Case study: Calibration of standards in Geography

In partnership with: The Royal Geographical Society, with IGB
Purpose

This case study focuses on the calibration of academic standards for research-based coursework in Geography. The calibration activity described was undertaken as a collaboration between the Higher Education Academy (now Advance HE) and the Royal Geographical Society (RGS).

The case study has been created as part of the Degree Standards Project and is for those in higher education interested in the process of calibration, particularly: higher education providers, academics in subject departments or who are leading programmes seeking to calibrate standards, and professional, statutory and regulatory bodies (PSRBs).

Introduction

As part of the Degree Standards Project, the calibration of academic standards has been explored through an approach to calibration which involves the ‘social moderation’ of student work with groups of subject external examiners. Social moderation can be effectively used as a process for obtaining consistency in understanding and applying academic standards through discussion of examples (i.e. of student work) and drawing on external reference points (e.g. subject benchmark statements).

The project team has worked in partnership with exemplar disciplines to undertake calibration activity and produce resources describing processes, which other subject, discipline or professional areas can emulate. In the longer term, the anticipated impact of this approach to developing calibration is the establishment of a community of practice around degree standards, and the increased comparability of academic standards across subject communities.

The activity described here involved external examiners in Geography participating in a one-day workshop in May 2017 to pilot calibration activities in three key areas:

- Disciplinary data on degree outcomes
- Social moderation of student work
- ‘Toolkit’ – checklist and information (external benchmarks)

This case study describes the process for the second of these areas – the social moderation of research-based coursework. A key outcome of this work has been the development of a Calibration Toolkit for Geography (available on the Degree Standards Project website), which can be used to support a calibration process specifically in relation to research-based coursework essays and reports. This toolkit aims to:

- raise awareness among geography academics and examiners about the potential variation in academic standards across different assessors
• build greater consistency in the judgement of student work among geography lecturers and examiners both within and across programmes and institutions
• discuss and create a shared understanding of criteria for the assessment of research-based coursework essays or reports, drawing on key reference documents: the Subject Benchmark Statement for Geography and the Frameworks for Higher Education Qualifications.

Approach and process

The workshop involved eight external examiners representing a broad range of experience and institutions. The intended outcomes for the participants were:

• A greater awareness of differences in academic standards across different assessors.
• A greater capability and confidence in their assessing and external examining roles.
• A shared interpretation of agreed key criteria for the assessment of research-based coursework essays or reports with reference to relevant national standards statements.

The approach adopted for the calibration process involved two main stages as follows.

Stage A: assessing student work

In advance of the workshop, the external examiners were each provided with five example assignments – research proposals for final-year dissertation projects. They were asked to verify the task validity against the benchmark statement for Geography, assess each against the benchmark standard (‘below threshold’, ‘threshold’, ‘typical’, ‘excellent’) and grade the proposals using their own institution’s marking scheme. They then submitted the grades and a brief reason – anonymously – through a VLE area set up for this.

Stage B: workshop activity

At the workshop, participants were presented with the outcomes of the pre-workshop activity of assessing student work as data tables to illustrate the range of benchmark standards that had been awarded to each assignment (see Appendix) and the wide variation in terms of grades given. Small-group discussions were then held to consider these results, with the aim of achieving a consensus on the grades for each assignment and for the examiners to reflect on what influenced their decisions.

A whole-group discussion was then held with a view to achieving a consensus regarding the grading decisions. A list of the common characteristics that
influenced their judgements was pulled together, which was used to support them in achieving a consensus on the marks.

**Evaluative summary**

Participant feedback to the workshop indicated that the external examiners appreciated the opportunities to discuss experiences and practice in a subject-specific setting and gain an insight into other assessment and marking schemes.

*Most useful aspect was being able to debate the idea of standards with colleagues from very different institutions to my own – and to take this further into a consideration of what resources would be useful in trying to develop a more common understanding of academic standards across the sector.*

*I think everyone there found it a valuable experience and learned something that they could potentially take back to their own institution, as well as contributing. So there is a clear role for EEs in terms of sharing good practice.*

*I think that using the benchmark was really useful and I will refer to it more in my role. I already had a good knowledge of B7 and my role as a critical friend is perhaps more data informed now.*

It was clear that the participants found the piloted activity for calibrating standards helpful, as it facilitated discussion and supported exploration of standards, as highlighted through the following participant comments:

*Social moderation was really interesting and should be something that most departments do as ‘best practice’ within their peer-support mechanisms. Recognition of these elements of marking practices and culture would be useful to staff and external examiners alike.*

*In terms of actually changing habits of mind (or at least in making tacit decisions more explicit), the social moderation activity was excellent – but it also showed that it takes some time to fully discuss multiple viewpoints around work (and its context) to reach a consensus on a mark.*

Dr Stephanie Wyse,¹ who was the main lead for the workshop at the workshop, has pointed to the real sense of ownership of the issues and preparedness to engage. She has also highlighted that the approach to consider relevant disciplinary standards and build consensus for these takes time and skills in

---

¹ Previous Project Coordinator for Learning, teaching and accreditation at the RGS.
facilitation. At times, the discussion of the student work and associated standards was diverted by the difficulties in applying the subject benchmark. In particular, the benchmark standards (‘below threshold’, ‘threshold’, ‘typical’, ‘excellent’) do not map to grades or degree outcomes, but external examiners are asked to assess comparability across grades or degree classes. However, participants had clearly valued using student work, which enabled discussion on what influences marking.

Discussion also highlighted multiple perspectives on standards and how they are used. Do we use them in the context of reward, stretch, challenge etc.? I began to develop a more multifaceted understanding of what a ‘standard’ means.

It highlighted to me how much your own context influences the mark you award. It became much clearer to me that we don’t simply mark a piece of work. Rather, we assess the work in the context of the student journey and the support package we feel we are offering our students.

Appendix: Range of benchmark standards awarded

<table>
<thead>
<tr>
<th>Marker</th>
<th>Student 1</th>
<th>Student 2</th>
<th>Student 3</th>
<th>Student 4</th>
<th>Student 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marker 1</td>
<td>Typical</td>
<td>Threshold</td>
<td>Below threshold</td>
<td>Below threshold</td>
<td>Typical</td>
</tr>
<tr>
<td>Marker 2</td>
<td>Excellent</td>
<td>Typical</td>
<td>Threshold</td>
<td>Typical</td>
<td>Typical</td>
</tr>
<tr>
<td>Marker 3</td>
<td>Excellent</td>
<td>Threshold</td>
<td>Typical</td>
<td>Excellent</td>
<td>Typical</td>
</tr>
<tr>
<td>Marker 4</td>
<td>Threshold</td>
<td>Excellent</td>
<td>Typical</td>
<td>Typical</td>
<td>Threshold</td>
</tr>
<tr>
<td>Marker 5</td>
<td>Typical</td>
<td>Threshold</td>
<td>Threshold</td>
<td>Excellent</td>
<td>Typical</td>
</tr>
<tr>
<td>Marker 6</td>
<td>Threshold</td>
<td>Excellent</td>
<td>Below threshold</td>
<td>Typical</td>
<td>Threshold</td>
</tr>
<tr>
<td>Grade</td>
<td>59</td>
<td>59</td>
<td>38</td>
<td>90</td>
<td>45</td>
</tr>
</tbody>
</table>