q1	50	49	44	12	40	45	55	54	N/A	51	50	50	45
Marker 3 Q2	59	52	43	0	57	48	61	51	N/A	44	49	50	46
Marker 3 Total	55	51	44	6	49	47	58	53	N/A	48	50	50	46
Marker 4 Q1	54	40	45	20	45	62	54	58	55	N/A	N/A	N/A	48
Marker 4 Q2	50	46	50	10	56	64	70	50	50	N/A	N/A	N/A	50
Marker 4 Total	52	43	48	15	51	63	62	54	53	N/A	N/A	N/A	49
Marker 5 Q1	43	34	47	N/A	47	51	53	55	60	N/A	N/A	N/A	49
Marker 5 Q2	42	37	48	N/A	55	57	58	50	N/A	N/A	N/A	N/A	50
Marker 5 total	43	36	48	N/A	51	54	56	53	60	N/A	N/A	N/A	50
Marker 6 Q1	42	38	38	7	40	38	56	57	55	57	63	53	45
Marker 6 Q2	44	42	40	50	56	56	60	53	62	62	64	50	54
Marker 6 Total	43	40	39	29	48	47	58	55	59	60	64	52	49
Marker 7 Q1	50	48	38	35	50	38	N/A	N/A	N/A	N/A	N/A	N/A	43
Marker 7 Q2	50	48	37	35	45	55	N/A	N/A	N/A	N/A	N/A	N/A	45
Marker 7 Total	50	48	38	35	48	47	N/A	N/A	N/A	N/A	N/A	N/A	44
Marker 8 Q1	40	38	35	5	37	48	44	48	55	45	42	44	40
Marker 8 Q2	45	43	40	40	52	55	58	43	58	55	47	44	49
Marker 8 Total	43	41	38	23	45	52	51	46	57	50	45	44	44
Marker 9 Q1	45	40	40	6	56	62	49	67	60	48	58	54	49
Marker 9 Q2	53	45	37	5	63	48	64	55	56	57	60	57	49
Marker 9 Total	49	45	40	31	56	54	62	50	56	55	56	51	50
Ave of Assessor Group 2 marks for Q1	48	43	40	21	50	51	57	54	58	54	55	50	48
Ave of Assessor Group 2 marks for Q2	48	40	34	10	31	36	54	61	61	66	72	53	47
Ave total exam mark for Assessor Group 2	46	40	34	60	53	58	58	57	53	60	64	49	53
Internal Q1	48	40	34	10	31	36	54	61	61	66	72	53	47
Internal Q2	46	40	34	60	53	58	58	57	53	60	64	49	53

Table 2(b)													
Assessor Group 2													
Paper	A	B	C	D	E	F	G	H	I	J	K	L	Ave
Internal exam total	47	40	34	35	42	47	56	59	57	63	68	51	50
Deviation Q1	-1	1	5	2	12	12	-3	-3	-2	-13	-18	-3	-1
Deviation Q2	3	5	6	-29	3	-4	4	-7	3	-5	-9	2	-3
Deviation exam Total	1	3	6	-14	8	4	1	-5	1	-9	-13	-1	-2

Using the same student work for both exercises provided the opportunity not only to compare the marks awarded by Assessor Group 1 with those awarded by the additional markers comprising Assessor Group 2, but also provided the opportunity to examine the variations between the marks awarded by the original internal marker and markers 1 to 4 from Assessor Group 1 in Exercise 1 with the marks that they awarded for the same pieces of work in Exercise 2. See Appendix 1 for Table 2(c), which compares course work marks from Exercises 1 and 2, and Table 2(d), which compares the exam script marks from Exercises 1 and 2.

None of the markers showed any very significant deviations in the marks awarded to the course works between Exercises 1 and 2; Marker 2 showed the greatest overall shift, awarding on average 7% less the second time these course works were marked. These averages, however, mask large variations between the first and second markings of certain course works. Marker 4 awarded 58% to course work 3 in Exercise 1, and 37% in Exercise 2. The original internal marker shifted from the mark of 69% awarded to course work 3 to 50%. In all there are 8 instances here where the deviation between a marker's first grading and second grading of the same course work is 10% or more.

There was greater consistency between the first marking and remarking of the exam scripts. In total there were only 3 instances across 60 remarked exam scripts (12 per scripts, 5 markers) where the variation between the grade awarded for an exam paper as a whole in Exercise 1 deviated from that in Exercise 2 by 10% or more.

The views of the researchers at this point, i.e. at the pilot stage (bearing in mind the relatively limited numbers of markers and the numbers of scripts), may be summarised thus:

- there was no marked difference between markers from old and new universities in the level of marks awarded, the original hypothesis;
- there was no difference between the genders;
- there was no difference as to whether the markers were old or young;

- there was no difference as to whether the markers were inexperienced or not;
- there was some difference between marks of the same examiners after a gap of time;
- there was evidence that careful and reasonable markers given the same guidance and the same script could produce differing grades.

It was this last point that prompted the research project team to extend the scope of the inquiry further. It was recognised, however, that there were several evident drawbacks with the research so far. First, the module looked at was only one of 18 or so undertaken on a Law degree; secondly, whilst there was a 'control' in the sense that we had the internal marks for comparison, there was no 'absolute' mark; thirdly, the institutions' assessment policies differed. Furthermore, the coverage of subjects such as ELS varied widely, for example the interaction of common law and equity may be taught at one institution but not at another. It was therefore difficult to make true comparisons. Another difficulty was that at some institutions modules were taught throughout the year with one assessment event in May or June ('long thin' modules), whereas others were taught in a modularised and semesterised scheme whereby assessment took place at the end of each semester ('short fat' modules).

10. Broadening the Hypothesis

The researchers determined that the project should proceed. Interesting issues were coming to the fore, even though at that time it was clear on the evidence to hand that the suspected differences between marking at old and new universities did not exist. Because of the wide variation in the practice of assessment it was decided to widen the scope of the enquiry and to look at another foundational law module, Criminal Law. There were several matters favouring this choice: ELS was a first year (level 1) module whereas at all the participating institutions Criminal Law was a second year (level 2) module (and thus subject to greater external scrutiny); the team had expertise in Criminal Law (two were writers of student textbooks and had many years' marking experience in the subject at several universities as internal markers, moderators, and external examiners); and markers could be easily recruited into the project.

The decision was taken to formulate a criminal law problem question that could be included on a level 2 Criminal Law paper at 3 of the participating HEIs – the question is set out in Appendix 2. The first cohort of students attempted the question at the end of semester 1 in January 2004. Students at the two remaining HEIs attempted the paper containing the question at examinations held in June 2004. The resulting student exam answers provided the material for Exercises 3, 4 and 5, detailed below.

11. An exploration of intuitive marking: Exercise 3 – Assessor Group 3 York ALT

Assessor Group 3 comprised 38 volunteers attending the Comparative Marking Workshop held at the Association of Law Teachers' (ALT) Conference, 2004. Members of Assessor Group 3 were self-selecting, comprising Criminal Law teachers, law teachers who had never taught criminal law, teachers from pre and post 1992 HEIs, and some law teachers from overseas institutions. Members of this group were provided with copies of answers to two questions (Paper 'A' and Paper 'B') set on a Criminal Law examination at a pre-1992 HEI. A briefing was provided (Appendix 3), but no marking criteria were used (as was the case with the original internal marking of the work). Members of Assessor Group 3 completed a questionnaire providing information on experience of marking Criminal Law work and nature of their 'home' HEI (see Appendix 3). In addition to identifying the most appropriate grade, members of the group were asked to indicate the lowest and highest marks they would be prepared to agree to for each of the two questions. The results are set out in Tables 3(a) to 3(f). For each exam answer the internally moderated mark awarded students at the participating HEI was 56%.

11.1 Exercise 3: Assessor Group 3

11.1.1 Results of all Assessor Group 3 markers for Papers A & B

Data in Tables 3(a) and 3(b) (set out in full in Appendix 4) shows the range of marks (lowest, ideal and highest) members of Assessor Group 3 would have been prepared to award for Paper A, the average of these three marks, and the range of marks (i.e. the difference between the lowest and highest grade). Table 3(b) similarly shows the results in relation to Paper B.

For Paper A the average 'ideal' grade came out at 53.2% - reassuringly close to the internal marker's 56%. Further, 28 of the 38 markers awarded an ideal grade within 6% (i.e. between 50% and 62%) of the internal grade. 15 of the ideal grades were in the same degree classification as the internal mark. More worrying, however, is the data on the range of marks. The ideal mark range runs from 25% to 73%; the lowest mark range runs from 20% to 60% and the highest mark range runs from 25% to 71%. For Paper B the average ideal grade comes out at 50.3% - still in the same degree classification band as the internal mark, but only just. Again, 15 markers placed Paper B in the same degree classification as the internal marker, but only 21 markers agreed on an ideal mark within 6% either side of the internal mark.

The Research Project team had assumed, given the diversity of the membership of Assessor Group 3 that it would be necessary to disaggregate the results from Exercise 3, in order to consider error deviation amongst significant sub-groupings, such as criminal law lecturers, non-criminal law lecturers, and pre and post –1992 HEI staff. This data is set out in Tables 3(c) to 3(f).

11.1.2 Exercise 3: Results extracted for Criminal law markers only

Table 3(c)	Paper A				
Criminal	Law	Markers	Only		
	Lowest Grade for Paper A	Ideal grade for Paper A	Highest grade for Paper A	Average of range	Range
	30	38	42	36.7	12
	36	40	44	40.0	8
	44	47	50	47.0	6
	43	47	51	47.0	8
	49	53	58	53.3	9
	54	57	59	56.7	5
	55	62	65	60.7	10
	62	64	66	64.0	4
	60	65	68	64.3	8
	60	65	70	65.0	10
Averages	49.3	53.8	57.3	53.5	8

Table 3(c)	Paper B				
Criminal	Law	Markers	Only		
Marker	Lowest Grade for Paper B	Ideal grade for Paper B	Highest grade for Paper B	Average of range	Range
	30	35	39	34.7	9
	30	36	40	35.3	10
	40	40	40	40.0	0
	43	40	46	43.0	3
	38	44	48	43.3	10
	42	45	47	44.7	5
	48	50	52	50.0	4
	47	50	55	50.7	8
	48	55	58	53.7	10
	48	55	59	54.0	11
Averages	41.4	45	48.4	44.9	7

The Research Project team had expected this sub-set of Assessor Group 3 to produce the grouping of marks closest to the internal grade and with the smallest variation between highest and lowest grades. Although the overall average ideal grade for Paper A is 53.8%, close enough to 56% to suggest that teachers within the same subject grouping can intuitively identify the

appropriate grade band for a piece of work within their expertise, it is again the range of marks offered by this sub-set of the Assessor Group that causes concern. The lowest ideal grade suggested for Paper A is a borderline fail at 39%. The highest is a solid 2:1. Only 2 of the Criminal Law markers comprising this sub-set of Assessor Group 3 gave an ideal mark in the same degree classification as the internal marker. The range, from the lowest mark any member of this sub-set of markers would have awarded and the highest, is 40%. As might be expected from looking at the results for all members of Assessor Group 3 in relation to Paper B, the Criminal Law teacher sub-set arrived at a lower average ideal mark for Paper B (45%). This indicates a significant deviation from the internal marker's grade of 56%, the range of ideal marks this time encompassing a clear fail at 35% and a solid 2:2 at 55%. Only 2 members of the Criminal Law teacher sub-set arrived at a grade for Paper B that was within 5% of the internal grade.

It was the data set out in Tables 3(c) and 3(d) in particular that informed the next stage of the Research Project – criteria referenced marking for experienced Criminal Law teachers – considered further below.

11.1.3 Exercise 3: Results extracted for non-criminal law markers only

Disaggregating the data to isolate the grades awarded by the non-criminal law teachers for Papers A and B (set out in full in Tables 3(d) and 3(e) at Appendix 5) revealed, contrary to the project team's expectations, that the non-criminal law markers, overall, arrived at average ideal grades closer to the internal marks for both Papers A and B than was the case with the Criminal Law teacher sub-set. The range of ideal marks, however, showed huge variations; 35% to 70% for Paper A, and 25% to 73% for Paper B. Interestingly, 14 markers in the non-criminal law teacher sub-set awarded an ideal mark in the same degree classification band as the internal marker for Paper A, and 11 did so for Paper B. 19 of these markers awarded an ideal grade within 6% either way of the internal mark for Paper A and 17 did so in respect of Paper B. This could be seen as providing support for the view that, even without expert subject knowledge, assessors familiar with HEI standards for degree level work are able to recognise certain factors as being common to all student work falling within a given degree classification – such as structure of argument, clarity of expression and reasoning, reference to and use of authorities in supporting argument etc.

11.2 Exercise 3 - Summary of grades awarded by reference to marker HEI background

It will be recalled that part of the original hypothesis was the question of whether there might be any significant difference in the way identical pieces of student work were graded by assessors from pre and post 1992 Universities. Exercise 3 provided a limited opportunity to test this, given that members of Assessor Group 3 were asked to identify the nature of their 'home' institution. The results of this disaggregation are set out in Table 3(f), below.

Table 3(f)	Exercise 3				
	Pre 92 HEI Marker Average Low	Pre 92 HEI Marker Average Ideal	Pre 92 HEI Marker Average high	Pre 92 HEI Marker Average of Range	Pre 92 HEI Marker Range
Paper A	40.70	42.30	47.00	43.33	6.30
Paper B	44.00	49.30	53.30	48.87	9.30
Average for both A and B	42.35	45.80	50.15	46.10	7.80
	Post 92 HEI Marker Average Low	Post 92 HEI Marker Average Ideal	Post 92 HEI Marker Average high	Post 92 HEI Marker Average of Range	Post 92 HEI Marker Range
Paper A	49.33	54.78	58.13	54.08	8.80
Paper B	45.56	49.80	53.33	49.56	7.77
Average for both A and B	47.45	52.29	55.73	51.82	8.29
	Other 92 HEI Marker Average Low	Other 92 HEI Marker Average Ideal	Other 92 HEI Marker Average high	Other 92 HEI Marker Average of Range	Other 92 HEI Marker Range
Paper A	43.80	50.40	54.80	49.67	11.00
Paper B	48.40	53.60	58.00	53.33	9.60
Average for both A and B	46.10	52.00	56.40	51.50	10.30

These figures suggest that pre-1992 HEI markers are less generous in the awarding of credit – although the average for both sectors in respect of the ideal grade for Paper B is remarkably consistent at 49.3% and 49.8%. The 12% disparity in the ideal averages for Paper A is of more concern. The pre-1992 HEI markers regarding the script as, ideally, a poor pass – the post 1992 HEI markers seeing it, ideally, as a safe 2:2. It should be borne in mind that the script itself was that of a candidate sitting a Criminal Law examination at a pre-1992 HEI. Clearly the average ideal marks awarded by the post-1992 HEI markers are far closer to the 56% awarded for both papers by the internal marker.

12. Intuitive and criterion-referenced marking

Following the completion of Exercise 3, the Research Project team decided to concentrate on the issue of whether or not the provision of marking criteria could be relied upon to eliminate the more extreme examples of grade error variation encountered during the research exercises where participants were allowed to adopt a largely intuitive approach to the awarding of credit.

Assessment of student performance can serve a variety of purposes – broadly these tend to fall within three categories. Diagnostic, formative (in the sense that it provides feedback on progress and thus alters learner behaviour), and summative. Of course, assessment can serve any or all of these functions to varying degrees, but it is summative assessment that is likely to have most impact on a student's progression and ultimate qualification. The basic purpose of summative assessment is to arrive at an informed conclusion as to the standard achieved by the student, whether expressed as 'pass/fail', 'competent/not competent', by reference to a particular banding, such as '1st', '2:1' etc, or as a precise percentage. The grade awarded effectively performs the function of a form of freely convertible currency within the academic community – a symbol that fellow academic practitioners understand, and also serves as an endorsement of the student in his or her relationships with other stakeholders (such as employers and providers of postgraduate and vocational stage training) who act in good faith on these exercises in academic judgment.

As indicated, in the introduction to this paper, implicit in the hypothesis underpinning this Research Project is the assumption that a group of assessors employing an intuitive approach to marking the same piece of work will award marks within a wider range compared to a group of assessors required to follow a set of explicit marking criteria.

Adopting an intuitive approach to awarding credit for assessments assumes the existence of widely held and commonly understood norms regarding appropriate grading. The development of these cultural norms regarding the awarding of appropriate credit for assessment can result from academics inheriting a degree of cultural certainty from their own experiences in higher education, whether as students (the marks they were themselves awarded), or as a result of the pressure of academic peer group behaviour in team teaching and moderation exercises. As academics become more experienced they are likely to become more distrustful of constructs that override their instincts as to the grade that a piece of work might deserve. As Kathryn Ecclestone (2001: 924) observes:

'...the transition from being a novice, through stages of "advanced beginner" and "competent" to become "expert", is characterised partly by a declining dependence on rules, routines and explicit deliberation. Experts become more intuitive and less deliberative, and are less able to articulate the tacit knowledge on which much of their decision making has come to depend. They also become more reluctant to

deliberate since deliberation is associated with beginner status and devalued by peer group norms...'

Where a module/unit leader on an LLB core module/unit undertakes the task of marking summative assessments it is unlikely that he or she will be undertaking this task alone. The reality for many is that they will be co-ordinating a team of full-time and hourly paid colleagues, with varying degrees of experience as subject markers. As the data produced by Exercises 1, 2, and 3 (considered above) indicates, relying on the intuition of team members as the basis for awarding credit may well create a risk that there will be gross variations between markers in the awarding of credit.

Ensuring the integrity and credibility of the assessment process therefore necessitates the development of some form of marking scheme to ensure consistency of approach.

Criterion-referenced marking schemes are in many ways a response to the perceived arbitrariness of the intuitive approach, although there is hugely varying practice in the design of such schemes. At one extreme it can indicate that there is a detailed and precise allocation of marks for each element of an assessment, possibly leaving very little discretion to the marker other than the value judgement as to whether or not the required material was covered. This is a model to be found, for example, on many vocational stage courses. At the other extreme a criterion-referenced scheme might broadly indicate that parts (a) and (b) of a question carry up 25% of the marks each, whilst part (c) carries the remaining 50% (see Appendix 6). Within these sub-sections intuitive marking can resurface. Criterion referenced marking can also encompass the indication of the required standard to be reached in order for marks to be awarded within a particular banding. Thus, to achieve a mark of 50% a candidate may only need to display an 'understanding' of a topic, whilst a candidate seeking to be awarded 70% must provide evidence of having 'critically assessed' the concept (see Appendix 7). Questions arise as to whether these criteria ought to be published to students at the time the assessment is set and, if they are, the extent to which they determine student behaviour in attempting the assessment. Inevitably, with unseen written examinations, the scope for providing content related assessment criteria to candidates in advance (beyond the use of generic statements on standards) is extremely limited without compromising the nature of the assessment itself.

The conventional wisdom is that adherence to clear and detailed marking criteria helps to reduce the incidence of error variation, but it should be noted that the production of such schemes by teaching teams does not necessarily mean they are applied consistently. Some academics may see the production of marking criteria as a QAA hoop through which they have to jump – a bureaucratic 'fig-leaf' to cover what is in effect intuitive marking. Even those adopting such schemes in good faith may lack the time to apply them thoroughly. A scheme may be defective if the criteria have been poorly drafted. In addition, there may be no opportunity to review the efficacy of such schemes through teaching team review exercises.

13. Exercise 4 – the Criminal Law Marking Workshop

By July 2004 the Research Project team had access to the Criminal Law examination scripts containing answers to the common examination problem (set out in Appendix 2) from all three participating HEIs. In order to explore the extent to which reference to detailed marking criteria, and the submission of marking to a moderation process might prevent error variation, a Workshop event for Criminal Law markers was convened. The team of markers assembled for this exercise – Assessor Group 4 – comprised experienced Criminal Law teachers from a range of UK HEIs. Eight examination scripts were selected. In each case the candidate's answer to the problem question set out at Appendix 2, was photocopied for the markers, all internal marker comments being removed. All the answers were handwritten. Marking criteria and a standards statement were provided to each marker who also received an oral briefing on these documents. (Appendices 6 and 7). As in Exercise 3 (above) the markers were asked, in addition to identifying the most appropriate grade, to indicate the lowest and highest marks they would be prepared to agree to for both of the two questions; see instruction sheet at Appendix 8.

13.1 Exercise 4 - Unmoderated criterion referenced marking exercise

Table 4(a) sets out the unmoderated first internal mark for each of the 8 answers used in Exercise 4, along with the final internal moderated marks, as agreed with the external examiner, and the average marks awarded by Assessor Group 4 markers at the Workshop event. A detailed breakdown of all of the marks awarded by each of the 11 Assessor Group 4 markers is set out in Table 4(b) in Appendix 9.

Table 4(a)									
Paper Number	792	663	888	075	703	953	962	980	Ave
Internal first mark	40.0	45.0	35.0	48.0	80.0	60.0	70.0	40.0	52.3
Moderated internal mark	40.0	30.0	35.0	48.0	80.0	62.0	70.0	40.0	50.6
Average Assessor Group 4 ideal mark	27.6	41.1	27.1	38.2	59.2	56.9	70.7	27.8	43.6
Average Assessor Group 4 ideal mark deviation from first internal mark	-12.4	-3.9	-7.9	-9.8	-20.8	-3.1	0.7	-12.2	-8.7
Average Assessor Group 4 ideal mark deviation from internally moderated mark	-12.4	11.1	-7.9	-9.8	-20.8	-5.1	0.7	-12.2	-7.0

This exercise effectively provided for 176 points of marking comparison (i.e. 8 scripts, remarked by 11 markers, with results compared with the internal mark

and the moderated mark awarded by the participating HEI). In only 4 instances did a marker in Assessor Group 4 award the same grade as that recorded by the internal or the HEI moderation process. In 8 instances there was a variation of more than 30% between the Assessor Group marker's ideal grade and that recorded at internal first marking or moderation. In 28 cases there was a variation of between 20 and 30%.

Overall, markers in Assessor Group 4 were significantly less generous in their gradings than the internal marker and moderators. Looking more closely at the pass/fail boundary reveals some interesting findings. The participating HEI approved failing grades for both script 663 and script 888. The Assessor Group 4 markers generally supported these judgements – only one felt that, ideally, 888 should have passed, whilst 8 felt that 663 should have been awarded a pass. Against this, however, is the fact that the internal markers and moderators were content to award 980 a bare pass, whilst not one of the Assessor Group 4 markers would have done so based on their ideal grades. Furthermore, 10 of the 11 Assessor Group 4 markers would have failed paper 792, also awarded a bare pass on the internal marking, and 8 would have opted to fail paper 075, which was awarded a clear pass by internal markers. Given that all members of Assessor Group 4 were experienced in the assessment of Criminal Law on undergraduate programmes the project team felt that these results might have had implications for the effectiveness and reliability of the moderation process that had produced the original moderated grades at the participating HEI. Interviews with those participating in the Criminal Law marking Workshop, however, indicated that there were other factors that might have been significant:

- the workshop provided a more relaxed and reflective environment in which to work than might normally be the case when markers are operating under the pressures of an examination marking schedule – this was particularly the case where marking was carried out in the context of detailed criteria and standards;
- there was no clear consensus on the extent to which the awarding of credit to a particular cohort of students is objective, in the sense of being referenced against objective criteria, and the extent to which it is subjective, i.e. the extent to which the marks awarded reflect the performance of the student against the performance of the cohort as a whole. It was acknowledged that this was just as likely to lead to a marker being more generous to a script because it was simply the best of a poor batch, as it was to lead a marker to take a more austere view of a poor script amongst many good ones;
- Where unit leaders took prime responsibility for first marking scripts and/or leading internal moderation there may be a tendency to adopt a generally benign view of student failings, based on a knowledge of the difficulties students had encountered with the subject during the year, or simply a recognition that a failure rate above a certain level was likely to be regarded as aberrant.

13.2 Exercise 4 – Criterion referenced expert Criminal Law mock moderation meeting

The second part of the Criminal Law Marking Workshop involved setting up mock exam moderation meetings (see Appendix 10). Two members of Assessor Group 4 were assigned to chair the moderation meetings and the remaining markers were allocated to one of the two groups by drawing lots. The aim of the exercise was to explore the extent to which the variation in marks awarded would be 'smoothed' by the process, and to see what the effect would be on scripts at the pass/fail borderline. The results from this process are set out in Table 4(c).

Table 4(c)									
Paper Number	792	663	888	75	703	953	962	980	Ave
Average Assessor Group 4 unmoderated ideal mark	26.9	40.5	26.6	36.4	58.0	56.6	70.5	26.9	31.1
Mark agreed by moderation group A	34.0	48.0	29.0	34.0	60.0	58.0	75.0	30.0	33.5
Mark agreed by moderation group B	24.0	38.0	25.0	39.0	60.0	57.0	70.0	27.0	30.9
Moderation Group A deviation from internal first mark	-6.0	3.0	-6.0	-14.0	-20.0	-2.0	5.0	-10.0	-4.5
Moderation Group A deviation from HEI final moderated mark	-6.0	18.0	-6.0	-14.0	-20.0	-4.0	5.0	-10.0	-3.4
Moderation Group A deviation from average Assessor Group 4 ideal mark	7.1	7.5	2.4	-2.4	2.0	1.4	4.5	3.1	2.3
Moderation Group B deviation from internal first mark	-16.0	-7.0	-10.0	-9.0	-20.0	-3.0	0.0	-13.0	-7.1
Moderation Group B deviation from HEI final moderated mark	-16.0	8.0	-10.0	-9.0	-20.0	-5.0	0.0	-13.0	-5.9
Moderation Group B deviation from average Assessor Group 4 ideal mark	-2.9	-2.5	-1.6	2.6	2.0	0.4	-0.5	0.1	-0.2
Average of Moderation Group A and B moderated marks	29.0	43.0	27.0	36.5	60.0	57.5	72.5	28.5	32.2

Table 4(c)									
Paper Number	792	663	888	75	703	953	962	980	Ave
Moderation Group A and B average moderated mark deviation from internal first mark	-11.0	-2.0	-8.0	-11.5	-20.0	-2.5	2.5	-11.5	-5.8
Moderation Group A and B average moderated mark deviation from HEI final moderated mark	-11.0	13.0	-8.0	-11.5	-20.0	-4.5	2.5	-11.5	-4.6
Moderation Group A and B average moderated mark deviation from average Assessor Group 4 ideal mark	2.1	2.5	0.4	0.1	2.0	0.9	2.0	1.6	1.1

These results indicate that only in relation to paper 663 (Moderation Group B) did the moderation turn a script that, on the basis of average ideal marks, Assessor Group 4 felt was a pass into a fail. Between the 2 Moderation Groups, however, there was a marked difference in the treatment of scripts. Note that Moderation Group A arrived at a final mark that was higher than the Assessor Group 4 average in 7 out of 8 cases. By contrast Moderation Group B arrived at a final mark below the Assessor Group 4 ideal average in respect of half of the scripts. Script 663 remained the only script in respect of which the two Moderation Groups could not agree an outcome: Group A settling on a clear pass (almost 2:2) grade; Group B deciding that it should remain a fail on 38%.

The other key line of data is the comparison between the average of the agreed marks for Moderation Groups A and B, and the original moderated mark awarded by the participating HEI. In every case, saving script 962, the Workshop Moderation Exercise resulted in a lower grade than that actually awarded to the student. In the case of paper 703 the disparity was 20%. There were three scripts, 792, 075, and 980, that the Workshop moderators deemed clear fails, that nevertheless were awarded passing grades following the participating HEI moderation process. The Project Team are of the view that the moderation stage of the workshop indicates not only that effective moderation is essential in identifying and to some extent nullifying error variation in grading, but further research might be warranted into the effectiveness of moderation process at HEIs where large cohorts are being assessed under severe time constraints. In particular there may be issues here for reflection for smaller law schools, where there is sometimes only one fully qualified internal marker in a given subject. Effective moderation in such circumstances may be difficult to achieve.

14. Exercise 5: Parallel marking in Criminal Law

Setting the same question on the Criminal Law paper at 3 HEI's presented the Research Project team with an additional opportunity for the comparison of

marking standards. The Criminal Law subject leaders at the three institutions (designated HEI 1, 2 and 3) each selected at random 20 exam scripts where the candidate had attempted the Criminal Law problem used in Exercises 3 and 4 (see Appendix 2). As with Exercises 3 and 4, all internal marker comments were removed and the answers photocopied. Marking criteria used at the originating institution were supplied. Each of the subject leaders (referred to here collectively as Assessor Group 5) then marked a batch of 20 scripts supplied by the other subject leaders. In each case the pass mark was 40% and the question was assessed as a Year 2/level 2 assessment. Each subject leader marked the scripts according to his or her 'home' institution's internal moderation processes. The detailed tables of results from this exercise are set out at Appendix 11 (Tables 5(a) – (c)). Table 6 summarises the outcomes of the exercise

Table 6				
Parallel Marking Exercise Data Summary				
Average marks awarded by HEI 1 to its own scripts	54.2			
Average marks awarded by HEI 2 to its own scripts	48.8			
Average marks awarded by HEI 3 to its own scripts	53.8			
Average of marks awarded to HEI 1 scripts by HEI 2 and 3	45.2			
Average of marks awarded to HEI 2 scripts by HEI 1 and 3	47.1			
Average of marks awarded to HEI 3 scripts by HEI 1 and 2	41.9			
	Average of marks awarded to HEI 1 scripts	Average of marks awarded to HEI 2 scripts	Average of marks awarded to HEI 3 scripts	Average of marks awarded by each HEI
Average of marks awarded by HEI 1	54.2	45.2	43.3	47.6
Average of marks awarded by HEI 2	42.6	48.8	40.5	44.0
Average of marks awarded by HEI 3	47.8	49.0	53.8	50.2
Average of marks awarded	48.2	47.6	45.9	

to each HEI				
-------------	--	--	--	--

The summary data indicates that, assuming comparability of marking standards, the HEI 1 scripts were the superior grouping, and the HEI 2 selection the weakest. If this were the case, one would have expected to see this borne out in the remarking by the other participating institutions of the HEI 1 and HEI 2 script. Table 6 shows that the HEI 2 scripts emerge as those attracting the highest average marks from the HEI 1 and 3 markers, whilst the HEI 3 scripts attract the lowest average mark from the HEI 1 and 2 markers. In every case the remarking produced lower average grades. The project team noted that this seemed to be a feature of its remarking exercises; see Exercise 4 (above).

Exercise 5 involved a total of 120 instances of remarking. In only 4 instances did markers in Assessor Group 5 match the mark awarded by the internal marker at the originating HEI. Regarding the pass/fail boundary, Tables 5(a)-(c) (set out at Appendix 11), indicate that HEIs 2 and 3 only shifted two HEI 1 scripts from pass to fail (although not the same two), and similarly HEI 1 and 3 only shifted two HEI 2 passes to failing grades (this time the same two). It was the remarking of scripts supplied by HEI 3 that produced the most significant changes. HEI 1 failed 5 scripts passed at HEI 3, and HEI 2 failed 7. Closer inquiry reveals, however, that there are some very significant grade variations between HEI grades approved by exam boards and those arrived at in Exercise 5. There are 14 instances where the remarking grade was over 20% higher or lower than that awarded at the originating HEI. A disparity made all the more alarming given that all three markers were very experienced in assessing student work in the discipline concerned. In 64 instances, (i.e. over a third of the pieces marked) the disparity was at least 10% - enough to represent a difference of class of degree. The Research Project team feel that this again points to the crucial role played by effective moderation.

15. Conclusions

As with all research it is important that the resulting data is not made to bear an unjustified meaning. David Hume's (1711-1776) identification of the difficulties of deriving statements of value from statements of fact remains no less valid today than when he enunciated it in Book III, part 1, section 1 of his *Treatise of Human Nature*:

“In every system of morality, which I have hitherto met with, I have always remarked, that the author proceeds for some time in the ordinary ways of reasoning and establishes, or makes observations concerning human affairs; when all of a sudden I am surprised to find, that instead of the usual copulations of propositions is, and is not, I meet with no proposition that is not connected with an ought, or ought not. This change is imperceptible; but is however, of the last consequence. For as this ought, or ought not, expresses some new relation or affirmation, 'tis necessary that it should be observed and explained; and at the same that a reason should be given; for what

seems altogether unconceivable, how this new relation can be a deduction from others, which are entirely different from it.”

The Research Project team has not, therefore, sought to make conclusive judgments about the relative application of standards applied within post and pre-1992 institutions, or the extent to which students understand the outcomes and criteria by reference to which judgment is made about their knowledge and skills. To be able to answer such questions would have required the collection of a highly complex data set outside the scope of this research. Neither have we sought to make judgments about the dynamics of assessment processes as manifested in the working of marking teams, nor how the skills needed to make valid judgments about student achievement form part of actual staff development policies within law schools.

A further factor that must be kept in mind is the fact that the participants in the various marking exercises that are described in this research were in the main self-selecting and that no attempt was made to ensure that they were representative of those who work in HEI law schools. Different outcomes may have been achieved if another mode of selection had been used, for instance, by utilising established marking teams from a single institution. It is with these caveats that the following findings and recommendations must be viewed.

16. Findings

16.1 Intuitive understanding of standards

The first finding that the Research Project team arrived at was that, on the basis of the marking exercises carried out in the early stages of the project, there cannot be said to be an intuitive understanding, amongst those participating in the exercises that made up the research project, of the standards to be applied in respect of the work that was the subject of this study. Not surprisingly, when markers were asked to grade work on the basis of their own intuition without the benefit of assessment criteria and follow up moderation there was variation in the grades awarded in respect of the work that formed the basis of the exercise. As stated above the heterogeneous nature of the group of markers who undertook this exercise must be acknowledged and could have had an impact on the outcome. Such variation may not have occurred in respect of a more homogeneous group, of the type to be commonly found in the form of long established marking teams within HEI law schools.

However, this finding would lend powerful support for the use of assessment criteria, and justification for the precepts to be found in the Section of the QAA Code of Practice relating to assessment. Nevertheless, in the opinion of the research team the development and use of assessment criteria are in themselves not enough to ensure the uniform application of standards.

16.2. Use and understanding of assessment criteria insufficient on its own.

A further finding was that the use and understanding of criteria did not eradicate marker variation. It is often argued that it is sufficient for members of marking teams to be briefed on the assessment criteria to be applied when marking student work to ensure the uniform application of standards. This was not a hypothesis accepted by the research team and support was received through the recreation of marking exercises where graders received such briefing where it was not followed by post marking evaluation, discussion and statistical analysis.

16.3 Agreement as to student achievement dependant on rigorous moderation

Uniform application of standards was only achieved through the use of a rigorous process of moderation such as was applied in the final stage of the research. This involved all members of the marking team being thoroughly briefed by the author of the questions used to test student knowledge. In addition members of the team were given a written explanation of the areas of law students were expected to refer to and the level of analysis they were expected to demonstrate. As described above they were also given an appropriate period of time to consider and reflect on the answers they were evaluating. As indicated there was still some variation, but this was eradicated through the deployment of a process of moderation, which involved each student's performance being subject to detailed and intense discussion by the two marking teams.

17. Recommendations

17.1 Resourcing of assessment processes

The major recommendation arising out of this research is the importance of that the process of assessment is adequately resourced. Markers must be given the support and time needed to carry out the process of assessment in a considered manner. This must involve ensuring that they have a full understanding of the assessment criteria to be applied and the outcomes students are required to show competence in. This must then be followed up by a moderation process in which student performance is subject to considered analysis and debate by members of the marking team. In a period of mass higher education it is vital that these needs are observed. Grading should not be a rushed exercise and it is imperative that properly resourced teams undertake it.

17.2. Use of statistical tools to assist assessment

What should also be considered is the guidance to be gained from using statistical tools to assist the process of assessment. The existence of error variation can be detected through the use of what to a statistician is a relatively simple calculation. Although the research did not look at how

widespread the use of such tools is within HEI law schools we suspect that they are relatively rarely used. Given the insight that they can provide we would suggest that such use becomes more widespread.

17.3 Further Research

(a) A question that the research did not answer is the extent to which there are shared understandings of standards across the sector. The project did not, for instance, seek to evaluate the extent to which there was an understanding of the extant guidance on levels of graduate achievement as found in the Benchmark Statement for Law. Whether there is such sector wide understanding and the extent to which law schools prioritise it through the induction of new and the development of existing staff awaits a further research project. What this research project tends to bears out is that with such development, reinforced by rigorous and properly resourced processes of assessment and moderation, there can be a common understanding and application of standards.

(b) Further research may be necessary into the nature and usefulness of marking criteria and guidance for internal and external markers, and the extent to which they assist in the prevention of assessment grade error variation.

(c) The lack of a highly defined prescriptive approach to the coverage of the Foundations of Legal Knowledge in law degrees, and the extent to which this hampers comparisons of marking standards, also provides a useful avenue for further research.

17.4. Staff development

We are similarly confident of the benefits to be gained from markers from different institutions meeting to undertake grading exercises. It is through such exercises, which attracted participants at all stages of the research project, that cross institution appreciation of standards can be generated, and we recommend that they become a standard feature of the work of UKCLE as it seeks to disseminate good practice in the teaching and assessment of law.

References.

- Association of Law Teachers and Committee of Heads of University Law Schools and the Society of Public Teacher of Law, (1997) *Graduate Standards in Law: Report of a Feasibility Study Conducted for the Quality Assurance Agency for Higher Education*
- Cuthbert P, (2001) "Large Student Groups: some simple techniques for monitoring marking", *Educational Developments*, Issue 2:2 May 2001, pp17-20
- Ecclestone K, (2001) "'I know a 2:1 when I see it": understanding criteria for degree classifications in franchised university programmes", *Journal of Further and Higher Education*, Vol 25, No3, pp301-313
- Gibbs G, (1992) *Improving the Quality of Student Learning*, Technical and Educational Services Ltd.
- QAA, (2000 May) Code of Practice for the Assurance of Academic Quality and Standards in Higher Education.

APPENDICIES

Appendix 1 – Exercise 2 Tables 2(c) and 2(d)

Data from Exercise 2 comparing marks awarded in Exercise 1 and Exercise 2 by the same markers in respect of the same course works (Table 2(c)), and examination answers (Table 2(d)).

Table 2(c)					
	CW 1	CW 2	CW 3	CW 5	Aves
Marker					
Marker 1 exercise 1	35.0	80.0	57.0	58.0	57.5
Marker 1 exercise 2	48.0	68.0	66.0	56.0	59.5
Marker 1 deviation from exercise 1 to 2	13.0	-12.0	9.0	-2.0	2.0
Marker 2 exercise 1	46.0	68.0	58.0	73.0	61.3
Marker 2 exercise 2	35.0	70.0	58.0	53.0	54.0
Marker 2 deviation from exercise 1 to 2	-11.0	2.0	0.0	-20.0	-7.3
Marker 3 exercise 1	35.0	80.0	56.0	60.0	57.8
Marker 3 exercise 2	42.0	70.0	55.0	67.0	58.5
Marker 3 deviation from exercise 1 to 2	7.0	-10.0	-1.0	7.0	0.8
Marker 4 exercise 1	40.0	72.0	58.0	58.0	57.0
Marker 4 exercise 2	40.0	74.0	37.0	55.0	51.5
Marker 1 deviation from exercise 1 to 2	0.0	2.0	-21.0	-3.0	-5.5
Original internal mark	45.0	75.0	69.0	58.0	61.8
Internal marker's remarking	40.0	65.0	50.0	52.0	51.8
Original marker deviation from exercise 1 to 2	-5.0	-10.0	-19.0	-6.0	-10.0
Average 1st mark*	39.0	75.0	57.3	62.3	58.4
Average 2nd mark**	41.0	69.4	53.2	56.6	55.1
Average Deviation	2.0	-5.6	-4.1	-5.7	-3.3
* excluding original internal mark					
** including internal markers remarking					

Table 2(d)													
Candidate	A	B	C	D	E	F	G	H	I	J	K	L	Aves
Marker 1 1st Mark Q1	54.0	38.0	35.0	5.0	56.0	49.0	54.0	58.0	66.0	60.0	59.0	50.0	48.7
Marker 1 1st Mark Q2	40.0	38.0	25.0	35.0	65.0	60.0	65.0	63.0	54.0	53.0	58.0	49.0	50.5
Marker 1 1 st Mark Total	47.0	38.0	30.0	20.0	60.5	54.5	59.5	60.5	60.0	56.5	58.5	49.5	49.5
Marker 1 2nd Mark Q1	46.0	35.0	30.0	5.0	50.0	40.0	55.0	68.0	70.0	57.0	60.0	50.0	47.2
Marker 1 2nd Mark Q2	48.0	49.0	32.0	30.0	67.0	55.0	70.0	50.0	54.0	55.0	65.0	56.0	52.3
Marker 1 2nd Mark Total	47.0	42.0	31.0	17.5	58.5	47.5	62.5	59.0	62.0	56.0	62.5	53.0	49.6
Marker 1 Deviation from Exercise 1 to 2	0.0	4.0	1.0	-2.5	-2.0	-7.0	3.0	-1.5	2.0	-0.5	4.0	3.5	0.0
Marker 2 1st Mark Q1	53.0	46.0	34.0	10.0	13.0	47.0	60.0	55.0	58.0	52.0	57.0	58.0	45.3
Marker 2 1st Mark Q2	55.0	45.0	24.0	54.0	66.0	58.0	63.0	47.0	58.0	55.0	56.0	43.0	52.8
Marker 2 1st Mark Total	54.0	45.5	29.0	32.0	39.5	52.5	61.5	51.0	58.0	53.5	56.5	50.5	48.5
Marker 2 2nd Mark Q1	50.0	44.0	37.0	7.0	20.0	45.0	45.0	57.0	57.0	62.0	54.0	48.0	43.8
Marker 2 2nd Mark Q2	50.0	47.0	30.0	46.0	55.0	52.0	55.0	45.0	57.0	58.0	48.0	48.0	49.4
Marker 2 2nd Mark Total	50.0	45.5	33.5	26.5	37.5	48.5	50.0	51.0	57.0	60.0	51.0	48.0	46.4
Marker 2 Deviation from Exercise 1 to 2	-4.0	0.0	4.5	-5.5	-2.0	-4.0	11.5	0.0	-1.0	6.5	-5.5	-2.5	-2.0
Marker 3 1st Mark Q1	50.0	48.0	30.0	15.0	40.0	38.0	57.0	60.0	58.0	56.0	55.0	53.0	46.7
Marker 3 1st Mark Q2	48.0	44.0	31.0	0.0	57.0	72.0	69.0	61.0	57.0	49.0	58.0	46.0	49.6
Marker 3 1st Mark Total	49.0	46.0	30.5	7.5	48.5	55.0	63.0	60.5	57.5	52.5	56.5	49.5	47.9
Marker 3 2nd Mark Q1	50.0	49.0	44.0	12.0	40.0	45.0	55.0	54.0	N/A	51.0	50.0	50.0	45.5
Marker 3 2nd Mark Q2	59.0	52.0	43.0	0.0	57.0	48.0	61.0	51.0	N/A	44.0	49.0	50.0	46.4
Marker 3 2nd Mark Total	54.5	50.5	43.5	6.0	48.5	46.5	58.0	52.5	N/A	47.5	49.5	50.0	45.7

Table 2(d)													
Candidate	A	B	C	D	E	F	G	H	I	J	K	L	Aves
Marker 3 Deviation from Exercise 1 to 2	5.5	4.5	13.0	-1.5	0.0	-8.5	-5.0	-8.0	N/A	-5.0	-7.0	0.5	-1.2
Marker 4 1st Mark Q1	52.0	40.0	45.0	0.0	38.0	52.0	56.0	54.0	54.0	52.0	58.0	50.0	45.9
Marker 4 1st Mark Q2	56.0	60.0	50.0	50.0	65.0	70.0	68.0	54.0	66.0	55.0	64.0	50.0	59.8
Marker 4 1st Mark Total	54.0	50.0	47.5	25.0	51.5	61.0	62.0	54.0	60.0	53.5	61.0	50.0	52.7
Marker 4 2nd Mark Q1	54.0	40.0	45.0	20.0	45.0	62.0	54.0	58.0	55.0	N/A	N/A	N/A	48.1
Marker 4 2nd Mark Q2	50.0	46.0	50.0	10.0	56.0	64.0	70.0	50.0	50.0	N/A	N/A	N/A	49.6
Marker 4 2nd Mark Total	52.0	43.0	47.5	15.0	50.5	63.0	62.0	54.0	52.5	N/A	N/A	N/A	48.8
Marker 4 Deviation from Exercise 1 to 2	-2.0	-7.0	0.0	-10.0	-1.0	2.0	0.0	0.0	-7.5	N/A	N/A	N/A	-2.8
Original internal mark Q1	48.0	40.0	34.0	10.0	31.0	36.0	54.0	61.0	61.0	66.0	72.0	53.0	47.2
Original internal mark Q2	46.0	40.0	34.0	60.0	53.0	58.0	58.0	57.0	53.0	60.0	64.0	49.0	53.0
Original internal mark total	47.0	40.0	34.0	35.0	42.0	47.0	56.0	59.0	57.0	63.0	68.0	51.0	49.8
Internal marker's re-marking Q1	42.0	38.0	38.0	7.0	40.0	38.0	56.0	57.0	55.0	57.0	63.0	53.0	45.3
Internal marker's re-marking Q2	44.0	42.0	40.0	50.0	56.0	56.0	60.0	53.0	62.0	62.0	64.0	50.0	53.5
Internal marker's re-marking total	43.0	40.0	39.0	28.5	48.0	47.0	58.0	55.0	58.5	59.5	63.5	51.5	49.1
Internal marker's deviation from original to Exercise 2	-4.0	0.0	5.0	-6.5	6.0	0.0	2.0	-4.0	1.5	-3.5	-4.5	0.5	-0.7
Average first mark total*	51.0	44.9	34.3	21.1	50.0	55.8	61.5	56.5	58.9	54.0	58.1	49.9	
Average remark total	49.3	44.2	38.9	18.7	48.6	50.5	58.1	54.3	57.5	55.8	56.6	50.6	
Deviation	-1.7	-0.7	4.7	-2.4	-1.4	-5.3	-3.4	-2.2	-1.4	1.8	-1.5	0.8	
*excluding original internal mark ** including internal markers re-marking													

Appendix 2 – Criminal law problem question

ASSESSMENT QUESTION

Oliver has run short of cash. One warm summer's morning he is fortunate to find a £5 note in the street. He decides to pay a visit to the local department store, Sparks and Mensa. He walks into the menswear department, where he puts on a pair of trousers over the shorts he is wearing and removes all the tags and labels. He strides out into the next department, food, where he puts a whole salmon into his shopping bag. He also puts a chicken into the basket provided by the shop. He goes up to the checkout, intending to pay with the £5 note but instead he sees a cashier who he knows. He winks at her and she beckons him through without his paying for anything. He uses the £5 to pay for a book at the newsagent's next door.

Oliver has been charged with the offences of theft and burglary. Advise Oliver as to his liability for these offences.

Appendix 3 – Briefing for Exercise 3 ALT York 2004

ASSOCIATION OF LAW TEACHERS CONFERENCE YORK APRIL 2004

Comparative Marking Project

You have been provided with two examples of student work. They are both answers to a question set on a Criminal Law paper at a UK University. We want to explore the range of marks that such work might be awarded by examiners.

Context: candidates produced these answers under exam conditions. No additional guidance was provided, the papers were unseen, and no materials other than statute books were permitted. Internal markers were not provided with any marking guidance.

Instructions: please mark each script out of 100 – 40 being the pass mark. Allow approximately 15 minutes to assess each script, putting such feedback comments as you feel appropriate on each script. Make sure you fill in the boxes on this cover sheet indicating your grades, and providing some information about you as a marker.

Paper 1 - Your 'ideal' grade:

The minimum mark you would be willing to agree to give this answer:

The maximum mark you would be willing to agree to give this answer:

Paper 2 - Your 'ideal' grade:

The minimum mark you would be willing to agree to give this answer:

The maximum mark you would be willing to agree to give this answer:

About you - please circle the comment that most accurately describes your status

I am a:

Lecturer 'old' university

Lecturer 'new' university

Lecturer overseas

non of these

I am

an experienced Criminal law tutor

a novice criminal law tutor

not a criminal law tutor

not a law tutor

We are seeking volunteers to take part in the UKCLE funded marking workshop that forms part of this project. It will be held at the University of Wolverhampton on July 15th 2004. If you would like to take part or receive more information on this please contact Mike Molan on molanm@lsbu.ac.uk

Alternatively please fill in your details below and hand to a member of the project team at the end of this session.

NAME:

E-mail:

Contact number:

Appendix 4 - Exercise 3 – Complete data for ALT York 2004 intuitive marking exercise – all markers Papers A and B

Complete data for ALT York 2004 intuitive marking exercise – all markers

Paper A

Table 3(a)	Assessor Group 3			
Lowest Grade for Paper A	Ideal grade for Paper A	Highest grade for Paper A	Average of range	Range as a %
20	25	30	25.0	10
39	35	41	38.3	2
30	38	42	36.7	12
36	40	44	40.0	8
40	44	49	44.3	9
39	45	50	44.7	11
44	47	50	47.0	6
43	47	51	47.0	8
46	48	56	50.0	10
40	48	52	46.7	12
42	48	54	48.0	12
48	50	55	51.0	7
45	50	55	50.0	10
48	50	54	50.7	6
48	51	55	51.3	7
45	52	55	50.7	10
48	52	56	52.0	8
49	53	58	53.3	9
48	54	58	53.3	10
50	54	58	54.0	8
45	55	60	53.3	15
50	55	60	55.0	10
50	56	58	54.7	8
55	57	59	57.0	4
54	57	59	56.7	5
54	57	59	56.7	5
54	58	60	57.3	6
55	60	60	58.3	5
56	60	62	59.3	6
53	60	62	58.3	9
55	62	65	60.7	10
56	62	65	61.0	9
59	62	64	61.7	5
62	64	66	64.0	4
60	64	67	63.7	7
60	65	68	64.3	8
60	65	70	65.0	10
60	73	80	71.0	20

Average Lowest Grade for Paper A	Average Ideal grade for paper A	Average Highest grade for Paper A	Average of all grades for Paper A	Average of range
48.6	53.2	57.0	52.9	8.4

Exercise 3 – results of all markers for Paper B

Exercise 3 – Complete data for ALT York 2004 intuitive marking exercise – all markers Paper A

Table 3(b)	Assessor Group 3			
Lowest Grade for Paper B	Ideal grade for Paper B	Highest grade for Paper B	Average of range	Range as a %
30	35	40	35.0	10
30	35	39	34.7	9
30	36	40	35.3	10
36	40	44	40.0	8
43	40	46	43.0	3
40	40	40	40.0	0
34	41	43	39.3	9
40	42	45	42.3	5
38	44	48	43.3	10
38	45	52	45.0	14
42	45	47	44.7	5
38	46	48	44.0	10
43	46	49	46.0	6
42	46	50	46.0	8
42	48	53	47.7	11
42	48	56	48.7	14
48	50	52	50.0	4
47	50	55	50.7	8
45	50	55	50.0	10
47	50	55	50.7	8
46	52	56	51.3	10
45	53	58	52.0	13
50	53	54	52.3	4
52	54	56	54.0	4
49	55	60	54.7	11
50	55	57	54.0	7
50	55	59	54.7	9
48	55	58	53.7	10
48	55	59	54.0	11
50	56	59	55.0	9
55	58	62	58.3	7
58	60	62	60.0	4
55	60	65	60.0	10

Table 3(b)		Assessor Group 3		
Lowest Grade for Paper B	Ideal grade for Paper B	Highest grade for Paper B	Average of range	Range as a %
56	60	63	59.7	7
54	60	62	58.7	8
57	60	64	60.3	7
60	62	64	62.0	4
63	70	75	69.3	12
Average Lowest Grade for Paper B	Average Ideal grade for paper B	Average Highest grade for Paper B	Average of all grades for Paper B	Average of range
45.8	50.3	53.9	50.0	8.1

**Appendix 5 - Complete data for ALT York 2004 intuitive marking exercise
– Results extracted for Non-criminal law markers Papers A and B**

Exercise 3 – Complete data for ALT York 2004 intuitive marking exercise –
Results extracted for Non-criminal law markers only – Paper A

Table 3(d)			
Non-criminal law	Markers	Paper A	
Lowest Grade for Paper A	Ideal grade for Paper A	Highest grade for Paper A	Average of range
20	25	30	25.0
39	35	41	38.3
40	44	49	44.3
39	45	50	44.7
40	48	52	46.7
42	48	54	48.0
46	48	56	50.0
45	50	55	50.0
48	50	54	50.7
45	52	55	50.7
48	50	55	51.0
48	51	55	51.3
48	52	56	52.0
48	54	58	53.3
45	55	60	53.3
50	54	58	54.0
50	56	58	54.7
50	55	60	55.0
54	57	59	56.7
55	57	59	57.0
54	58	60	57.3
55	60	60	58.3
53	60	62	58.3
56	60	62	59.3
56	62	65	61.0
59	62	64	61.7
60	64	67	63.7
60	73	80	71.0
Averages	Averages	Averages	Averages
48.3	53.0	56.9	52.8

**Exercise 3 – Complete data for ALT York 2004 intuitive marking exercise
– Results extracted for Non-criminal law markers only – Paper B**

Table 3(e)			
Non-criminal law	Markers	Paper B	
Lowest Grade for Paper B	Ideal grade for Paper B	Highest grade for Paper B	Average of range
30	35	40	35.0
34	41	43	39.3
36	40	44	40.0
40	42	45	42.3
38	46	48	44.0
38	45	52	45.0
43	46	49	46.0
42	46	50	46.0
42	48	53	47.7
42	48	56	48.7
45	50	55	50.0
47	50	55	50.7
46	52	56	51.3
45	53	58	52.0
50	53	54	52.3
52	54	56	54.0
50	55	57	54.0
49	55	60	54.7
50	55	59	54.7
50	56	59	55.0
55	58	62	58.3
54	60	62	58.7
56	60	63	59.7
58	60	62	60.0
55	60	65	60.0
57	60	64	60.3
60	62	64	62.0
63	70	75	69.3
Averages	Averages	Averages	Averages
47.4	52.1	55.9	51.8

Appendix 6 – Marking Guidance for Criminal Law Marking Workshop

Exercise 4 – Criminal Law Marking Workshop

MARKING GUIDANCE

Criteria	What we want	Marks available
1. Knowledge and understanding	Theft of the £5 note – property – lost or abandoned – if abandoned no theft – if lost – appropriation – ITPD is evident – main issue then dishonesty – consider s.2(1)(a)-(c) – then Ghosh	20%
1. Knowledge and understanding	Theft of trousers by removing labels – PBTA – appropriated – ITPD seems evident – dishonesty is obvious.	20%
1. Knowledge and understanding	Theft of salmon by putting it in his own bag – PBTA – clearly appropriation – did he have ITPD – dishonesty seems evident.	15%
1. Knowledge and understanding	Theft of chicken – PBTA – comes by it innocently – appropriates when in basket – but no mens rea until he leaves the store without paying	20%
1. Knowledge and understanding	Burglary – did he intend to steal before entering the store – if so s.9(1)(a) – otherwise could be s.9(1)(a) burglary where he moves into the food department if he had an intention to steal when he did so. Section 9(1)(a) liability could become s.9(1)(b) where he actually steals.	25%
	Credit awarded for knowledge and understanding must reflect the extent to which the following skills are evidenced by the candidate's answer	
2. Legal skills	<ul style="list-style-type: none"> evidence of ability to accurately analyse a factual problem and isolate the issues of criminal liability to which it gives rise evidence of ability to persuasively and cogently argue how the relevant criminal law, should be applied to the resolution of a given problem, applying rules of precedent and statutory interpretation as appropriate evidence of engagement with, and ability to critically assess, the arguments for reform of the law where relevant 	

Criteria	What we want	Marks available
3. Transferable skills	<ul style="list-style-type: none">• has the candidate demonstrated an ability to evaluate and test opposing arguments with a view to rationalising stated preferences for a given line of reasoning, supported by authority and appropriate evidence?	
4. Practical skills	<ul style="list-style-type: none">• has the candidate demonstrated an ability to conduct basic independent problem solving tasks?	

Appendix 7 – Guidance on marking standards for Criminal Law Marking Workshop

Exercise 4 – Criminal Law Marking Workshop

The awarding of credit: standards

Pass 80%+

In the context of: (a) the level at which the unit is being assessed (L2); and (b) the University's policies on the assessment of candidates with special needs, work awarded credit in the starred first class 80%+ band will normally display many of the following features:

- Complete coverage of the assessment criteria, evidencing comprehensive knowledge, a confident and deep understanding of the subject, and excellence in analysis and evaluation of the relevant material;
- No significant errors or omissions.
- Excellence in the exposition of opinion, supported by reasons, based on evidence or authority
- Originality of exposition or treatment of the issues
- Excellent integration and structuring of materials
- Very good use of English –few if any grammatical errors

Pass 70% to 79%

In the context of: (a) the level at which the unit is being assessed (L2); and (b) the University's policies on the assessment of candidates with special needs, work awarded credit in the first class 70% to 79% band will normally display many of the following features:

- Complete coverage of the assessment criteria, evidencing comprehensive knowledge, a confident understanding of the subject, and very good analysis and evaluation of the relevant material;
- No significant errors or omissions.
- Very good exposition of opinion, supported by reasons, based on evidence or authority
- Very good integration and structuring of materials
- Very good use of English –few if any grammatical errors

Pass 60% to 69%

In the context of: (a) the level at which the unit is being assessed (L2); and (b) the University's policies on the assessment of candidates with special needs, work awarded credit in the upper second class 60% to 69% band will normally display many of the following features:

- Substantial coverage of the assessment criteria, evidencing (as appropriate to the assessment set) a sound knowledge of the area being assessed, a good understanding of the subject, and good analysis and evaluation of the relevant material;
- Only minor errors or omissions.
- Good exposition of opinion, supported by reasons, based on evidence or authority
- Good integration and structuring of materials
- Good use of English – only minor grammatical errors and spelling mistakes.

Pass 50% to 59%

In the context of: (a) the level at which the unit is being assessed (L2); and (b) the University's policies on the assessment of candidates with special needs, work awarded credit in the lower second class 50% to 59% band will normally display many of the following features:

- Generally sound in terms of coverage of assessment criteria, evidencing (as appropriate to the assessment set) an adequate knowledge of the area being assessed, a basic understanding of the subject, and a largely competent analysis and evaluation of the relevant material;
- One or two major omissions in terms of content coverage.
- Competent exposition of opinion, supported by reasons, based on evidence or authority
- Competent integration and structuring of materials
- Appropriate use of English, with only a few significant errors in spelling and grammar.

Pass 40% to 49%

In the context of: (a) the level at which the unit is being assessed (L2); and (b) the University's policies on the assessment of candidates with special needs, work awarded credit in the third class 40% to 49% band will normally display many of the following features:

- Satisfactorily meets the minimum requirements of the assessment criteria, evidencing (as appropriate to the assessment set) a basic knowledge of the area being assessed, failure to understand some key points being assessed, weakness in the analysis and evaluation of the relevant material;
- Numerous errors, omissions and/or unnecessary detail
- Weaknesses in, or lack of evidence of, the candidate's ability to express opinion, supported by reasons, based on evidence or authority
- Significant flaws in the integration and structuring of material
- Poor use of English – weak grasp of grammar – some incoherent statements.

NB: A candidate who has not directly addressed an assessment question, or has addressed a related question but not that which has been set by the examiner may not be awarded credit falling within this band unless there is sufficient evidence that the minimum learning outcomes have been achieved.

Fail 35-39%

In the context of: (a) the level at which the unit is being assessed (L2); and (b) the University's policies on the assessment of candidates with special needs, work awarded credit in the threshold (normally compensatable) fail 35-39% band will normally display many of the following features:

- Narrowly fails to meet the minimum requirements of the assessment criteria, evidencing (as appropriate to the assessment set) some key weaknesses regarding knowledge of the area being assessed, failure to understand or engage with the main points of the assessment, significant weaknesses in the analysis and evaluation of relevant material;
- Numerous errors, omissions and/or unnecessary detail.
- Significant weaknesses in, or lack of evidence of, the candidate's ability to express opinion, supported by reasons, based on evidence or authority
- Very significant flaws in the integration and structuring of material
- Poor use of English – weak grasp of grammar – a significant number of incoherent statements.

NB: A candidate who has not directly addressed an assessment question, or has addressed a related question but not that which has been set by the examiner may not be awarded credit falling within this band unless there is sufficient evidence that the candidate has made a significant (albeit unsuccessful) attempt at achieving the minimum learning outcomes.

Fail 30%-34%

In the context of: (a) the level at which the unit is being assessed (L2); and (b) the University's policies on the assessment of candidates with special needs, work awarded credit in the threshold (not normally compensatable) fail 30%-34%

band will normally display many of the following features:

- Fails to meet the minimum requirements of the assessment criteria, evidencing (as appropriate to the assessment set) significant key weaknesses regarding knowledge of the area being assessed, serious failure to understand or engage with the majority of the points being assessed, very significant weaknesses in (or no) analysis and evaluation of relevant material;
- Numerous errors, omissions and/or unnecessary detail.
- Very significant weaknesses in, or lack of evidence of, the candidate's ability to express opinion, supported by reasons, based on evidence or authority
- Little evidence of ability to integrate and structure material appropriately

- Very poor use of English – very weak grasp of grammar – a significant number of incoherent statements.

NB: A candidate who has not directly addressed an assessment question, or has addressed a related question but not that which has been set by the examiner may not be awarded credit falling within this band unless there is sufficient evidence that the candidate has made a significant (albeit unsuccessful) attempt at achieving the minimum learning outcomes.

Fail 0%-29%

In the context of: (a) the level at which the unit is being assessed (L2); and (b) the University's policies on the assessment of candidates with special needs, work awarded credit in the irredeemable fail 0%-29% band will normally display many of the following features:

- Clearly fails to meet the minimum requirements of the assessment criteria in any meaningful way, evidencing (as appropriate to the assessment set) very significant key weaknesses regarding knowledge of the area being assessed, very serious failure to understand or engage with the majority of the points being assessed, absence of analysis and evaluation of relevant material;
- Insufficiency of material – what is provided may be correct but it may be too brief.
- Failure to apply relevant law in any meaningful way to the question set
- Significant irrelevancies and non-sequiturs.
- Failure to integrate and structure material appropriately
- Major deficiencies in written English.

Notes

1. Knowledge = relevant knowledge in terms of the syllabus and the assessment criteria.
2. Assessment criteria = testing the achievement of specific unit outcomes to a minimum standard and beyond.
3. The assessment criteria will invite students to achieve more than the minimum standards set by the unit outcomes. The unit outcomes only define the pass/fail standard.
4. Answering examination questions in note form should not necessarily be a bar to the awarding of credit within any of the above bands provided the remaining criteria are satisfied.

Appendix 8 – Criminal Law Marking Workshop individual marker record sheet

UKCLE SPONSORED COMPARATIVE MARKING PROJECT

UNIVERSITY OF WOLVERHAMPTON THURSDAY JULY 15TH 2004

Individual marker record sheet

Allocated marker reference number:

Home Institution:

Number of years as a criminal law marker:

Script number	Highest mark you could agree to	Lowest mark you could agree to	Your preferred grade

Appendix 9 - Exercise 4 – Criminal Law Marking Workshop – complete results data for Assessor Group 4

Table 4(b)									
Paper Number	792	663	888	075	703	953	962	980	Ave
Assessor									
Marker 1 Low	20	40	15	40	60	52	60	15	37.8
Marker 1 Ideal	28	42	20	45	67	58	67	20	43.4
Marker 1 High	34	46	30	50	72	66	72	30	50.0
Marker 1 ideal mark deviation from first internal mark	-12.0	-3.0	-15.0	-3.0	-13.0	-2.0	-3.0	-20.0	-8.9
Marker 1 ideal mark deviation from internally moderated mark	-12.0	12.0	-15.0	-3.0	-13.0	-4.0	-3.0	-20.0	-7.3
Marker 2 low	10.0	35.0	5.0	15.0	63.0	40.0	60.0	10.0	29.8
Marker 2 ideal	20.0	38.0	10.0	25.0	68.0	48.0	65.0	20.0	36.8
Marker 2 high	30.0	48.0	15.0	30.0	75.0	46.0	70.0	30.0	43.0
Marker 2 ideal mark deviation from first internal mark	-20.0	-7.0	-25.0	-23.0	-12.0	-12.0	-5.0	-20.0	-15.5
Marker 2 ideal mark deviation from internally moderated mark	-20.0	8.0	-25.0	-23.0	-12.0	-14.0	-5.0	-20.0	-13.9
Marker 3 low	15.0	30.0	10.0	20.0	52.0	50.0	62.0	5.0	30.5
Marker 3 ideal	20.0	35.0	15.0	25.0	54.0	55.0	65.0	10.0	34.9
Marker 3 high	25.0	38.0	20.0	30.0	58.0	58.0	68.0	15.0	39.0
Marker 3 ideal mark deviation from first internal mark	-20.0	-10.0	-20.0	-23.0	-26.0	-5.0	-5.0	-30.0	-17.4
Marker 3 ideal mark deviation from internally moderated mark	-20.0	5.0	-20.0	-23.0	-26.0	-7.0	-5.0	-30.0	-15.8
Marker 4 low	25.0	45.0	25.0	30.0	53.0	52.0	65.0	30.0	40.6
Marker 4 ideal	30.0	50.0	30.0	35.0	63.0	57.0	75.0	35.0	46.9
Marker 4 high	35.0	55.0	35.0	40.0	68.0	60.0	80.0	38.0	51.4
Marker 4 ideal mark deviation from first internal mark	-10.0	5.0	-5.0	-13.0	-17.0	-3.0	5.0	-5.0	-5.4
Marker 4 ideal mark deviation from internally moderated mark	-10.0	20.0	-5.0	-13.0	-17.0	-5.0	5.0	-5.0	-3.8

Table 4(b)									
Paper Number	792	663	888	075	703	953	962	980	Ave
Marker 5 low	12.0	32.0	30.0	50.0	60.0	55.0	72.0	15.0	40.8
Marker 5 ideal	20.0	40.0	33.0	52.0	66.0	58.0	74.0	24.0	45.9
Marker 5 high	30.0	48.0	34.0	56.0	69.0	60.0	79.0	30.0	50.8
Marker 5 ideal mark deviation from first internal mark	-20.0	-5.0	-2.0	4.0	-14.0	-2.0	4.0	-16.0	-6.4
Marker 5 ideal mark deviation from internally moderated mark	-20.0	10.0	-2.0	4.0	-14.0	-4.0	4.0	-16.0	-4.8
Marker 6 low	10.0	15.0	20.0	41.0	63.0	47.0	66.0	17.0	34.9
Marker 6 ideal	16.0	24.0	33.0	45.0	65.0	52.0	71.0	28.0	41.8
Marker 6 high	26.0	30.0	35.0	48.0	68.0	55.0	74.0	33.0	46.1
Marker 6 ideal mark deviation from first internal mark	-24.0	-21.0	-2.0	-3.0	-15.0	-8.0	1.0	-12.0	-10.5
Marker 6 ideal mark deviation from internally moderated mark	-24.0	-6.0	-2.0	-3.0	-15.0	-10.0	1.0	-12.0	-8.9
Marker 7 low	30.0	40.0	35.0	45.0	60.0	56.0	65.0	30.0	45.1
Marker 7 ideal	35.0	48.0	42.0	50.0	63.0	60.0	68.0	35.0	50.1
Marker 7 high	40.0	50.0	50.0	55.0	65.0	65.0	70.0	40.0	54.4
Marker 7 ideal mark deviation from first internal mark	-5.0	3.0	7.0	2.0	-17.0	0.0	-2.0	-5.0	-2.1
Marker 7 ideal mark deviation from internally moderated mark	-5.0	18.0	7.0	2.0	-17.0	-2.0	-2.0	-5.0	-0.5
Marker 8 low	35.0	38.0	25.0	30.0	45.0	50.0	70.0	35.0	41.0
Marker 8 ideal	35.0	42.0	25.0	30.0	45.0	58.0	75.0	35.0	43.1
Marker 8 high	35.0	42.0	29.0	34.0	49.0	60.0	75.0	39.0	45.4
Marker 8 ideal mark deviation from first internal mark	-5.0	-3.0	-10.0	-18.0	-35.0	-2.0	5.0	-5.0	-9.1
Marker 8 ideal mark deviation from internally moderated mark	-5.0	12.0	-10.0	-18.0	-35.0	-4.0	5.0	-5.0	-7.5
Marker 9 low	20.0	40.0	25.0	35.0	45.0	55.0	69.0	30.0	39.9
Marker 9 ideal	25.0	40.0	30.0	36.0	50.0	60.0	70.0	30.0	42.6
Marker 9 high	29.0	45.0	34.0	40.0	50.0	60.0	75.0	35.0	46.0
Marker 9 ideal mark deviation from first internal mark	-15.0	-5.0	-5.0	-12.0	-30.0	0.0	0.0	-10.0	-9.6
Marker 9 ideal mark deviation from internally moderated mark	-15.0	10.0	-5.0	-12.0	-30.0	-2.0	0.0	-10.0	-8.0

Table 4(b)									
Paper Number	792	663	888	075	703	953	962	980	Ave
Marker 10 low	30.0	47.0	25.0	30.0	40.0	50.0	60.0	25.0	38.4
Marker 10 ideal	40.0	53.0	30.0	37.0	50.0	55.0	69.0	34.0	46.0
Marker 10 high	49.0	59.0	39.0	45.0	59.0	62.0	74.0	39.0	53.3
Marker 10 ideal mark deviation from first internal mark	0.0	8.0	-5.0	-11.0	-30.0	-5.0	-1.0	-6.0	-6.3
Marker 10 ideal mark deviation from internally moderated mark	0.0	23.0	-5.0	-11.0	-30.0	-7.0	-1.0	-6.0	-4.6
Marker 11 low	30.0	34.0	25.0	35.0	55.0	60.0	75.0	30.0	43.0
Marker 11 ideal	35.0	40.0	30.0	40.0	60.0	65.0	79.0	35.0	48.0
Marker 11 high	39.0	44.0	34.0	45.0	65.0	70.0	85.0	39.0	52.6
Marker 11 ideal mark deviation from first internal mark	-5.0	-5.0	-5.0	-8.0	-20.0	5.0	9.0	-5.0	-4.3
Marker 11 ideal mark deviation from internally moderated mark	-5.0	10.0	-5.0	-8.0	-20.0	3.0	9.0	-5.0	-2.6
Average Assessor group 4 low mark	21.5	36.0	21.8	33.7	54.2	51.5	65.8	22.0	38.3
Average Assessor Group 4 ideal mark	27.6	41.1	27.1	38.2	59.2	56.9	70.7	27.8	43.6
Average Assessor group 4 high mark	33.8	45.9	32.3	43.0	63.5	60.2	74.7	33.5	48.4
Average Assessor Group 4 ideal mark deviation from first internal mark	-12.4	-3.9	-7.9	-9.8	-20.8	-3.1	0.7	-12.2	-8.7
Average Assessor Group 4 ideal mark deviation from internally moderated mark	-12.4	11.1	-7.9	-9.8	-20.8	-5.1	0.7	-12.2	-7.0

Appendix 10 - Criminal Law Marking Workshop Moderation exercise record sheet

UKCLE SPONSORED COMPARATIVE MARKING PROJECT

UNIVERSITY OF WOLVERHAMPTON THURSDAY JULY 15TH 2004

Simulated moderation meeting

Chair of moderation group:

Members of moderation group by reference number:

Script number	Highest mark group could agree to	Lowest mark group could agree to	Final moderated grade for script

Appendix 11 - Exercise 5 – Criminal Law Parallel Marking – complete results data for Assessor Group 5

Table 5(a)							
	Parallel	Marking	Exercise	HEI 1	Scripts		
Candidate	HEI 1 exam board agreed grade	HEI 2 marking HEI 1	HEI 2 Deviation from HEI 1	HEI 3 marking HEI 1	HEI 3 deviation from HEI 1	Ave of HEI 2 and HEI 3	Deviation of HEI 2 & 3 Ave from HEI 1
HEI 1 Paper 1	65.0	61.0	-4.0	50.0	-15.0	55.5	-9.5
HEI 1 Paper 2	50.0	24.0	-26.0	35.0	-15.0	29.5	-20.5
HEI 1 Paper 3	58.0	47.0	-11.0	51.0	-7.0	49.0	-9.0
HEI 1 Paper 4	46.0	47.0	1.0	54.0	8.0	50.5	4.5
HEI 1 Paper 5	38.0	35.0	-3.0	38.0	0.0	36.5	-1.5
HEI 1 Paper 6	55.0	46.0	-9.0	49.0	-6.0	47.5	-7.5
HEI 1 Paper 7	62.0	45.0	-17.0	73.0	11.0	59.0	-3.0
HEI 1 Paper 8	55.0	27.0	-28.0	53.0	-2.0	40.0	-15.0
HEI 1 Paper 9	58.0	47.0	-11.0	47.0	-11.0	47.0	-11.0
HEI 1 Paper 10	58.0	52.0	-6.0	30.0	-28.0	41.0	-17.0
HEI 1 Paper 11	35.0	20.0	-15.0	32.0	-3.0	26.0	-9.0
HEI 1 Paper 12	62.0	57.0	-5.0	50.0	-12.0	53.5	-8.5
HEI 1 Paper 13	52.0	54.0	2.0	40.0	-12.0	47.0	-5.0
HEI 1 Paper 14	60.0	57.0	-3.0	54.0	-6.0	55.5	-4.5
HEI 1 Paper 15	48.0	48.0	0.0	40.0	-8.0	44.0	-4.0
HEI 1 Paper 16	56.0	48.0	-8.0	42.0	-14.0	45.0	-11.0
HEI 1 Paper 17	60.0	47.0	-13.0	61.0	1.0	54.0	-6.0
HEI 1 Paper 18	45.0	54.0	9.0	58.0	13.0	56.0	11.0

An examination of the incidence of 'error variation' in the grading of law assessments.

HEI 1 Paper 19	56.0	44.0	-12.0	46.0	-10.0	45.0	-11.0
HEI 1 Paper 20	65.0	63.0	-2.0	53.0	-12.0	58.0	-7.0
Ave	54.2	46.2	-8.1	47.8	-6.4	47.0	-7.2

Table 5(b)	Parallel	Marking	Exercise	HEI 1	Scripts		
Candidate	HEI 2 exam board agreed grade	HEI 1 marking HEI 2	HEI 1 Deviation from HEI 2	HEI 3 marking HEI 2	HEI 3 deviation from HEI 2	Ave of HEI 1 and HEI 3	Deviation of HEI 1 & 3 ave from HEI 2
HEI 2 Paper 1	65.0	53.0	-12.0	62.0	-3.0	57.5	-7.5
HEI 2 Paper 2	56.0	44.0	-12.0	51.0	-5.0	47.5	-8.5
HEI 2 Paper 3	22.0	35.0	13.0	41.0	19.0	38.0	16.0
HEI 2 Paper 4	35.0	22.0	-13.0	40.0	5.0	31.0	-4.0
HEI 2 Paper 5	35.0	30.0	-5.0	43.0	8.0	36.5	1.5
HEI 2 Paper 6	52.0	40.0	-12.0	48.0	-4.0	44.0	-8.0
HEI 2 Paper 7	43.0	40.0	-3.0	42.0	-1.0	41.0	-2.0
HEI 2 Paper 8	40.0	30.0	-10.0	20.0	-20.0	25.0	-15.0
HEI 2 Paper 9	66.0	53.0	-13.0	65.0	-1.0	59.0	-7.0
HEI 2 Paper 10	53.0	45.0	-8.0	50.0	-3.0	47.5	-5.5
HEI 2 Paper 11	52.0	48.0	-4.0	48.0	-4.0	48.0	-4.0
HEI 2 Paper 12	50.0	48.0	-2.0	53.0	3.0	50.5	0.5
HEI 2 Paper 13	63.0	55.0	-8.0	63.0	0.0	59.0	-4.0
HEI 2 Paper 14	42.0	58.0	16.0	60.0	18.0	59.0	17.0
HEI 2 Paper 15	66.0	66.0	0.0	53.0	-13.0	59.5	-6.5
HEI 2 Paper 16	62.0	75.0	13.0	68.0	6.0	71.5	9.5

Table 5(b)							
	Parallel	Marking	Exercise	HEI 1	Scripts		
Candidate	HEI 2 exam board agreed grade	HEI 1 marking HEI 2	HEI 1 Deviation from HEI 2	HEI 3 marking HEI 2	HEI 3 deviation from HEI 2	Ave of HEI 1 and HEI 3	Deviation of HEI 1 & 3 ave from HEI 2
HEI 2 Paper 17	63.0	60.0	-3.0	52.0	-11.0	56.0	-7.0
HEI 2 Paper 18	30.0	18.0	-12.0	40.0	10.0	29.0	-1.0
HEI 2 Paper 19	40.0	25.0	-15.0	35.0	-5.0	30.0	-10.0
HEI 2 Paper 20	40.0	58.0	18.0	45.0	5.0	51.5	11.5
Ave	48.75	45.15	-3.6	48.95	0.2	47.05	-1.7

Table 5(c)							
	Parallel	Marking	Exercise	HEI 1	Scripts		
Candidate	HEI 3 exam board agreed grade	HEI 1 marking HEI 3	HEI 1 Deviation from HEI 3	HEI 2 marking HEI 3	HEI 2 deviation from HEI 3	Ave of HEI 1 and HEI 2	Deviation of HEI 1 & 2 ave from HEI 3
HEI 3 Paper 1	52.0	49.0	-3.0	18.0	-34.0	33.5	-18.5
HEI 3 Paper 2	57.0	48.0	-9.0	47.0	-10.0	47.5	-9.5
HEI 3 Paper 3	42.0	30.0	-12.0	52.0	10.0	41.0	-1.0
HEI 3 Paper 4	67.0	40.0	-27.0	43.0	-24.0	41.5	-25.5
HEI 3 Paper 5	61.0	60.0	-1.0	47.0	-14.0	53.5	-7.5
HEI 3 Paper 6	60.0	50.0	-10.0	34.0	-26.0	42.0	-18.0
HEI 3 Paper 7	59.0	53.0	-6.0	44.0	-15.0	48.5	-10.5
HEI 3 Paper 8	62.0	60.0	-2.0	47.0	-15.0	53.5	-8.5
HEI 3 Paper 9	44.0	37.0	-7.0	32.0	-12.0	34.5	-9.5
HEI 3 Paper 10	61.0	58.0	-3.0	55.0	-6.0	56.5	-4.5
HEI 3 Paper 11	62.0	47.0	-15.0	23.0	-39.0	35.0	-27.0

Table 5(c)							
	Parallel	Marking	Exercise	HEI 1	Scripts		
Candidate	HEI 3 exam board agreed grade	HEI 1 marking HEI 3	HEI 1 Deviation from HEI 3	HEI 2 marking HEI 3	HEI 2 deviation from HEI 3	Ave of HEI 1 and HEI 2	Deviation of HEI 1 & 2 ave from HEI 3
HEI 3 Paper 12	50.0	35.0	-15.0	43.0	-7.0	39.0	-11.0
HEI 3 Paper 13	66.0	48.0	-18.0	51.0	-15.0	49.5	-16.5
HEI 3 Paper 14	38.0	28.0	-10.0	52.0	14.0	40.0	2.0
HEI 3 Paper 15	36.0	34.0	-2.0	58.0	22.0	46.0	10.0
HEI 3 Paper 16	49.0	48.0	-1.0	17.0	-32.0	32.5	-16.5
HEI 3 Paper 17	62.0	67.0	5.0	46.0	-16.0	56.5	-5.5
HEI 3 Paper 18	55.0	38.0	-17.0	37.0	-18.0	37.5	-17.5
HEI 3 Paper 19	38.0	15.0	-23.0	11.0	-27.0	13.0	-25.0
HEI 3 Paper 20	55.0	20.0	-35.0	52.0	-3.0	36.0	-19.0
Ave	53.8	43.3	-10.6	40.5	-13.4	41.9	-12.0