CAS1 Calibrating Academic Standards Workshop Report

Friday 4 August, Griffith University, Gold Coast

Contents

Tables

Report prepared by: Penny Jose
15 August 2017

Introduction

The Meeting the Standard Peer-Review of Assessment pilot project established a sustainable sector wide model for peer review of assessment in tourism, hospitality and events (TH&E). The referencing process focussed on student achievement of the TH&E Learning and Teaching Academic Standards. However, the project found that there was a lack of a shared understanding and consensus about assessments and outcome standards in comparable units being taught at the same AQF level, but in different institutions. One of the recommendations of the project was to introduce calibration workshops to help reduce grader variability.

Calibration workshops have been found to reduce variability across assessors and build assessor confidence. Freeman reported that with just a single workshop, the standard deviation of markers' judgements could be halved (O'Connell et al 2015). Calibration: 'The' key for assuring learning when benchmarking (Freeman 2015) is a short (8 min) video which outlines the reasoning behind and process adopted in the *Achievement Matters* project for the Accounting Standards on which the first TH&E workshop was based.

In the national context, the requirement for external referencing and benchmarking for all courses came into effect on 1 January 2017 (Higher Education Standards Framework (Threshold Standards) Clause 5.3 (HESF)) (TEQSA, 2015). In the United Kingdom, a major review by the Higher Education Academy recommended a commitment to 'the maintenance of degree output standards and their **reasonable comparability** across the UK higher education system' (HEFCE 2015). In response, a Degree Standards project 2016-2021 will develop proposals to explore approaches to the calibration of marking practices, but not to develop common marking criteria for all providers. The body of material collected from calibration workshops will form the basis of the review of the 2015 Tourism, Hospitality and Events Learning and Teaching Academic Standards (LTAS) after five years.

Aims

The calibration process is a multistage intervention in which external peer reviewers examine an assessment task and samples of student work against one or more of the threshold learning standard domains before, during and after a calibration workshop (O'Connell et al 2015). In this case, the domain selected for calibration was *Problem Solving*.

The workshop and calibration process aimed to determine:

1. What is the quantifiable difference in grader variability on the assessment of learning outcomes in *Problem Solving*?
2. Does participation in the workshops lead to reduced disparity in the assessment of the students' learning outcomes in *Problem Solving*?
3. Does participation in the workshops lead to greater confidence by reviewers in their ability to assess students' skills in *Problem Solving*?

Method

Participant selection

The first Calibration of Academic Standards workshop (CAS1) was held at Griffith University, Gold Coast campus on the afternoon of Friday 4 August, immediately after the CAUTHE Mid-Year Meeting. An invitation to CAUTHE's 34 Chapter directors and Affiliate members to participate in the calibration of assessment process yielded 13 participants from 11 institutions; including nine universities, one technical and further education (TAFE) institution and one private provider. In addition, seven observers attended

the calibration workshop from two universities, one private provider and one accreditation agency (Appendix 1). The workshop aimed to calibrate one of the learning standards for *Problem Solving* for Bachelor level (AQF 7) tourism, hospitality and events graduates.

Review process

Prior to the workshop, participants were divided into teams and asked to review and rate the validity of a de-identified assessment task (in terms of its ability to assess *Problem Solving* at AQF 7) and two de-identified samples of student work for that assessment task. The participants provided feedback and justification for their ratings using the *Qualtrics* online survey system. Once all reviewers had entered their evaluations, the results of all reviewers and their comments were shared with all team members. The tool reported key metrics (highest, lowest, mean, median, mode, and standard deviation) as well as all verbatim comments. These anonymised results were published and circulated to participants prior to the workshop as a key resource so that each participant could compare their assessments with that of their peers. Participants reported that the pre-workshop review took between one to three hours.

The workshop involved participants meeting together to discuss and reflect upon their judgments and the justification for their assessment of the validity of the assessment task and the samples of student work. The aim being that this respectful sharing of results and the discussion about why they gave the grades they did will help the assessors achieve consensus in applying the *Problem Solving* standard. In groups of five or six, participants firstly established that the assessment task was valid for demonstrating achievement of the standard. They then repeated the consensus reaching process through open discussion, benchmarking the two pieces of student work against the *Problem Solving* standard. Finally, participants repeated the consensus-reaching activity with a new anonymised sample of student work; exploring and discussing their results and reasons until there was agreement on the standard for a graduate. The workshop process took two and a half hours to complete.

Timeline

The tasks, responsibilities and timeline for CAS1 is shown in Table 1.

Table 1: Calibration of academic standards timeline

Task	Resp	Date
Send invitation to join the workshop to CAUTHE Chapter directors & project participants	Penny	16 May 2017
Choose TLO(s) to calibrate (and the AQF level) & identify an assessment item and a range of (3) marked examples	Paul	18 May 2017
Set up a survey space (*Qualtrics)* where colleagues provide feedback on assessment items prior to the workshop	Penny	14 July 2017
Open pre-workshop survey in *Qualtrics &* share the items (*Dropbox*)	Penny	17 July 2017
Deadline for reviews	All	31 July 2017
Publish and circulate aggregated results	Penny	2 August 2017
Calibration of Academic Standards (CAS1) workshop	All	4 August 2017

The workshop agenda is attached in Appendix 2.

Results

Assessment task

General feedback was that the assessment task was innovative, creative and motivating for the students and a 'great' assessment item. Participants agreed that the task was valid for demonstrating achievement of the national standard for problem solving in tourism, hospitality and events at AQF level 7.

There appeared to be some confusion around the purpose of the rating scale, which is intended to rate the degree to which the task met the standard, rather than to grade the task. Despite a high standard deviation (SD) of 18 and a relatively high mean of 79 for the quantitative results (Table 2), the qualitative comments about the quality of the work were noticeably uniform and consistent. This result was confirmed in the workshop discussions, during which participants agreed that the task, which involved identifying a problem and proposing a solution using the literature to support the ideation, creativity and innovation processes, was a good method to assess problem solving. It would be useful at future workshops to ask participants to re-rate the assessment task post-calibration to confirm the calibration process had been effective.

The following specific feedback on the assessment task was provided:

- Clearly link the intended learning outcomes to the standard being addressed.
- Use the exact wording from the standards in the description of the domain for the fields of tourism, hospitality and events (as appropriate).
- Incorporate language from the standard in the marking rubric e.g. analysis, synthesis, evaluation.
- Ensure the weightings in the rubric reflect the importance of the domain being addressed.
- Ensure the weighting of the assessment task reflect the volume of work required by students.

Student work

There was an overall range from a low of 40 to a high of 90 in the ratings for student work with a mean of 73 and SD 14. The results show that despite the stated need for calibration, student work at either end of the rating scale, either excellent or poor, is obvious and declares itself. Student sample 1 was rated as a good piece of work, with a mean of 80, a lower SD 9 and gap between minimum and maximum ratings of 26. Work in the 'middle', such as student sample 2 with a mean of 67, is more difficult to judge and yielded a higher SD 15 and a range of 50 between the minimum and maximum rating.

There again appeared to be some confusion around the purpose of the rating scale, which is intended to rate the degree to which the samples of student work demonstrated achievement of the standard, rather than grading the work. After discussion, the participant who had given the low rating of 40 for student work 2, agreed that the work had met the standard and the rating should have been higher. The participant whose rating was significantly higher than other team members (90), realised they had been 'bedazzled by the technology' and a lower rating would have been more appropriate. While the students had been somewhat innovative, it was generally agreed that they had provided random ideas and solutions which appeared to be 'reverse engineered' to fit the problem. There was some evidence that they 'collected data', and undertook some analysis and synthesis, but clearly did not adequately 'evaluate' their solutions. Participants noted that it is important to be expert in both subject content and submission media (in this case a video) when assessing student work.

Table 2 shows pre- and post-workshop descriptive statistics (mean scores, range and standard deviations) on the assessment item and student work samples.

Table 2: Comparisons of reviewers' scores on problem solving pre- and post-calibration

Activity	Task	Min	Max	Mean	SD	Differ	Count
Pre-calibration	Assessment item	50	100	79	**18**	50	10
	Student work 1	65	91	80	**9**	26	10
	Student work 2	40	90	67	**15**	50	10
	Student work 1 and 2	40	90	73	**14**	50	10
Post-calibration	Student work 3	60	91	71	**10**	30	10

0-50 NA = Not Acceptable / 50-100 A = Acceptable

Post-calibration

Following discussion, participants were asked to review and rate assessment item student work 3. The work was rated in the middle of the range with a mean of 71. The SD 10 was significantly less than the SD 14 for the combined student work prior to calibration. These results represent a 29% reduction in SD for the combined student work, and a decrease in the gap between minimum and maximum ratings of 30. Further, it represents a 35% reduction in in SD when compared with student work 2 (which had a similar middle of the range mean of 73) and a gap between minimum and maximum ratings of 50. This tends to confirm the value in undertaking the calibration workshop to improve assessor consistency.

Reviewer confidence

Participants found the calibration workshop process valuable and enjoyable. They felt that they would be able to apply the learnings in their own subjects and would be keen to share the experience with their colleagues in their home institutions. Table 3 shows the confidence rating pre- and post-calibration increased from 42% to 56% for 'strongly agree', and decreased for 'somewhat agree' from 48% to 39% and 'neither agree nor disagree' from 10% to 6%. Participants agreed that as a result of being involved with the process they were generally more confident in rating:

- the capacity of assessment task requirements to allow students to demonstrate the national learning standard for problem solving in tourism, hospitality and events.
- students' problem solving ability benchmarked against the national standard.

The calibration workshop activity, including reaching consensus on the pre-workshop rating exercise and reflecting on additional student work in the context of the agreed national learning standards, contributed to an enhanced understanding of standards for problem solving that might apply locally.

Table 3: Results for pre- and post-calibration reviewers' confidence

Question (n=10)	Strongly agree	Somewhat agree	Neither agree nor disagree
Pre-calibration			
I am confident rating the capacity of assessment task requirements to allow students to demonstrate the national tourism, hospitality and events learning standard for problem solving.	6	4	0
I am confident that my feedback, explaining my ratings and offering suggestions, will be useful to the assessor.	3	6	1
I am confident rating tourism, hospitality and events student's problem solving ability as benchmarked against the national standard.	5	4	1
The pre-workshop activity, requiring me to reflect on the students work in the context of the agreed national learning standards, changed my understanding of standards for problem solving that might apply locally.	3	5	2
Having to enter my ratings as scale in the survey tool required me to formalise my judgement.	4	5	1
Having to write my feedback in the survey tool caused me to reflect on the reasons for my judgement.	4	5	1
Total	**25**	**29**	**6**
	42%	**48%**	**10%**
Post-calibration			
Following the calibration workshop, I am confident rating the capacity of assessment task requirements to allow students to demonstrate the national learning standard for problem solving in tourism, hospitality and events.	2	4	0
Following the calibration workshop, I am confident rating students' problem solving ability benchmarked against the national standard.	4	2	0
The calibration workshop activity, including reaching consensus on the pre-workshop rating exercise and reflecting on a third sample in the context of the agreed national learning standards, changed my understanding of standards for problem solving that might apply locally.	4	1	1
Total	**10**	**7**	**1**
	56%	**39%**	**6%**

Conclusions

The results found some wide variation in the ratings and possibly some confusion about the intended focus of the task. However, there were some rich comments and reflection in the comments in both the survey feedback and workshop discussions. The calibration process determined:

1. There was a quantifiable difference in grader variability on the assessment of learning outcomes in 'problem solving'.
2. Participation in the workshop lead to reduced disparity in the assessment of the students' learning outcomes in 'problem solving'.
3. Participation in the workshop generally lead to greater confidence by reviewers in their ability to assess students' skills in 'problem solving'.

Next steps

Participants agreed that calibration workshops should be held twice yearly at the CAUTHE annual conference and mid-year meeting. It was proposed that future calibration workshops might examine multiple domains. Participants were encouraged to 'volunteer' assessment tasks and samples of student

work that might be suitable for this purpose. The project team will develop a plan and timeline to implement this process so that all five domains at AQF levels 7 and 9 can be covered over the next five years. This program of activity will be held in conjunction with ongoing peer-review of assessment.

Acknowledgements

References

Freeman, M. (2015). Calibration: 'The' key for assuring learning when benchmarking. Retrieved 26 April 2017, from http://www.screencast.com/users/mark.freeman/folders/Snagit/media/78afe369-0284-4b3e-bdfd-853f7b72a8b3

O'Connell, B., De Lange, P., Freeman, M., Hancock, P., Abraham, A., Howieson, B., et al. (2015). Does calibration reduce variability in the assessment of accounting learning outcomes? Assessment & Evaluation in Higher Education. Retrieved from http://www.tandfonline.com/doi/full/10.1080/02602938.2015.1008398

Higher Education Standards Framework (Threshold Standards) (2015) TEQSA.

Appendix 1: Calibration of assessment participants

Participants were invited from CAUTHE Chapter and Affiliate member institutions, many of whom were involved with the previous Setting the Standard: Establishing Threshold Learning Outcomes (TLOs) for Tourism, Hospitality and Events Higher Education and Meeting the Standard for Tourism, Hospitality and Events projects.

Team 1

1. Associate Professor Pierre Benckendorff (The University of Queensland)
2. Dr David Beirman (University of Technology Sydney)
3. Dr Kelly Phelan (The University of Queensland)
4. Associate Professor Elspeth Frew (La Trobe University)
5. Dr Aaron Tham (University of the Sunshine Coast)
6. Dr Tamara Young (University of Newcastle)
7. Ms Sally-Anne Leigh (Academies Australasia Polytechnic) (pre-workshop review only)

Observers

8. Dr Margee Hume (Kaplan Business School)
9. Ms Anastasia Yeark (Kaplan Business School)

Team 2

10. Dr Paul Whitelaw (William Angliss Institute TAFE)
11. Dr Naomi Dale (University of Canberra)
12. Ms Marcela Fan (William Angliss Institute TAFE)
13. Assoc Prof Laurie Murphy (James Cook University) (by SKYPE)
14. Prof Marianna Sigala (University of South Australia)
15. Ms Esther Teo (Academies Australasia Polytechnic)
16. Dr Mieke Witsel (Southern Cross University)

Observers

17. Ms Pauline Tang (THE-ICE)
18. Ms Jane Gentle (THE-ICE)
19. Assoc Prof Erica Wilson (Southern Cross University)
20. Dr Anne Hardy (University of Tasmania)
21. Dr Dale Sanders (Edith Cowan University)

Project Manager

22. Mrs Penny Jose (CAUTHE Secretariat)

Appendix 2: Workshop agenda

Meeting the Standard: calibration and peer-review project
Calibrating Academic Standards Workshop 1: Problem Solving

Griffith University, Parklands Dr, Southport
G11_3.62 Learning Commons, Seminar Room
4 August 2017

Agenda

12.00pm – 1.00pm	Lunch	
1.00pm – 1:15pm	Welcome	
1.15pm – 2:00pm	Session 1	Assessment task validity
2.00pm – 2.15pm	Break	
2:15pm – 3.00pm	Session 2	Student work – samples 1 & 2
3.00pm – 3.30pm	Session 3	Student work (post) – sample 3
3.30pm – 4.00pm	Drinks	