

Marking and Moderation: a guide for discussion

We need to talk about marking...

Marking students work fairly, objectively and consistently is an expectation of those who mark in higher education. However the literature (Bloxham&Boyd, 2006) and students' perceptions as captured in recent NSS results (OfS, 2021) indicate we are falling short of this ideal.

This briefing paper is intended to support discussion at module, programme or departmental level about marking and moderation processes, to provide reassurance for both markers and students about the robustness of the process.

It is for module or programme teams and/or the department as a whole to agree what is most appropriate as an approach in their context(s).

For the purposes of this paper, it is assumed that the assessment task and associated assessment criteria are already decided. Further support on assessment design is available¹.

How should I/we prepare for marking?

We may feel that time spent preparing for marking would be better spent actually marking. But what if, as a marking pair or team, we are not quite on the same page? This then emerges at the later moderation phase and results in having to adjust marks or, even worse, mark everything again!

“The considerable body of research into marking in higher education supports the claim that marking in the sector is unreliable” p.95 (McConlogue, 2020)

A shared understanding?

Taking part in an activity to share markers' understanding and application of assessment criteria has been found to improve consistency

¹ <https://www.lancaster.ac.uk/od-and-ed/educational-development/assessment-practice>

² We mean here the whole marking team where there are 3 or more markers including postgraduates who teach (GTAs)

³ The suggested minimum sample is 3 items to provide sufficient information for the marking team to discuss matters

amongst pairs or teams of markers² (Pokorny & Warren 2021), thus reducing the degree of variation in marks awarded. “Calibration within the higher education sector is an approach that aims to ensure consistent standards for judging the quality of student work” (Bloxham et al 2018). The same term is commonly used when this process is conducted *internally* within an institutionally defined marking team.

The principles remain the same, whereby the power of peer discussion is harnessed to reach a common understanding of standards and thresholds among markers based upon discussion of the marking of a sample³ of live or historic scripts or other assessment artefacts in a calibration event prior to the process of marking and moderating the full cohort⁴.

This helps to induct those who are new to this type of assessment and/or to the marking team, as well as provide opportunity for re-calibration for experienced colleagues.

Examples of how Lancaster colleagues' conduct such events are available⁵.



It is important markers are clear about which Intended Learning Outcomes this assessment enables students to demonstrate. This supports discussion and agreement about the focus, style and amount of feedback. Further guidance on assessment feedback is available¹.

of consistency, discrepancy and query with regards to academic standards.

⁴ Where historic examples are available this activity “may take place at times separated from the normal deadlines of marking student work.” (Sadler, 2013)

⁵ <https://www.lancaster.ac.uk/ceda> <needs to be created>

The calibration event is also an opportunity to clarify what to do if a piece of work is submitted that is outside the marking scheme e.g. passing on to the module convenor.

Similarly, to ensure that everyone is clear about what steps to take in cases of poor academic practice or deliberate attempts to pass off others' work as the students own (Nuttall, 2007). The lone marker

If you are not part of a pair or team of markers, you may find it helpful to review a sample of the previous years' marking and feedback, particularly if this is your first experience of marking this piece of assessment work.

How should I actually go about marking?

Following a calibration of standards event, the use of the mark scheme, rubric or marking criteria is helpful in assuring inter- and intra-marker reliability (Bloxham et al 2011). This is particularly helpful for large amounts of marking which is likely to happen over more than a single sitting and/or with a greater number of colleagues in the marking team.

Suggestions for managing the marking load include:

- Where there are multiple questions, mark the same question across all scripts, rather than each script in full as "this is faster and more reliable" Brown *et al.* (1997)
- Have, or develop, a comment bank for common points of feedback to share among all markers on the same module to build consistency and help new markers
- Leverage the technology⁶
- Plan your time to meet the University commitment for return of work
- Check for any departmental requirements on timings e.g. for administrative colleagues to process the assignments
- Review how the marking has gone in practice to inform next year's approach

Reasonable adjustments

Reasonable adjustments as defined by a student's Inclusive Learning Support Plan (ILSP) may also apply to marking of work and an assessment cover sheet will provide further guidance to markers and must be taken into account.

Holistic or criterion based marking

More experienced markers will have a 'sense of' the criteria as a whole and may complete the marking sheet after deciding on the appropriate mark (see Ashwin, 2011 and Bloxham and Boyd, 2011). This is where a check and balance is provided by the **moderation process**.



We need to talk about moderation...

For this briefing paper, we are considering the situation where moderation occurs *after* assessment marking has taken place⁷.

It is important because it ensures that there are not significant variations in application of the marking criteria, resulting in similar pieces of work achieving very different results (Nuttall, 2006). Having transparent moderation processes in place is also "likely to increase students' confidence in marking" (ibid.) and potentially reduce the number of queries that may arise from perceived bias or unfairness, as well as helping staff to feel confident about their judgement in the face of such queries.

Moderation takes place in two contexts, that of internal moderation, possibly across a marking team, and external moderation, usually by the External Examiner.

Internal moderation should take place before students receive provisional marks. These marks receive final ratification by the external examiner at an assessment board.

The following section outlines the four options for moderation according to Lancaster University's Manual of Academic Rules and Procedures, General Regulations for Assessment and Award GR 2.4.2 (pg 9).

⁶ Further guidance from [Embrace Digital](#)

⁷ Moderation by colleagues and external examiners of exam questions at the time they are set is also a common HE practice.

Moderation options

The marking team will need to be clear about which moderation option has been selected at the assessment design stage and what this means for their marking and moderation process.

Unseen Double marking

Unseen double marking, where student work is independently assessed by a second marker without the knowledge of marks assigned by the first marker;

Second marking

Second marking, where student work is assessed by more than one marker, but the second marker knows the mark allocated by the first marker

Sampling

Sampling, where second markers review a representative sample of work first-marked by other colleagues for the purpose of checking the consistent application of marking criteria and moderating marks awarded (a sample from a collection of n scripts should involve five scripts or the square root of n scripts, whichever is the greater); where more than one marker is involved, the square root rule should apply separately to each marker).⁸

Thought needs to be given as to who will do this review of first marked work. It could be someone who is part of the marking team with more experience, for example the course convenor moderating the marks and feedback from their team of tutors. Alternatively, it could be someone from the department who has not been directly involved with the marking so far.

Trends

Analyses of marking trends, where work is marked by only one marker, undertaking a comparative analysis of marking trends to compare individual students' consequential marks on an individual course with their average mark on all their other courses.

Resolution of marks

After moderation has taken place there may be a need to resolve any significant differences in grading where double or second marking is used.

“GR 2.4.3 [...] departments must follow a clear procedure for determining final marks and grades where the two markers are in disagreement, and there must be a clear audit trail to show how the final mark or grade was reached.” (pg10)

What this procedure is in practice can be determined and documented locally to provide the required “clear audit trail”.

External moderation

Every programme at Lancaster University will have an External Examiner and further information about the selection, appointment and duties is detailed on the Academic Standards and Quality website (ASQ, 2022). Crucially, their role is to provide “impartial and independent advice and comment on the institution's standards, and on students' achievement in relation to those standards” (ibid.). External examiners will see a sample of marked work, including all fails for all modules for which they are responsible.

To what extent do your students understand your marking and moderation processes?

Lancaster University's General Regulations for Assessment and Award states “GR2.4.1 There will be agreed grading and marking criteria for all types of assessment and these will be made available to students at the appropriate times.” (pg9)

Is there more we can do than simply make grading and marking criteria “available”?

Actively engaging students with the assessment criteria as part of our teaching strategy, asking them to apply the criteria to a number of pieces of work, has been shown to significantly improve the students' performance when they subsequently undertake the task (Price et al., 2001). In addition, “this improvement continues in other similar pieces of work in which similar criteria are used at least a year later.” Rust (2002).

⁸ For example taking each marker's N scripts, the square root of 16 scripts is 4 and so 5 scripts must be sampled; square root of 100 scripts is 10 and so 10 scripts must be sampled;

Could such opportunities be designed into your programme curriculum?

Vignettes of practice from across Lancaster University are available to share ideas and the range of existing practice for calibration activities amongst marking teams and with students. There is also advice from [Advance HE Calibration of Academic Standards project](#).

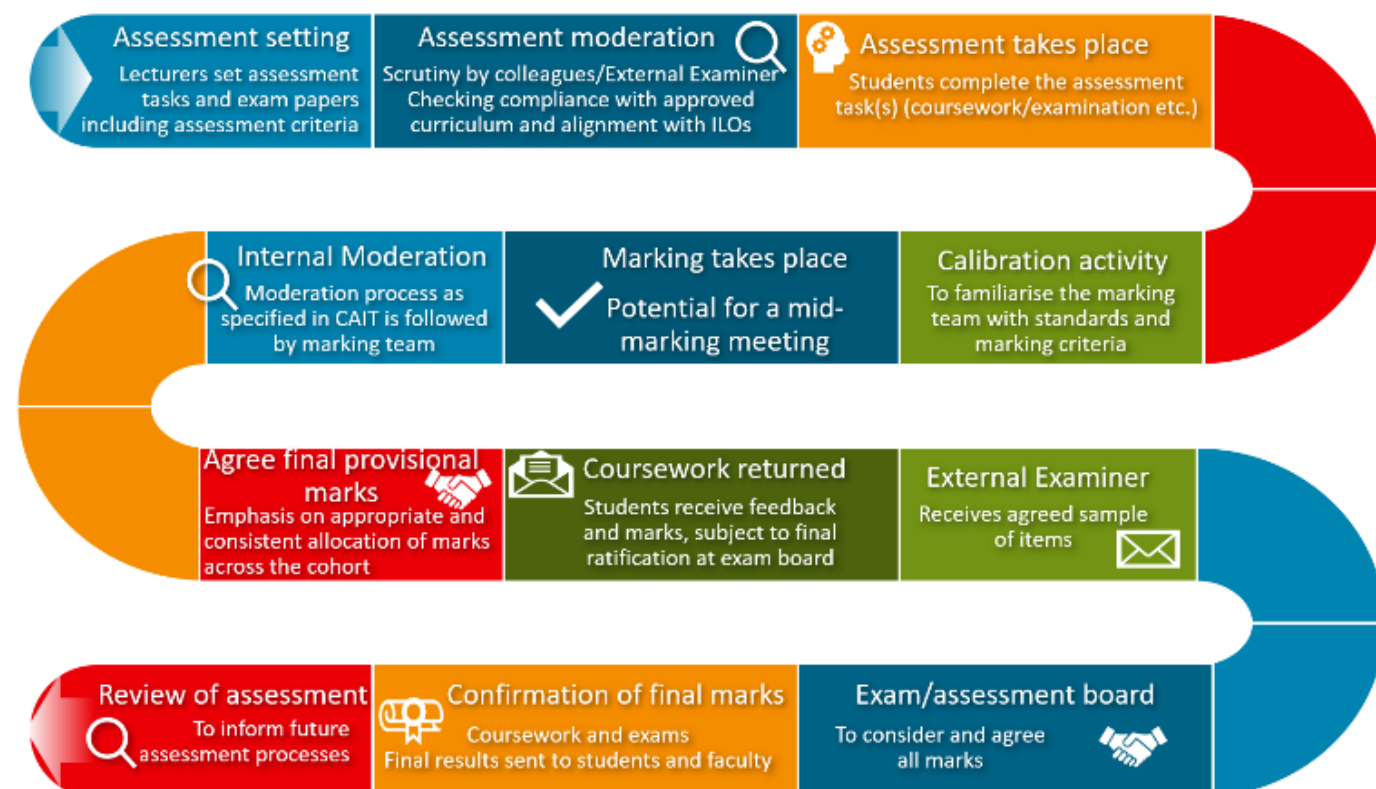
The end of the process?

Returning marks and feedback to students may appear to be the end of the process of marking and moderation.

However, this can also be a good opportunity to consider any implications this has for:

- this cohort and subsequent modules
- teaching and assessment of future cohorts
- your marking and moderation process and practices

Indicative roadmap of assessment setting, marking and moderation



What is your process?

To help you/your marking team understand the various steps in your process for marking and moderation, an indicative 'roadmap' is shown below to prompt discussion about when, where and how some of these steps take place in your context. Note that this version of the map has calibration taking place with live, rather than historic, scripts, which may allow calibration to occur at an alternative time for the marking team.

It is also an opportunity to consider when and how you inform students about and/or engage them in, the process. [Click here for some examples of how others do this.](#)

Scholarly Underpinning

Past and current efforts to ensure integrity of assessment marking – in the UK sector and LU – focused first on development of sector standards via shared texts such as [Subject Benchmark Statements](#) and the Framework for Higher Education Qualifications (FHEQ), plus institutional programme specifications, learning outcomes, and assessment criteria / grade descriptors. These documents are then interpreted through processes of first marking and moderation by internal and external examiners.

The current approach has limits, neatly summarised by the Higher Education Academy (HEA) in 2018:

‘... although the sector has produced vast quantities of agreed written standards, achieving a shared interpretation of their meaning is a very different matter. Repeated studies over many years demonstrate considerable inconsistency in academics’ judgements about student performance and variation in the meaning they accord to written standards. Studies of external examiners have found similar inconsistency. This is not a criticism of examiners but a recognition that the language of standards always needs a level of interpretation and individuals differ in the meaning they accord to them. Such fluidity in standards leaves the sector open to charges of grade inflation as institutions reference sector norms (for example, proportion of firsts) rather than agreed national standards.’
(Bloxham, S., Reimann, N. and Rust, C. (2018))

In an effort to reduce variability within and between institutions and subject communities the HEA (now part of AdvanceHE) developed a [Degree Standards Project](#) in 2019, which among other things explored and promoted the use of calibration.

Calibration can be defined as follows:

‘Calibration within the higher education sector is an approach that aims to ensure consistent standards for judging the quality of student work. A ‘calibrated’ academic is able to make grading judgments consistent with those of calibrated academics in other institutions across the UK. The aim of calibration is to achieve comparability of academic standards across institutions and stability of standards over time.’

(Bloxham, S., Reimann, N. and Rust, C. (2018))

The rationale for this approach was articulated by the Higher Education Quality Council (HEQC) in 1997:

‘Consistent assessment decisions among assessors are the product of interactions over time, the internalisation of exemplars, and of inclusive networks. Written instructions, mark schemes and criteria, even when used with scrupulous care, cannot substitute for these.’ (Higher Education Quality Council (HEQC), 1997, quoted in Bloxham, S., Reimann, N. and Rust, C. (2018))

While calibration in the literature is predicated on *extra*-institutional subject-defined groups jointly agreeing academic standards, this term can also be used to refer to the process when conducted *internally* within an institutionally defined marking team. The principles remain the same, whereby the power of peer discussion is harnessed to reach a common understanding of standards and thresholds among markers based upon discussion of the marking of a sample of live or historic scripts of other assessment artefacts in a semi structured environment prior to the process of marking and moderating the full cohort.

Sadler (2013) makes the case for the shift from moderation towards calibration, and there is research evidence that appears to support the effectiveness of such approaches. For example, O’Connell et al (2016) observed the impact of consensus moderation on marking practices in Accounting, and found that it both decreased variability, and improved marker confidence in judgements about standards. Further, a study by Mason & Roberts (2023) emphasises that a safe environment needs to be fostered to encourage openness and dialogue in consensus moderation to reduce power differentials between markers, particularly when disagreement on mark allocation has been identified.

In another study, Bamber (2014) involved postgraduate students in marking calibration exercises. This also had the effect of achieving common understanding of standards amongst markers – in this instance, both academics and students – and from a student perspective heightened their understanding about how judgements were being applied to their work, improving their confidence in the marking system.

References and further reading

AdvHE (2019) [Calibration of Academic Standards | Advance HE \(advance-he.ac.uk\)](#)

Ashwin et al (2011) Chapter 13: Assessment does it make a contribution to learning? In *Reflective Teaching in Higher Education*, Bloomsbury

ASQ (2022), External Examiners pages, <http://www.lancaster.ac.uk/academic-standards-and-quality/external-examiners> accessed 23/03/2022

ASQ (2022) General Regulations for Assessment and Award, Manual of Academic Rules and Procedures <https://www.lancaster.ac.uk/media/lancaster-university/content-assets/documents/student-based-services/asq/marp/General-Regs.pdf> accessed 31/03/2022

Bamber, M. (2015) The impact on stakeholder confidence of increased transparency in the examination assessment process. *Assessment & Evaluation in Higher Education*. 40(4), pp. 471-487

Bennett, R. and Kane, S. (2014) Students' interpretations of the meanings of questionnaire items in the National Student Survey. *Quality in Higher Education*. Vol. 20, No. 2, pp. 129–164

Bloxham, S., & Boyd, P. (2007). *Developing effective assessment in higher education: A practical guide*. McGraw-Hill Education (Chapter 6)

Bloxham, S., Boyd, P. & Orr, S. (2011) Mark my words: the role of assessment criteria in UK higher education grading practices, *Studies in Higher Education*, 36:6, 655-670, DOI: [10.1080/03075071003777716](https://doi.org/10.1080/03075071003777716)

Bloxham, S., Hughes, C. and Adie, L. (2016) What's the point of moderation? A discussion on the purposes achieved through contemporary moderation practices. *Assessment & Evaluation in Higher Education*, 41(4) pp. 638-653

Bloxham, S., Reimann, N. and Rust, C. (2018) *Calibration of Standards: What, Why and How?* Higher Education Academy.

Herbert I.P., Joyce, J. & Hassal, T. (2014) "Assessment in Higher Education: The Potential for a Community of Practice to Improve Inter-marker Reliability," *Accounting Education*, Taylor & Francis Journals, vol. 23(6), pages 542-561, December.

Mason & Roberts (2023) Consensus moderation: the voices of expert academics, *Assessment & Evaluation in Higher Education*, DOI: [10.1080/02602938.2022.2161999](https://doi.org/10.1080/02602938.2022.2161999)

Nuttall, M. (2007) Chapter 8: Moderation: towards greater consistency in marking in Bloxham, S. & Boyd, P. *Developing Effective Assessment in Higher Education: a Practical Guide*, McGraw-Hill Education, 2007. *ProQuest Ebook Central*, <http://ebookcentral.proquest.com/lib/lancaster/detail.action?docID=332673>

O'Connell, B., De Lange, P., Freeman, M., Hancock, O., Abraham, A., Howieson, B. and Watty, K. (2016) Does calibration reduce variability in the assessment of accounting learning outcomes?' *Assessment and Evaluation in Higher Education*, 41(3), pp. 331-359

Office for Students (OfS), NSS results 2021, [NSS data: provider-level - Office for Students](#) accessed 6th April 2022

Pokorny and Warren (2021) *Enhancing Teaching Practice in Higher education*, 2nd edition

Rust (2002) The Impact of Assessment on Student Learning: How Can the Research Literature Practically Help to Inform the Development of Departmental Assessment Strategies and Learner-Centred Assessment Practices? *Active Learning in Higher Education*; 3; 152

Sadler, D. R. (2013) Assuring academic achievement standards: from moderation to calibration. *Assessment in Education: Principles, Policy and Practice*. 20(1) pp. 5-19

UK Quality Code Advice and Guidance: Assessment [Assessment \(qaa.ac.uk\)](#)

Watty, K., Freeman, M., Howieson, B., Hancock, P., O'Connell, B., de Lange, P., & Abraham, A (2014) Social moderation, assessment and assuring standards for accounting graduates, *Assessment & Evaluation in Higher Education*, 39:4, 461-478, DOI: [10.1080/02602938.2013.848336](https://doi.org/10.1080/02602938.2013.848336)