

Adie, L., Lloyd, M., & Beutel, D. (2013). Identifying discourses of moderation in higher education. *Assessment and Evaluation in Higher Education*. <http://dx.doi.org/10.1080/02602938.2013.769200>

An investigation of moderation practices within a faculty of education

Denise Beutel, Lenore Adie*, Margaret Lloyd

Faculty of Education

Queensland University of Technology,

Brisbane, Australia

*E-mail address:

lenore.adie@qut.edu.au

ABSTRACT

Moderation of student assessment is a critical component of teaching and learning in contemporary universities. In higher education, moderation is usually governed by university-wide policies and practices. However, in Australia, moderation processes will now be guided by the new national university accreditation authority, Tertiary Education Quality and Standards Agency (TEQSA). In light of this reform, the purpose of this qualitative study was to identify and analyse current moderation practices operating within a faculty of education at a large urban university in eastern Australia.

Keywords: Moderation, higher education, assessment

1. INTRODUCTION

Moderation may be described as a process in which members of a teaching team develop shared understandings of assessment requirements, criteria, standards, and the evidence that demonstrates differing qualities of performance (Adie, Lloyd & Beutel, 2013). The purpose of moderation is to ensure that assessment aligns with established criteria, learning outcomes and standards; its processes are equitable, fair and valid; and judgements are consistent, reliable, and based on evidence within the task response (Adie, Lloyd & Beutel, 2011). The moderation process involves discussion of assessment tasks, criteria, standards and judgement decisions to ensure the validity and reliability of assessments, with the aim of improving the quality of the teaching/learning experience. In sum, moderation is a critical component of effective teaching and learning.

In Australia, moderation processes in higher education have been typically located within individual institutions with universities given the responsibility for developing their own specific policies and practices. However, with the introduction of the new national university accreditation authority, TEQSA (Tertiary Education Quality Standards Authority) (TEQSA, 2012) radical changes to moderation processes are being mandated. Under these new arrangements, universities will be required to declare:

details of moderation and any other arrangements that will be used to support consistency and reliability of assessment and grading across each subject in the course of study, noting any differences in these processes across delivery methods, delivery sites, and/or student cohorts (TEQSA, 2012, p. 30)

This reform is intended to move towards heightened accountability and greater transparency in

the tertiary sector, as well as entrenching evidence-based practice in the management of Australian academic programs. This formalising of systemic moderation of assessment in Australian universities is likely to upset a culture of practice in which moderation is part of the teaching and learning process but is not currently overt. This paper will describe and analyse current moderation practices being used in one Faculty of Education in an Australian university. Further, we will attempt to provide recommendations and guidelines for academics to navigate through the moderation process and to make decisions about practice which will suit their contexts.

2. LITERATURE

An investigation into any Australian university's policy and guidelines for assessment will reveal a set of processes and procedures outlining the conduct of moderation of assessments within the university. Yet, these guidelines on the need for and conduct of moderation have seemingly not countered the problems identified in the literature such as a lack of shared understanding and inconsistent application of criteria within and across courses (Sadler, 2010; Sanderson & Yeo, 2011). For example, while Van der Schaaf, Baartman & Prins (2011) found that when assessment criteria were available consistency in the judgement of student portfolios increased and uncritical acceptance of others' decisions decreased, Bloxham, Boyd and Orr (2011) found that experienced markers used criteria as a post-hoc validation for their holistic judgements. Furthermore, a study by Goos and Hughes (2010) found, through an online survey of 380 academics, that managerial accountability inhibited assessment practices with academics choosing to stay within safe and easily managed modes of assessment. Sadler (2010) identified the notion of a causal link between Learning Outcomes and assessment task standards descriptors as one reason for the inconsistencies with judgement decisions. He contends that Learning Outcomes and standards descriptors serve different purposes, and that each is prone to different interpretations by multiple users regardless of the specificity of description. Academics write increasingly detailed criteria yet fail to capture the essential qualities that will identify a particular standard of performance. One element missing from these discussions is the absence of leadership, particularly in the sharing of understandings and interpretation of standards, and the provision and revision of assessment criteria. A confident leader may identify ineffective practice and enact positive interventions. Our research sought to explore lecturers' understandings of the moderation process and their perceptions of what supported and inhibited their practice of moderation with the intent of contributing to the evidence of effective moderation practices.

3. RESEARCH DESIGN

This qualitative study was conducted within a Faculty of Education at a large university in eastern Australia. Funded by a university Faculty Teaching and Learning grant, the research was designed to investigate and analyse the moderation practices currently operating within the Faculty. The specific aim was to determine the different practices, processes and procedures of moderation that were being used, and to inform next steps in promoting efficient and effective moderation practices. The study was designed and conducted prior to the release of the new TEQSA requirement for moderation to be made explicit in university course documentation.

Semi-structured interviews were conducted with 25 academic teaching staff from a potential of 90 full time faculty members. The interviews included questions about the frequency, nature and topics of moderation discussions. Further interview questions focused on

how criterion-referenced assessment was used to inform the moderation process and on how consistency and comparability of assessment judgements could be improved within units in education courses. The participants included unit coordinators in core units in the undergraduate and graduate diploma teacher education programs across the faculty, as well as tutors and sessional academics. Some participants adopted differing roles and discussed more than one unit (a semester program of study) in the interviews. When categorised by role, the participants were unit coordinators ($n=21$) and tutors ($n=8$, including two sessional academic staff). Further, within the sample interviewed, there were some instances ($n=6$) where individual academics had sole responsibility for assessment and moderation within units. Details of the sample are provided in the table below. It is important to note that some units are offered in multiple courses.

Table 1: Description of sample

Course representation by unit

Bachelor of Education (Early Childhood)	Bachelor of Education (Primary)	Bachelor of Education (Secondary)	Graduate Diploma in Education			
			Early Years	Primary	Middle Years	Senior Years
11	9	8	2	3	2	5

As the aim of the research was to collect, collate and analyse a range of processes and procedures of moderation currently being used within the Faculty, the interview subjects were selected purposefully (Cohen, Manion, & Morrison, 2011) across a broad range of courses (here, a specialised degree) and units and to a lesser extent, across assessment types. We were most interested in differing instances of moderation, that is, where a unit coordinator worked with a number of tutors across campuses, where students from differing courses were enrolled, and where an individual had sole responsibility for the assessment and moderation within a unit. We also sought to represent atypical instances, for example, where (i) an integrated assessment item was offered across three units in one course; (ii) units were offered in multiple ways, namely, as core in one course but elective in another; (iii) students from different year levels were enrolled in the one unit; or (iv) units were offered in differing time periods, that is, over a semester or a shorter intensive block.

The interviews were transcribed and then analysed later by each researcher independently. The data were analysed iteratively with broad themes emerging after repeated readings of the data. This process brought inter-rater reliability to the findings.

4. RESULTS AND DISCUSSION

Four distinctly different moderation practices for the marking of written assessments were identified from the data. We describe these practices as: sample marking; blind marking; conference moderation and coordinator as arbiter (See Figure 1). In sample marking, all members of a teaching team individually marked an agreed number of submitted assessment items (perhaps 4-6) and later met and moderated with other tutors to develop consistent understandings of criteria and standards. Following this, the team members individually marked the remainder of the assignments allocated to them. Finally, as a team or with a buddy/peer, they revisited the highest and lowest grades or agreed benchmarks. This practice focused on understanding the criteria and relevant standards. In the case of small units, in which one person

had sole responsibility for teaching and marking student work, the advice of another expert in the field was sought to ensure the right standard was being applied and to assist with extreme and borderline grades.

The moderation practice of blind marking differed in that all members of a teaching team independently marked the same task without an awareness of the mark or grade awarded by other markers. The assignment marked was usually a task from a previous student cohort. After independent marking, the teaching team met to share the grade or marks awarded and to discuss the qualities they were valuing in the task. Following these discussions, team members individually marked the assignments of their own tutorial group. The high and low grades or agreed benchmarks were then revisited and reviewed with a buddy/peer. Pairings were often assigned by the unit coordinator and were made between experienced and inexperienced staff. It is of interest to note that, while some tutors expressed appreciation for the guidance that they received, there was also some concern with blind marking because it was seen to be disconcerting and at times led to confrontations between tutors. This practice tended to focus on inconsistencies rather than points of agreement, and tutors felt that they were being judged rather than the assessment task.

The third moderation practice was termed conference moderation. In this practice, each member of a teaching team marked all assignments in his/her tutorial group and then brought a representative of each grade or agreed benchmarks to a team meeting. Although a variety of practices were identified, this was the most common notion and practice of moderation, yet it also revealed many issues with moderation as a consensus forum rather than a rigorous analysis of the qualities that were being valued as indicative of a standard.

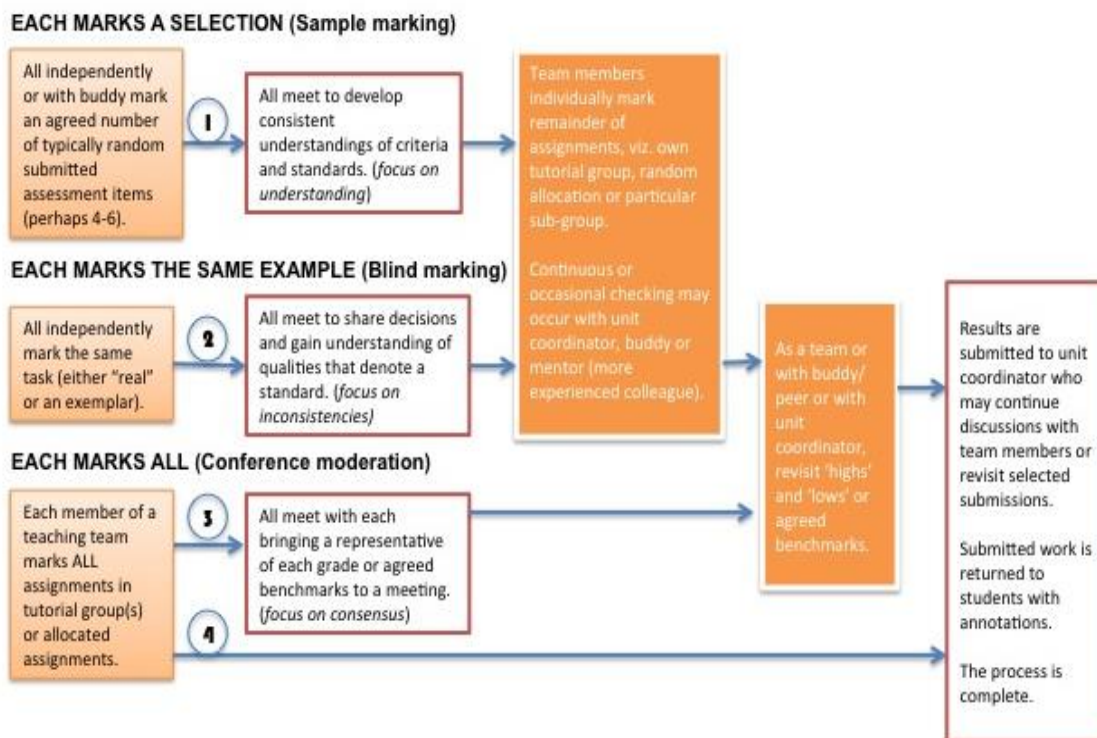


Figure 1. Moderation Practices in the Faculty of Education (2011)

In the fourth moderation practice that was identified, the unit coordinator acted as arbiter. In one instance, the markers were not part of the teaching team and had no connection to the unit other than through the assessment item. The unit coordinator worked individually with the markers on a continuous basis with frequent checking of student work. The focus was on consistency and fairness. In another instance, the teaching team were asked to submit a spreadsheet of results to the unit coordinator who then perused the distribution of the awarded grades. The focus of this practice was on the achievement of a normative outcome (bell curve) rather than on examining the quality of student work. The final example was the moderation of the “top” grades, the “fails,” and the “borderline” grades where the peer tutor or the unit coordinator “signed off” on the awarded grades.

The moderation of oral presentations and examinations were identified by the participants as especially problematic to moderate. To counter problems in moderating oral presentations, participants discussed efforts to include more than one of the teaching team observing the presentation, as well as recording it for later consideration. Most of the participants discussed taking detailed notes during presentations and stated that well-constructed criteria further assisted with this process. Examinations were not typically moderated with the reason given that the timeframe for completing and submitting final grades was too short. Where moderation of examinations did occur, it took the form of double marking of “failing” students’ scripts. It would seem also that moderation was on the assessment item, that is, the exam questions rather than focussing discussions on the qualities within student responses. Detailed marking guides were also provided to teaching teams to assist with marking examinations.

The clarity of assessment criteria was identified as a critical element in the moderation process. In the larger units, the ramifications of ambiguous or poorly constructed criteria for assessment were significant and impacted on both the teaching staff and students. Problems also occurred when tutors placed emphasis on different aspects of the assessment task. In smaller units, that is, those run by an individual or a small team, the criteria sheet frequently became a de facto moderation panel and the final arbiter in any determination of grades. Further, the criteria were used in follow-up conversations with students regarding the assessment of their work. Reference was made in the interviews to the usefulness of annotated exemplars to support an understanding of a standard. This had the effect of encouraging members of a teaching team to draw on evidence (founded in the criteria and standards descriptors) to support their judgements as opposed to subjectivity or tacit beliefs.

Criterion referencing clearly has a key role in the constructive alignment (Biggs & Tang, 2007) of assessment and broader learning outcomes (unit, course, university/graduate capabilities). A few participants made tacit reference to this alignment in terms of the need for a common understanding of the assessment task, how this links with the criteria sheet, and the teaching strategies employed in tutorials which includes the messages provided to students regarding the assessment task. It was, however, not widely referred to as part of the moderation process, nor with regard to understanding the standard required for a particular year level within a course.

4.1 Informing the practice of moderation

From our analysis and from the silences, omissions or contradictions in the data, we identified a number of areas which require attention as we work towards improving the practice of moderation. As Sadler (2010) identified, we also found that moderation was viewed mostly in terms of an assessment item with little reference to whole unit (subject) moderation or the alignment of assessment standards across a course. The determination of the final grade tended to be a numeric calculation rather than an on-balance judgement in relation to the unit learning

outcomes suggesting an analytic rather than a holistic approach to judgement making for an overall grade.

Definitions of moderation and descriptions of practice that appeared overly concerned with marks/grades suggested the residual effects of a system that previously emphasised normative and performative outcomes. References to a “bell curve” or marking to a grade distribution were made by some participants. This was typically associated with conversations of not wanting to be called to justify results, particularly awarding “too many” grades of Distinction or High Distinction. Unit coordinators used standard deviation and distribution of marks within and across tutorials to call for adjustments of student grades. While the distribution of marks can provide insight into the standard being applied to marking by a tutor, it is important that this information is understood as only part of the story, and that other factors must be considered before grades are adjusted. Grades must be viewed in terms of the quality of work, and opportunities for success provided by the tutor.

The participants revealed a number of difficulties in enacting moderation processes. These related to: time; the teaching team; sessional staff and working across different student cohorts. The time required for quality moderation discussions to take place was the most frequently cited difficulty. This was particularly evident in large units in which the teaching team members often worked across different campuses and/or sessional staff were involved thus finding a suitable time and place for the whole team to meet proved challenging. Developing shared understandings of assessment and standards are crucial to effective moderation processes and take time to establish (Sadler, 2010). Changes to the teaching teams and differences in the teaching and assessment teams required much time and effort each semester to build shared understandings. A further complication arose when new teaching team members were inexperienced as these novices required induction into the culture of university assessment as well as the unit assessment and standards. Participants identified that, in some instances, markers did not teach in the units and, as such, were not involved in the establishment of shared understandings of assessment and standards throughout the semester. The use of sessional staff presented issues in developing a shared culture of assessment. Several participants commented that payment of sessional staff for moderation meetings meant that meetings times needed to be limited which inhibited extended and meaningful conversations around assessment.

Working across different student cohorts provided another identified challenge to effective moderation. For example, it needed to be ensured that a similar standard applied to all groups according to the criteria sheet and that values relating to knowledge of a student, for example, the effort put into the assessment or the background of the student was irrelevant with regard to the assessment grade. When consideration of elements beyond the criteria sheet was given, this created an ad hoc system of assessment within a unit resulting in equities and rendering the criteria sheet that was meant to be guiding the moderation practice, obsolete.

5. CONCLUSION

Our study revealed that there is at present, liminal understanding of moderation as an integral part of teaching and learning, and differentiated understanding as to why or how moderation should occur and how circumstances may affect the type of practice adopted. We believe that the steps towards heightened accountability through the new TEQSA requirements for universities need to be responded to with informed practice that enhances the teaching and learning processes within higher education. As a way forward, we propose that moderation

needs to be understood as occurring through conversations before, during and after assessment. The moderation process must be supported with a renewed focus on the development of comprehensive and clear guidelines for teaching staff and students, and well-constructed criteria sheets, annotated samples, exemplars, and marking guides for exams. Further, the expert model of moderation (with unit coordinator as arbiter) needs to be replaced with a view of moderation as capacity building through sharing practices and the development of a culture of assessment and moderation. While unit coordinators need to be leaders, to support the continuation of practice and consistency of practice across an entire course, unit coordinators need to be involved in activities that promote the development of shared understandings of assessment standards and the qualities that denote those standards.

These findings highlight the need for ongoing substantive conversations around moderation. With this start, we hope to open up avenues for further critique of the value of moderation processes in higher education when working within standards-based assessment practices. We also hope to have provided academics with a starting point from which to review their current practice.

6. REFERENCES

- Adie, L., Lloyd, M., & Beutel, D. (2011). *QUT Faculty of Education: Moderation practices in the Faculty of Education*. Unpublished report, Teaching and Learning Development Grant, QUT.
- Adie, L., Lloyd, M., & Beutel, D. (2013). Identifying discourses of moderation in higher education. *Assessment and Evaluation in Higher Education*. DOI:10.1080/02602938.2013.769200
- Biggs, J., & Tang, C. (2007). *Teaching for quality learning at university*. Maidenhead, UK: McGraw-Hill / Open University Press.
- 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.
- Cohen, L., Manion, L., & Morrison, K. (2011). *Research methods in education* (7th ed.). New York: Routledge.
- Goos, M. & Hughes, C. (2010). An investigation of the confidence levels of course/subject coordinators in undertaking aspects of their assessment responsibilities. *Assessment & Evaluation in Higher Education*, 35(3), 315-324.
- Sadler, D. R. (2010). *Assuring academic achievement standards at Griffith University*. Consultation Paper (2nd ed.). Retrieved from <http://app.griffith.edu.au/assessment-matters/pdfs/assuring-academic-achievement-standards-second-edition.pdf>
- Sanderson, G., & Yeo, S. (2011). *Moderation for fair assessment in transnational learning and teaching. Final report*. Retrieved from <http://www.olt.gov.au/system/files/resources/PP8-906%20UniSA%20Sanderson%20Final%20Report%202011.pdf>
- TEQSA (Tertiary Education Quality and Standards Authority). (2012, January 29). *Application guide: Application for accreditation of a higher education course of study (AQF Qualification)*. Commonwealth of Australia. Retrieved from http://www.teqsa.gov.au/sites/default/files/Course%20Accred%20Guide_final_080212.pdf
- Van der Schaaf, M., Baartman, L., & Prins, F. (2011). Exploring the role of assessment criteria during teachers' collaborative judgement processes of students' portfolios. *Assessment & Evaluation in Higher Education*, DOI:10.1080/02602938.2011.576312