

# Calibration Toolkit for Musical Recitals: Introduction

Welcome to the Musical Recitals Toolkit, which is designed to assist you as a music academic and external examiner in calibrating your academic standards with those of your subject community. The toolkit:

- supports you in discussing and creating a shared understanding of the criteria used to assess student performances, leading to greater consistency in your understanding and application of academic standards
- refers you to published national reference documents for academic standards as you make your judgements.

You can use the toolkit within and across departments, and while you are an aspiring or experienced external examiner.

The overall aim of calibration is to enable you to make consistent grading decisions that are:

- commensurate with the respective levels of academic achievement they represent
- comparable across course and institutional boundaries at any one point and over time
- consistent with disciplinary and societal expectations of higher education graduates.

## Background

A recent review of external examining (HEFCE, 2015) noted there is a formal requirement and a public expectation that UK awards are comparable, at least at the level of threshold academic standards. This public expectation illustrates the increased interest in university assessment practices seen in recent years as the profile of UK awards has shown considerable upward drift, particularly in the proportion of first-class degrees that are awarded. While changes in mark profiles are likely to be the result of a complex range of factors, consistent marking that reflects national standards is vital for the reputation of UK higher education, fairness to students and the confidence of employers and professional and statutory bodies.

Internationally, there is a similar emphasis on maintaining standards. An Australian study reports that a 'Reliable, consistent assessment process that produces comparable assessment grades between assessors and institutions is a core activity and an ongoing challenge with which universities have failed to come to terms' (O'Connell et al., 2016, p. 331).

In the UK it is recognised that external examiners are part of the quality assurance measures designed to safeguard academic and quality standards. However they typically provide the only external quality assurance activity designated to consider actual student achievement. The Finch Report on external examining (2012, p. 5) considered that external examiners 'ensure that there is benchmarking against comparable programmes at the point when students are being awarded their qualifications'.

However, the recent HEFCE review (2015, p. 88) was much less confident about the contribution of the external examining system to safeguarding standards because:

'The assessment decisions of external examiners and others in the higher education setting exhibit very poor reliability in applying standards to student work, although examiners are often unaware of this; It is not possible to explicitly state most academic standards so they are interpreted similarly by all users although often examiners, other academics and those managing quality assurance may assume that explicit standards are genuinely shared;(with the proviso of b) Some examiners do not make use of published reference points for standards.'

The 2015 review drew on the research evidence regarding assessment, marking and external examining to make recommendations about how to improve the contribution of external examining to assuring academic standards. It argued that there is an urgent need to promote a shared understanding of what constitutes quality in student work, and that we cannot rely on written statements of standards (e.g. assessment criteria, qualification descriptors and benchmarking statements) to assure this consistency. We need to interpret the key terms used in these written statements consistently across markers. Dialogue provides a means through which assessors can establish a common vocabulary and set of meanings in relation to the criteria they are using to judge the academic standards of pieces of work. The intent with calibration discussions is to share the rationales for making judgements about quality, identifying meanings-in-use for the principal explanatory terms. Consequently, the 2015 review recommended regular calibration of external examiners' standards to take place in disciplinary communities across the range of institutions where programmes of a similar nature are studied. This recommendation is similar to the conclusion of a HEQC report on external examining from over 20 years ago.

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 (HEQC, 1997).

## Purpose of the toolkit

This toolkit is designed to support the calibration process in relation to final-year undergraduate music recitals. It aims to:

- raise awareness among music academics and external examiners about the potential variation in academic standards
- build greater consistency in the judgement of student work among music academics and external examiners both within and across programmes and institutions
- discuss and create a shared understanding of criteria for the assessment of recitals, drawing on key reference documents, particularly the [Subject Benchmark Statement for Music and the Framework for Higher Education Qualifications](#).

The intended outcomes for those participating in the calibration process are:

- greater awareness of differences in academic standards across assessors
- greater capability and confidence when drawing on accepted academic standards in their assessing and external examining roles
- a shared interpretation of agreed key criteria for the assessment of recitals, with reference to relevant national standards statements.

## Overview of the process of calibration

The approach to calibration described here takes the form of a workshop. Before the workshop, delegates are asked to independently mark and briefly comment on four final-year recitals (8–10-minute recitals of piano, horn and popular music, and a 30-minute programme of pieces on a cello), then upload their marks and comments anonymously.

An agreed judgement of these assignments has been made by a group of 24 music assessors from across all UK conservatoires and several university music departments in advance of the calibration workshop (called the 'agreed grades'), but these marks are withheld from the delegates at this stage. Before the workshop the facilitator compiles an overview of the results. At the workshop small groups attempt to arrive at a consensus grade for the recitals related to the learning outcomes as expressed in the Subject Benchmark Statement for Music, noting the characteristics of the recital that were particularly influential when it came to making decisions at the borderline.

Finally, the small group grades and provides comments and these are shared in the whole group along with the agreed grades. The facilitator manages a cross-group discussion aimed at achieving consensus about the grade for each recital and the influential characteristics.

These above stages (small group and whole group) are repeated for recitals 2, 3 and 4 (depending on the time available).

The workshop concludes with a summary of its achievements, encouraging delegates to consider the implications of the day for their own standards and use the process with other assessment types in music, and in other forums such as their own or an examining department.

## What the toolkit contains

- A set of amendable PowerPoint slides for running the workshop.
- The detailed programme for running the online preparatory stage and the workshop, including use of software, timings, instructions to delegates, resources and when to use the PowerPoint slides. This also includes the contextual information regarding the recitals – the module learning outcomes (Appendix A).
- Three videos: 'Why Calibration?', 'Running the Calibration' and Jane Ginsborg
- Four videos of final-year undergraduate recitals.
- An agreed mark for each recital as determined by a group of experienced music assessors, and an explanation for the marks awarded. These relate to the contextual information and external reference points.
- Group panel marking template.

## References

Baume, D. et al. (2004) What is happening when we assess, and how can we use our understanding of this to improve assessment? *Assessment and Evaluation in Higher Education*, 29(4), 451–477.

Bloxham, S., den-Outer, B., Hudson, J. & Price, M. (2015): Let's stop the pretence of consistent marking: exploring the multiple limitations of assessment criteria. *Assessment & Evaluation in Higher Education*, DOI: 0.1080/02602938.2015.1024607.

Brooks, V. (2012) Marking as judgement. *Research Papers in Education*, 27, no. 1, 63–80.

- Elander, J. and Hardman, D. (2002) An application of judgement analysis to examination marking in psychology. *British Journal of Psychology* 93, 303–328.
- HEFCE (2015) A review of external examining arrangements across the UK. Bristol; HEFCE
- HEFCE (2016) Revised Operating Model for Quality Assessment, p. 36
- HEQC (1997) Graduate Standards Programme: final report. London: Higher Education Quality Council.
- Laming, D. (1990) The Reliability of a Certain University Examination Compared with the Precision of Absolute Judgements. *Quarterly Journal of Experimental Psychology* 42A (2), 239–54.
- Leach, L., Neutze, G. & Zepke, N. (2001) Assessment and empowerment: some critical questions. *Assessment and Evaluation in Higher Education*, 26(4), 293–305.
- O’Connell, B. et al (2016) Does Calibration Reduce Variability in the Assessment of Accounting Learning Outcomes? *Assessment and Evaluation in Higher Education*, v41 n3, 331–349.
- O’Hagan, S.R. and Wigglesworth, G. (2014) Who's marking my essay? The assessment of non-native-speaker and native-speaker undergraduate essays in an Australian higher education context. *Studies in Higher Education*, DOI: 10.1080/03075079.2014.896890.
- Read, B., Francis, B. and Robson, J. (2005) Gender, bias, assessment and feedback: analyzing the written assessment of undergraduate history essays. *Assessment and Evaluation in Higher Education*, 30(3), 241–260.
- Sadler, D.R. (2012): Assuring academic achievement standards: from moderation to calibration. *Assessment in Education: Principles, Policy & Practice* 20(1), 5–19, DOI:10.1080/0969594X.2012.714742.
- Shay, S.B. (2004) The Assessment of Complex Performance: A Socially Situated Interpretive Act. *Harvard Educational Review* 74, no. 3, 307–329.
- Thornes, J.E. (2012) External examiners and the continuing inflation of UK undergraduate geography degree results. *Area*, 44(2), 178–185.
- Universities UK, Guild HE. (2011) Review of external examining arrangements in Universities and Colleges in the UK; final report and recommendations (Finch report). London: Universities UK.

## Further reading

- Bloxham, S. and Price, M. (2013) External Examining: Fit for Purpose? *Studies in Higher Education*. doi: 10.1080/03075079.2013.823931.
- Higher Education Academy (2017) Assessment – a briefing paper. Prepared for the Professional Development Course for External Examiners.
- Newstead, S. E. and Dennis, I. 1994. Examiners examined: The reliability of exam marking in psychology. *The Psychologist* 7, 216-219.

Music Subject Benchmark Statement [http://www.qaa.ac.uk/docs/qaa/subject-benchmark-statements/sbs-music-16.pdf?sfvrsn=1f9af781\\_10](http://www.qaa.ac.uk/docs/qaa/subject-benchmark-statements/sbs-music-16.pdf?sfvrsn=1f9af781_10)  
Framework for Higher Education Qualifications (FHEQ)  
<http://www.qaa.ac.uk/docs/qaa/quality-code/qualifications-frameworks.pdf>