

What's the point of moderation? A discussion of the purposes achieved through contemporary moderation practices.

Sue Bloxham^{a*}, Clair Hughes^b and Lenore Adie^c

^a*Faculty of Education, Arts & Business, University of Cumbria, Lancaster, United Kingdom*

^b*Faculty of Health and Behavioural Sciences, University of Queensland, Brisbane, Australia*

^c*Faculty of Education, Queensland University of Technology, Brisbane, Australia*

Correspondence details:

Sue Bloxham (*corresponding author)
Emeritus Professor of Academic Practice
University of Cumbria
Green Head Farm
Low Bentham, Lancaster LA2 7ES
UK
Tel: 44 (0) 15242 62037
Email: s.bloxham@cumbria.ac.uk

Clair Hughes
Faculty of Health and Behavioural Sciences
The University of Queensland
St Lucia QLD 4072
Australia
Email: clair.hughes@uq.edu.au

Lenore Adie
School of Cultural and Professional Learning
Faculty of Education
Queensland University of Technology
Victoria Park Road
Kelvin Grove QLD 4059
Australia
Tel: +61 7 3138 8552
Email: lenore.adie@qut.edu.au

Sue Bloxham has taught and researched in higher education for many years, developing a particular interest in assessment, marking and moderation. She has also researched and published on matters such as student skills development, group assessment, feedback, course design and marking. She also has extensive experience in educational development.

Clair Hughes has focused on assessment in both her teaching and research in a number of different educational sectors and contexts. Her current interest is in developing the understanding of assessment tasks that can support assessment design and critique and enhance related assessment processes such as moderation. Her work has been shared widely through publications and conference presentations.

Lenore Adie is a senior lecturer and researcher with the Queensland University of Technology. Her research focuses on assessment and moderation processes as these contribute to supporting teachers' pedagogical practices and student learning. Lenore has worked as a primary school teacher and in administration positions in state and private schools within Queensland, Australia.

What's the point of moderation? A discussion of the purposes achieved through contemporary moderation practices.

An increasingly regulated higher education sector is renewing its attention to those activities referred to as 'moderation' in its efforts to ensure that judgements of student achievement are based on appropriate standards. Moderation practices conducted throughout the assessment process can result in purposes identified as equity, justification, accountability and community building. This paper draws on the limited studies of moderation and wider relevant research on judgement, standards and professional learning to test commonly used moderation practices against these identified purposes. The paper concludes with recommendations for maximising the potential of moderation practices to establish and maintain achievement standards.

Keywords: assessment; moderation; standards

Introduction

Assessment of student learning is a necessary and integral component of any education system. Yet designing quality assessment that generates credible evidence of student achievement and that can be consistently and reliably judged by those assessing the work is a contentious and complex task. In higher education, much has been written about quality assessment, the reliability of judgement-making and the consistency of standards within and across institutions. Internationally, quality assurance and quality improvement within higher education is increasingly in the spotlight (Johnson 2014). While countries such as the UK have operated in 'accountability-heavy conditions' for some time (Kohoutek 2014, 322), others are adopting a stronger emphasis on assessment standards and quality assurance through initiatives which include pan-European Standards and Guidelines for quality assurance, seven of which relate to quality assurance in higher education (Kohoutek 2014). Likewise in Australia, the establishment of national bodies such as the Tertiary Education Quality Standards Agency (TEQSA) and Higher Education Standards Panel (HESP) indicates a shift to tighter regulation.

One approach to quality assurance across many higher education institutions is to require some form of moderation, or verification, of assessment judgments and that is the focus of this paper. While there are a range of understandings of moderation, generally it is considered as a taken-for-granted approach to agreeing, assuring and checking standards (Bloxham, 2009). Systems such as the UK have practised forms of internal and external moderation as part of national quality assurance codes of practice for many years. In the UK, internal moderation uses institutional processes to test the quality of assessment and standards whereas external moderation contributes to that process but also seeks to align quality with national standards. Other systems such as Australia and various European countries have more varied and often less formalised moderation practices. In general, quality assurance of assessment lacks both research and theory (Kohoutek 2014, 310-311) with an inadequate knowledge base to underpin decisions regarding the most effective choice of moderation practices and their potential to assure standards.

The work presented in this paper represents an international collaboration in moderation research that was initiated as a result of several parallel investigations in Australia and the UK. It explores an expanding concept of 'moderation' and questions the stated benefits of much widely-accepted moderation practice. The aim is to present a conceptual review of common approaches to moderation with a view to considering how its effectiveness can be enhanced in relation to its stated purposes. It draws on evidence from Australian and UK research to discuss the question: 'To what extent can existing moderation practices contribute to the assurance of standards in higher education?' The term 'standards', when used in this paper, refers to standards of student achievement; that is, 'output' rather than 'input' or 'process' standards (Bloxham and Boyd 2012). Although it is recognised that broader definitions of 'learning' or 'academic' standards exist in higher education, our focus is on the moderation of student achievement.

In reviewing how effectively different moderation practices contribute to the assurance of standards we:

- identify the range of activities that comprise moderation in higher education;
- select, from within this range, a sample of moderation practices in current use for focused analysis;
- articulate the purposes of moderation;
- draw on limited studies of moderation and wider research on judgement, standards and professional learning to test these familiar methods against the identified purposes.

These stages of our enquiry form the structure of the paper which concludes with a set of recommendations for maximising the potential of moderation practices to establish and maintain achievement standards.

The range of activities

Whereas the term ‘moderation’ in the higher education sector has generally referred to post-judgement processes undertaken to negotiate agreement of grades, there is now evidence that a broader concept of moderation as a cyclic process is being adopted by a number of higher education institutions, for example Curtin (2014), Griffith (n.d.) and Edith Cowan (2013) Universities and the University of Tasmania (2014) in Australia, and Sheffield Hallam (2013) in the UK. This inclusive view of moderation links different moderation practices to stages in the assessment process, each of which requires a specific focus as illustrated in Table 1.

[Insert Table 1 around here]

Elements such as validation, professional accreditation and monitoring might be considered aspects of the wider quality assurance of assessment but they all make a contribution to decisions about achievement standards.

Selection of a sample of practices

The practices considered for analysis are partly drawn from those identified in the Assessing and Assuring Graduate Learning Outcomes (AAGLO) study (Barrie et al. 2014). This study explored the “landscape” of assessment practices across seven disciplines through extended telephone interviews with 48 academics from universities across all Australian states. The disciplines selected were those included in the first phase of a national Learning and Teaching Academic Standards project (Australian Learning and Teaching Council 2009) and the interviewees were purposefully selected as demonstrating active leadership in their discipline at the institutional or national level. In their responses to questions related to the assurance of standards, interviewees identified commonly used assurance practices which are grouped in Table 1 according to the stage of the assessment process at which they occur and the focus of attention. Table 1 also includes activities indicated in the UK Quality Code for Higher Education, chapter B6 (QAA 2013).

Examples of these practices will be found in many universities, but the way they are undertaken in practice will vary according to the aspect of assessment practice which is the focus of activity, the number and position of participants and the mode of communication employed. The interplay of these factors produces variations of each method which in turn influence the dominant purpose of the moderation practice. This analysis attempts to reflect these variations in practice. However as appraisal of every moderation practice listed in Table 1 is outside the scope of this paper a range of illustrative practices (marked with an asterisk) have been selected for analysis.

Purposes of moderation

Higher education institutions frequently imply various purposes for their moderation policies and guidelines. Though wording may vary in scope and detail purposes are generally defined simply in terms of assuring agreement on student achievement and marks awarded - a quality assurance function. In this study, we were interested to use an inclusive purpose which reflected the different ways moderation is conceptualised by those participating in moderation practices. In an investigation of university teachers' moderation practices Adie, Lloyd and Beutel (2013) identified four broad ways of thinking and talking about moderation that influenced the focus and thus the outcome of the moderation practice for individual participants. These were identified as equity, justification, accountability and community building, and are useful in articulating an inclusive purpose for moderation which is firmly related to assuring standards.

Equity indicates where the main goal of moderation is to achieve consistency and fairness for students. Moderation can promote equity when outcomes ensure that tasks enable all students to demonstrate intended outcomes, where all assessors are given opportunities to develop common understandings of standards, and when judgements about student work are consistent with agreed standards and comparable within and across institutions.

Justification is typified by a concern with confidence in decision-making and the ability to defend decisions to students and colleagues, the third purpose, *accountability*, is concerned with meeting systemic requirements in terms of processes (e.g. second-marking) and outcomes (e.g. acceptable grade distributions).

Community building emphasises the value of teacher collaboration in establishing and reviewing standards as revealed in assessment tasks, criteria, standards, and learning and teaching activities. Such activities have the potential to calibrate academics' judgements through building shared interpretations of criteria and standards (Sadler, 2013). *Community*

building also emphasises the capacity to improve assessment design and inform teaching through discussion of issues and sharing of practice.

These purposes were not distinguished as discrete but rather as overlapping, interconnected and agentic. ‘Agentic’ is being used here from a sociocultural perspective and describes the relations and effects of understanding moderation in a particular way which in turn influences actions. As used in this paper, the purposes reflect key potential aims for moderation directly related to assuring standards; that is to ensure consistency and fairness in standards (*equity*), to create confidence for academics in their grading decisions (*justification*), to align with systemic requirements (*accountability*) and to calibrate judgements and build shared interpretations of criteria and standards (*community building*). These purposes, therefore, serve as the analytical framework for considering the sample of moderation methods indicated in Table 1. The analysis draws on empirical evidence and scholarly debate in relevant fields of research to test the case for each moderation practice as an effective approach to assuring standards for students and other stakeholders.

Analysis of moderation practices

Peer scrutiny of assessment design

A number of school systems have for some time successfully undertaken processes for assuring the quality of assessment design (Cumming and Maxwell 2004). However the inclusion of peer scrutiny of assessment design as a formal mechanism for the assurance of standards is a relatively recent development in the higher education sector. It is now becoming increasingly common for assessment design to be “moderated” (for example see the university webpages listed in the Introduction to this paper). Practices associated with peer scrutiny have been found to be quite diverse and include anything from informal chats about tasks among a teaching team, to formal written submissions to institutional or professional committees (Barrie et al 2014). Not all peer scrutiny of assessment activity is

necessarily referred to as “moderation” in institutional documentation as some requirements fall under the category of Quality Assurance. The purposes achieved through peer scrutiny are largely dependent on the concerns of participants and the consequent focus of their attention. For example *equity* is promoted through a focus on ensuring that assessment tasks provide all students with equal opportunities to demonstrate their true standard of learning. Concerns with *equity* also relate to the consistency of mechanisms for responding to student assessment-related questions and for providing feedback on work in progress in order to avoid inadvertent advantage or disadvantage for individuals or class groups.

The AAGLO project (Barrie et al. 2014) found that some form of ‘moderation’ of *accountability* for assessment design was a common feature of institutional quality assurance requirements. However it also found that practices varied enormously among institutions with some interviewees reporting complete autonomy in assessment design decisions while others experienced detailed scrutiny of the alignment of tasks with learning outcomes and a requirement for formal responses to committee requests for adjustments. While the quality of assessment design is not a new concern, emerging approaches to peer scrutiny indicate a shift of purpose from *accountability* through the scrutiny of systemic requirements – number, length, variety and timing of assessment tasks – to *justification* in terms of features with a more significant impact on standards such as alignment with learning outcomes and adherence to an overall program assessment plan. The importance of such scrutiny to the assurance of standards is highlighted through reports that many teachers believe they are assessing a fuller range of learning outcomes than they actually are, that their assessment does not fully align with course learning outcomes, and that a focus on lower cognitive levels allows high grades to be achieved through the memorisation of course materials (Hailikari et al. 2014). In this case, We are using “course” to refer to a subject, module or unit of study

usually of a semester's duration and "programme" to refer to a series of courses required for the award of a degree.

Peer scrutiny of assessment design may not only boost academics' ability to demonstrate *accountability* through compliance with institutional requirements, but also their ability to *justify* this aspect of their practice to students and other stake-holders. The value of this form of moderation has been convincingly demonstrated through standards-based national projects funded by the Australian Office for Learning and Teaching (OLT) in disciplines as disparate as accounting (Hancock et al. 2013) and history (Brawley et al. 2013). Both projects incorporated collaborative activities that challenged participants to explain individual course assessment designs to peers in terms of alignment with nominated learning outcomes and resulted in task revisions that better enabled students to demonstrate targeted outcomes and consequently better enabled assessors to reach agreement on standards during later stages of the projects. Other more generic projects undertaken in the Australian context (Booth, 2014; Krause et al. 2013) have also generated peer feedback on assessment designs and the focus on work generated in final or capstone courses has encouraged scrutiny of assessment design beyond the course.

There are good arguments for the benefits of attention to assessment at the program level (for example, Knight 2000; TESTA 2010) but peer scrutiny has tended in the main to rely on exercises such as curriculum and assessment mapping rather than detailed review of assessment tasks against the program outcomes. As a consequence, moderation practices associated with assurance of the quality of assessment design, particularly as they relate to the whole student experience, are theoretically underdeveloped and have received little attention in the assessment literature (Kohoutek 2014).

Community building, though not commonly the primary focus of peer scrutiny practices, may be an inadvertently positive outcome of activities intended to achieve *equity*, *justification* or *accountability* of assessment design, particularly at the course level.

Second marking

Second marking of student work is a practice that has been designed primarily to ensure that grading decisions are fair for students, that standards of performance are appropriate, and that marking standards are defensible and consistent. For ease of reading, we are using ‘second marking’ to refer to a range of approaches where a second academic ‘marks’ a sample of work or, in the case of ‘double marking’, all the student work. Demonstrating some form of second marking (either blind or with first markers’ grades and comments visible) is a key element of the quality assurance of assessment in the UK (QAA 2013). In Australia, individual universities’ assessment policies may advocate double marking, especially of ‘A’ and fail grades (Krause et al. 2013).

A normal accompaniment to second marking is some form of ‘consensus moderation’ where grades are discussed and final outcomes agreed. For the purposes of this article, we will discuss ‘consensus moderation’ separately in order to highlight important differences.

Second marking perceives marker variation in standards as a potential asset to fairer marking by bringing different points of view to the judgement (Dracup 1997). It generally pays little attention to the appropriateness of the task or whether it has offered equal opportunities to all students. Nevertheless, research has found that academics have a strong faith in second marking to deliver fairness to students (Bloxham and Boyd 2012) and that tutors considered a mark ‘correct’ when awarded by both first and second markers (Handley, den Outer, and Price 2013). However, inexperienced assessors can become more concerned with ensuring that judgements are ‘fairly much what everybody else was awarding’ (Handley,

den Outer, and Price 2013, 894). This consequent focus on consistency poses a threat to *equity* as the potential for fairness becomes subjugated to securing grounds for *justification*.

The extent to which second marking is able to meet the purpose of justification is also contentious. It is not surprising that the appearance of a check on marking standards is valuable in defending judgements and in boosting staff confidence in dealing with students' queries (Swann and Ecclestone 1999). Likewise, demonstration that second marking has taken place is often sufficient to meet institutional requirements for moderation, that is its accountability purpose. However, very few empirical studies on the outcomes of second marking are available to support either justification or accountability purposes although Cannings, Hawthorne, Hood and Houston (2005) found considerable unexplained variation between the numerical marks of double markers. Research into marking in general indicates a considerable lack of reliability in grading across different higher education assessors (Brooks 2012; QAA/HEA 2013; O'Hagan and Wigglesworth 2014) and poses a challenge to the idea that two markers can agree on a more 'accurate' mark than first marking alone. Even if that is presumed, the notion of 'representative sample' is not applicable where students' responses can vary considerably within the same grade boundaries and a sample of assignments cannot be assumed to be 'indicative of the whole' (Partington 1994, 58).

Other challenges to *justification* and *accountability* purposes of moderation are evident. Satchell and Pratt (2010) demonstrate mathematically how second marking is likely to result in a less 'accurate' mark. First markers' grades have been found to be more 'accurate' than second markers, possibly because first markers are more likely to be an expert in the subject matter (Dracup 1997). The expectation of 'blind' second marking may actually change assessors' judgement, for example, encouraging cautious, 'defensive marking' where assessors seek to avoid being out of alignment with their colleagues (Hornby 2003). Murphy (1979) shows the influence on judgement decisions when the first assessor's mark is known

which suggests that the agreed grade may not be the result of independent decision-making on the part of two people.

Second marking could be said to contribute to *community building* through the opportunity it provides to observe other colleagues' marking standards and feedback. In addition, and more importantly, second marking may contribute to *community building* through the potential for dialogue that accompanies the resolution of grades, for example in the type of moderation meeting discussed in the next section.

Moderation discussion (consensus/social moderation)

It is not uncommon for some form of discussion or meeting to take place to agree on grades following second marking, 'pre-marking' of a sample of scripts by all markers or blind (cross institutional) review. We are referring to this as consensus moderation. Consensus moderation has been described as a 'post-judgement quality assurance' (Barrie, Hughes and Crisp, 2014).

While groups of academics discussing an assessment task may agree on a grade or a standard of performance, unless evidence is a part of that discussion, agreement may not serve the purposes of achieving fair outcomes for students. For example, if discussion between markers is limited to 'splitting the difference' which is a commonly used method (Orr 2007), the appearance of agreement over marks may be produced. However, this may be distorting standards toward the mean. Hand and Clewes (2000), Orr (2007) and Reimann et al. (2010) note the impact of power relations in moderation discussions where the views of the senior colleague will hold sway. It is difficult to see how such procedures are delivering *equity* for students.

University administrators may view consensus moderation as meeting the requirements for *accountability* and transparency in marking and it may provide staff with confidence that final marks have been clearly debated and agreed. However, the quality of the

discussion is important. As discussed in relation to *equity*, if the moderation dialogue is reduced to one of finding acceptable common ground (Dracup 1997), then it is unlikely to surface varying conceptions of the valued features of a response or different understandings of vague terms in the criteria. A problem with consensus moderation is that the purpose and outcome of participation may not be transparent at all. Gaining consensus on judgement decisions can mask the level of academic engagement that has occurred in the moderation discussion. Consensus can be reached by agreeing on marks, yet with little reference to the evidence used to *justify* decisions, or with unjustified reference to evidence not listed in the criteria (Orr and Bloxham 2013). While such consensus may satisfy the need for *accountability* and reassure markers in responding to students' queries, it hardly assures the fidelity of the final grade.

Genuine *justification* requires assessors to identify the evidence in student work and match it with the relevant standards. The limited research on the subject suggests that systematic consensus moderation practice involving thorough debates of standards and student work can lead to a greater agreement on grades and confidence in the alignment with established learning outcomes (for example, Watty et al. 2014). However, such moderation, though effective, is also expensive. It involves time to train internal and external moderators and to engage in calibration processes. The emphasis on 'light touch' in quality assurance discourse (QAA 2011) is therefore not compatible with moderation for *justification*.

An underlying assumption of much of the research on standards and marking is that standards are 'fuzzy' (Sadler 1987) and can only be truly found in student work as 'concrete referents' (Sadler 2013, 11). Dialogue can make more visible the tacit knowledge of criteria and standards which is often not evident in explicit criteria, and can enable greater sharing of understandings. Therefore, as stated in the previous paragraph, appropriately thorough consensus moderation can certainly contribute to *community building*.

Consideration of grade distributions

Moderation through the consideration of grade distributions is a response to concerns that students may be advantaged or disadvantaged through ‘hard’ or ‘soft’ marking or an inappropriate spread or bunching of marks. Grade distributions deemed to align with anticipated norms are generally taken as demonstrations of *accountability* and as requiring no further action.

Atypical or unexpected distributions on the other hand are responded to in several ways. Where academics are requested to explain their grade distributions student work can be used as evidence of the *justification* of allocated marks and of the soundness of the standards applied in the marking process.

Where student work itself is not taken into account it becomes apparent that compliance with a perceived norm or accountability is the dominant concern. The response in such circumstances can be either a requirement for assessors to undertake make some form of adjustment to grades or the formulaic rescaling of grades (to fit normal expectations, but not necessarily a normal curve) which is sometimes referred to as ‘statistical moderation’.

Consistency and fairness, the two aspects of *equity*, may be at odds in grade distribution moderation processes if consistency of grade distributions within or across cohorts is achieved at the expense of fairness in the application of standards. In an era of criterion referencing, changes to marks or grades without reference to the criteria and their enactment in student work cannot be said to deliver greater confidence in the appropriateness of marks or in the assurance of standards. The prevalence of this form of moderation can indicate a tension between institutional policy and assessment practice. Barrie et al. (2014) for example found it was not uncommon for moderation discussions to have grade distributions as their point of reference rather than the primary evidence of achievement

against the standards as demonstrated in student work, even though criterion-referenced assessment was espoused policy in the majority of institutions.

Grade distributions can however be useful in highlighting the unexpected and in initiating standards' discussions to determine whether grades and grade distributions are to be accepted, questioned or modified in line with traditional expectations. In such circumstances the consideration of grade distributions can achieve the purpose of *community building*. It is unlikely however that assessors are able to completely detach themselves from existing normative assumptions with the consequence that norm- and criterion-referencing can simultaneously pervade assessment practice with only a fuzzy distinction between them (Yorke 2008). Consistent and comparable grade distributions achieved through the manipulation of numbers alone are no assurance that student work has been judged equitably, that communities have built a shared understanding of standards or that the grades awarded are aligned with the standards.

External peer review (external examining)

External peer review, often known as 'external examining' or sometimes by terms such as 'inter-institutional peer review and moderation' (Krause et al. 2013), can take various forms. These vary from where the external assessor forms part of the 'first' judgement as in the external examining of a PhD thesis to the more common use of external peers to review the assessment of a programme or part programme through scrutiny of assessment tasks, student work, marks, and feedback. In this paper the focus is on the latter broad description of external peer review. Research on external examining and standards is relatively sparse (see Bloxham and Price 2013) and therefore much of the analysis of the potential of external examining is derived from broader studies of academic judgement and moderation.

A recent UK study of external examiners (QAA/HEA 2013) identified that external examiners differ by whether they perceived their task as upholding standards or upholding

standardised and fair procedures. When external examiners are asked to review assessment procedures, academics and universities can have some confidence in claiming that assessment processes have been externally verified, for example, that they provide students with appropriate guidance, adhere to the required moderation practices, and make consistent and legitimate decisions when a student's mark profile indicates a borderline case. External examiners adopting such a role, and using the stated standards of the programmes they were reviewing, tended to define their purpose in terms of fairness and consistency (QAA/HEA 2013).

However, such processes are normally easier to define and check in an explicit manner than assessment methods and grading decisions. The suggestion that we can derive confidence that *equity* has been assured because grading has been moderated by an experienced external assessor is founded on the spurious assumption that such external examiners hold appropriate and shared standards. In practice, the QAA/HEA study (2013) found that even experienced examiners from a range of universities displayed the lack of consensus evident in higher education marking research in general. In truly blind reviews of final year work, they tended to use a wide and diverse range of criteria and, even where similar criteria were shared, varying interpretations resulted in manifestly different standards and little inter-examiner agreement. The QAA/HEA findings are consistent with earlier work arguing that examiners are bounded by their social and cultural environment and expectations (Ross 2009), and that personal experience of both standards and quality assurance processes are more significant than formal reference points in providing the basis for examiners' decisions (Colley and Silver 2005; QAA 2005).

External examiners offer individual academics, programmes and universities a 'public defence' of their assessment processes and standards which provides for both *justification* and *accountability*. Bloxham and Boyd (2012) found that lecturers perceive external

examiners as the authority on standards and therefore their confirmation of marks is likely to give confidence in dealing with students. For a wider audience, external examiner's reports are used by institutions to support claims of high quality and a commitment to regulations and standards. As discussed in the preceding paragraphs, this *justification* in relation to standards may be based on supposition rather than evidence.

The UK approach and the Australian methods (Krause et al. 2013; Group of Eight 2014; Watty et al. 2014) all stress the contribution that external peer review can bring to promoting discussion on 'best practice in teaching and learning' (University of Sydney, 2014). Indeed, it can be argued that external examining and external peer review have the potential to promote *community building* in encouraging observation and discussion of standards across different institutions. External examiners/reviewers have the opportunity to see standards outside their own institution and internal staff can gain feedback on their standards from an external viewpoint. Such practices can contribute to the development of a shared understanding of standards in so far as external examiners recognise the socially situated nature of standards and the need for reference to explicit statements such as disciplinary learning outcomes, benchmarks and qualification frameworks. Therefore, while we are not arguing that external examiners assure *equity* or *justification*, there is a sense in which they can make a useful contribution to *community building* regarding standards across institutions.

The implications of this analysis are that if a process of external review is used as a form of moderation, there are considerable consequences in relation to the training of external reviewers, calibration processes and the time and resources to participate in significant scrutiny of material and dialogue with staff teams.

So what's the point of moderation?

Given the growing body of research that highlights the challenge of achieving reliable, appropriate, academic judgement in higher education (Brooks 2012; QAA/HEA 2013; O'Hagan and Wigglesworth 2014), there is a continued need to critically examine our claims for assessment standards. The purpose of this paper was to explore some common approaches to moderation in an attempt to answer the question: 'To what extent can existing moderation practices contribute to the assurance of standards in higher education?' The discussion in this article has, necessarily, had to make some assumptions about widespread moderation practices and we recognise that rather than there being one set process for each of these common methods, it is often the form and emphasis that individual variations take that determine the dominant purpose.

Having said that, there are some clear themes emerging from this investigation. Firstly, the findings suggest that many commonly used activities only meet the purposes of *justification* and *accountability* in terms of an apparent assurance of standards and offer little in terms of concrete evidence of calibrating standards or providing *equity* to students. Much of our exploration of this topic has led us to conclude that moderation practices provide largely a public image of systematic checking of standards. The moderation practices that have been analysed provide some assurance that institutional assessment procedures are in place (for example, anonymous marking, scrutiny of assignment design or double marking) and this can give confidence that due process has been observed. However we have to recognise that quality assurance through moderation of 'inputs' and 'processes' does not necessarily guarantee appropriate 'output' standards (Alderman 2009) or a satisfactory student assessment experience (Crook, Gross, and Dymott 2006).

Secondly, while purposes of *equity* are used to support the validity of these moderation practices, some common moderation practices may be distorting standards and encouraging bunched rather than appropriately differentiated grades. Fairness and

consistency may be at odds when consistency is achieved at the expense of fairness, for example when manipulating grade distributions or averaging first and second marker's judgements. In addition, efforts to standardise grades or come to an agreement over marks may encourage practices which favour quick or technical decision-making over careful discussion of evidence. Overall, we may hold indefensible assumptions regarding the power of moderation to deliver *equity, justification* and *accountability*.

Thirdly, moderation with a focus on *community building* has been shown to add to assessors' assessment literacy as well as knowledge of standards. This view is reinforced by empirical studies of face-to-face consensus building (Watty et al. 2014; Brawley et al. 2013; Hancock et al. 2013) and through teacher claims that such activities assist in the calibration of their standards (Bloxham and Boyd 2012). In addition, for people to benefit from moderation as a form of *community building*, those involved need to share an understanding of this purpose for the activity. This condition holds equally for external reviewers who may be unaware of their role's potential contribution to such an outcome. Universities need to interrogate their policies for moderation to determine their effectiveness in encouraging or requiring academics to develop a shared understanding of the standards, as well as agreement on how the evidence in the student work *justifies* the grade awarded.

Finally, the research reveals the complexity of moderation although it is likely that understanding of this may be limited in many university systems depending on tradition and local and national policy. The complexity of moderation has also meant that while this paper has explored the likely purposes of many common approaches, its scope has not permitted consideration of other relevant factors in the moderation debate such as feasibility, levels of engagement or impact on the student experience.

Recommendations

What this review of common moderation practices suggests is that few current approaches contribute significantly to the assurance of higher education standards. A key finding is the importance of a thorough process involving significant academic engagement and debate focussed on evidence aligned with standards descriptors. What alternatives are there?

Moderation in the assessment process

When there is little time available, as in university assessment timelines, the task of achieving appropriate, consistent and fair standards is enormously difficult. Consequently, our first recommendation is that moderation is built into the whole teaching and assessment process with a particular emphasis on pre-teaching moderation; that is, moderation which is informed by, and informs, the learning design. This has a potential range of benefits for staff and students. Early scrutiny and discussion of assessment tasks is possibly the least well developed method yet it can inform teaching and, very importantly, inform dialogue with students about assessment, supporting them to understand the required standards (Willis and Adie 2014). Handley, den Outer, and Price (2013) found that post teaching discussion of marking standards may be too late and students may already have been given a ‘wrong’ idea of how to go about the assignment by members of a teaching team.

In addition, early consideration of expected standards through discussing assessment tasks and, perhaps, reviewing examples of work from the previous cohort, may provide for better consistency in judgement across a teaching team. This cannot easily be obtained at the end of the course because marking deadlines are too tight for effective collaborative scrutiny of student work (Lawson et al. 2014; Albertyn, Kapp, and Frick 2007). In this scenario, end of course moderation could focus on close scrutiny of the hard borderline cases with detailed examination and discussion of the students’ work matched against the stated standards.

In this model, there is the potential for the limited resources available to be focused on moderation practices which can have some impact beyond demonstrating compliance or

completing unproven routines. In effect, we are suggesting flipping the process so that most moderation effort is applied at the beginning of courses, where it is more useful to students, likely to bring greater consistency in advice, and encourage an ‘assessment for learning conversation’ amongst staff. The approach certainly deserves further investigation.

Professional development and calibration of standards

The well conducted and thoughtful moderation practices we propose constitute professional learning and promote the enhancement of assessment design as well as system-wide consistency of judgement and maintenance of standards. In particular, dialogical, cross-university consensus moderation has the potential to support professional development through the calibration of disciplinary standards. We would argue that academics are more likely to see the value in effective moderation if they have greater awareness of the disparity in individuals’ standards, their situated nature and their diverse provenance (for example, in prior experience, values, and formal documentation). Such professional development is likely to be assisted by greater interactive moderation, for example, in assessing exemplars, or discussing marks and criteria.

In conclusion this position paper has drawn on existing literature and policy to review a commonly practised but rarely researched aspect of higher education assessment. Undoubtedly, it is limited in scope and evidence but we hope that it has commenced an important conversation regarding the ‘point’ of moderation which is most clearly understood when practices move beyond accountability to inform teaching and enhance student learning opportunities. There is a need for more empirical research on the effectiveness of different methods as well as greater informed debate in the sector about priorities and possibilities in an age of increasing global interest in achievement standards.

References

- Adie, L., M. Lloyd, and D. Beutel. 2013. "Identifying Discourses of Moderation in Higher Education." *Assessment and Evaluation in Higher Education* 38 (8): 968-977.
- Albertyn, R. M., C. A. Kapp, and B. L. Frick. 2007. "Taking the Sting out of Evaluation: Rating Scales for Thesis Evaluation." *South African Journal of Higher Education* 21 (8): 119-123.
- Alderman, G. 2009. "Defining and Measuring Academic Standards: A British Perspective." *Higher Education Management and Policy* 21 (3): 11-22.
- Australian Learning and Teaching Council. 2009. "Learning and Teaching Academic Standards (LTAS) project overview". <http://www.altc.edu.au/standards/overview>
- Barrie, S., C. Hughes, and G. Crisp. 2014. "Assessing and Assuring Graduate Learning Outcomes (AAGLO) Final Report". Office for learning and Teaching, Sydney. Australia. <http://www.itl.usyd.edu.au/projects/aaglo/overview.htm>
- Bloxham, S. 2009. "Marking and Moderation in the UK: False Assumptions and Wasted Resources." *Assessment and Evaluation in Higher Education* 34 (2): 209-220.
- Bloxham, S., and P. Boyd. 2012. "Accountability in Grading Student Work: Securing Academic Standards in a 21st Century Quality Assurance Context." *British Educational Research Journal* 38 (4): 615-634.
- Bloxham, S., and M. Price. 2013. "External Examining: Fit for Purpose?" *Studies in Higher Education*. doi: 10.1080/03075079.2013.823931.
- Booth, S. 2014. *National Peer Review Network: Peer Review of Assessment* (Powerpoint presentation). Office for Learning and Teaching. http://www.utas.edu.au/data/assets/pdf_file/0009/567540/PRNpowerpointPresentation25072014.pdf

- Brawley, S., J. Clark, C. Dixon, L. Ford, E. Nielsen, S. Ross, and S. Upton. 2013. *After Standards: Engaging and Embedding History Standards using International Best Practice to Inform Curriculum Renewal*. Sydney: Office for Learning and Teaching.
- Brooks, V. 2012. "Marking as Judgement." *Research Papers in Education* 27 (1): 63-80.
- Cannings, K. Hawthorne, K. Hood and H. Houston. 2005. Putting double marking to the test: a framework to assess if it is worth the trouble *Medical Education* 39: 299–308.
- Colley, H., and H. Silver. 2005. *External Examiners and the Benchmarking of Standards*. York: Higher Education Academy.
- Crook, C., H. Gross, and R. Dymott. 2006. "Assessment Relationships in Higher Education: The Tension of Process and Practice." *British Educational Research Journal* 32 (1): 95-114.
- Cumming, J., and G. Maxwell. 2004. "Assessment in Australian schools: Current Practice and Trends." *Assessment in Education: Principles, Policy and Practice* 11 (10): 89-108.
- Curtin University. 2014. "Moderation." Curtin University.
http://otl.curtin.edu.au/teaching_learning_practice/assessment/moderation.cfm
- Dracup, C. 1997. "The Reliability of Marking on a Psychology Degree." *British Journal of Psychology* 88 (4): 691-708.
- Edith Cowan University. 2013. "Moderation of Assessment." Edith Cowan University.
<https://intranet.ecu.edu.au/learning/support-and-resources/assessment-and-benchmarking/moderation-of-assessment>
- Griffith University. n.d. "Consensus Moderation." Griffith University.
<http://app.griffith.edu.au/assessment-matters/docs/consensus-moderation>
- Group of Eight. 2014 Quality Verification System: Assessment Review Guidelines. Turner ACT: Group of eight. Available at

<http://www.adelaide.edu.au/learning/reviews/benchmarking/external/QVS-2014-assessment-review-guidelines.pdf>

Hailikari, T., L. Postareff, T. Tuononen, M. Raisanen, and S. Lindblom-Ylänne. 2014.

“Student’s and Teachers’ Perceptions of Fairness in Assessment.” In *Advances and Innovations in University Assessment and Feedback*, edited by Caroline Kreber, Charles Anderson, Noel Entwistle, and Jan McArthur, 99-113. Edinburgh: Edinburgh University Press.

Hancock, P., M. Freeman, A. Abraham, P. De Lange, B. Howieson, B. O’Connell, and K.

Watty. 2013. *Achievement Matters: External Peer Review of Accounting Learning Standards*. Sydney: Office for Learning and Teaching.

Hand, L., and D. Clewes. 2000. “Marking the Difference: An Investigation of the Criteria

Used for Assessing Undergraduate Dissertations in a Business School.” *Assessment and Evaluation in Higher Education* 25 (1): 5-21.

Handley, K., B. den Outer, and M. Price. 2013. “Learning to Mark: Exemplars, Dialogue and

Participation in Assessment Communities.” *Higher Education Research and Development* 32 (6): 888-900.

Hornby, W. 2003. “Assessing Using Grade-Related Criteria: A Single Currency for

Universities?” *Assessment and Evaluation in Higher Education* 28 (4): 435-454.

Johnson, G. M. 2014. “Record of Assessment Moderation Practice (RAMP): Survey

Software as a Mechanism of Continuous Quality Improvement.” *Assessment and Evaluation in Higher Education*. doi: 10.1080/02602938.2014.911244.

Knight, P. T. 2000. “The Value of a Programme-wide Approach to Assessment.” *Assessment*

and Evaluation in Higher Education 25 (3): 237-251.

- Kohoutek, J. 2014. "European Standards for Quality Assurance and Institutional Practices of Student Assessment in the UK, the Netherlands and the Czech Republic." *Assessment and Evaluation in Higher Education* 39 (3): 310-325.
- Krause, K-L., G. Scott, K. Aubin, H. Alexander, T. Angelo, S. Campbell, M. Carroll, et al. 2013. *Assuring Final Year Subject and Program Achievement Standards Through Inter-university Peer Review and Moderation*. Sydney: Office for Learning and Teaching.
- Lawson, R., T. Taylor, E. French, E. Fallshaw, C. Hall, S. Kinash, S., and J. Summers. 2014. "Hunting and Gathering: New Imperatives in Mapping and Collecting Student Learning Data to Assure Quality Outcomes." *Higher Education Research and Development*. doi: 10.1080/07294360.2014.911249.
- Murphy, R. J. L. 1979. "Removing the Marks from Examination Scripts before Re-marking Them: Does it make a Difference?" *British Journal of Educational Psychology* 49 (1): 73-78.
- O'Hagan, S. R., and G. Wigglesworth. 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.
- Orr, S. 2007. "Assessment Moderation: Constructing the Marks and Constructing the Students." *Assessment and Evaluation in Higher Education* 32 (6): 645-656.
- Orr, S., and S. Bloxham. 2013. "Making Judgements about Students Making Work: Lecturers' Assessment Practices in Art and Design." *Arts and Humanities in Higher Education* 12 (2-3): 234-253.
- Partington, J. 1994. "Double-marking Students' Work." *Assessment and Evaluation in Higher Education* 19 (1): 57-60.

- QAA (Quality Assurance Agency). 2005. *Outcomes from Institutional Audits: External Examiners and their Reports*. Gloucester: Quality Assurance Agency.
- QAA (Quality Assurance Agency). 2011. "Risk and Regulation." *Talking about Quality*, Issue 1: 1-7. <http://www.qaa.ac.uk/en/Publications/Documents/Risk-and-regulation.pdf>
- QAA (Quality Assurance Agency). 2013. *The UK Quality Code for Higher Education, Chapter B6: Assessment of students and recognition of prior learning*. Gloucester: Quality Assurance Agency.
- QAA/HEA (Quality Assurance Agency/ Higher Education Academy). 2013. *External Examiners' Understanding and Use of Academic Standards*. Gloucester: QAA. Accessed February 5, 2014. <http://www.qaa.ac.uk/publications/information-and-guidance/publication/?PubID=2686#.VDIN-p1wb4g>
- Reimann, N., K. Harman, A. Wilson, and L. McDowell. 2010. "Learning to Assess in Higher Education." Paper presented at the *Higher Education Close Up Conference*, Lancaster July 20-22.
- Ross, V. 2009. "External Music Examiners: Micro-macro Tasks in Quality Assurance Practices." *Music Education Research* 11 (4): 473-484.
- Sadler, D. R. 1987. "Specifying and Promulgating Achievement Standards." *Oxford Review of Education* 13 (2): 191-209.
- Sadler, D. R. 2013. "Assuring Academic Achievement Standards: From Moderation to Calibration." *Assessment in Education: Principles, Policy and Practice* 20 (1): 5-19.
- Satchell, S., and J. Pratt. 2010. "The Dubiety of Double Marking." *Higher Education Review* 42 (2): 59-62.
- Sheffield Hallam University. 2013. *Moderation Policy and Operating Guidelines*. Sheffield Hallam University.

https://students.shu.ac.uk/regulations/exams_and_coursework/Moderation%20Policy%20and%20Operating%20Guidelines_2012-13.pdf

- Swann, J., and K. Ecclestone. 1999. "Litigation and Learning: Tensions in Improving University Lecturers' Assessment Practice." *Assessment in Education* 6 (3):357-375.
- TESTA (Transforming the Experience of Students Through Assessment). 2010. "Welcome to TESTA." The University of Winchester. <http://www.testa.ac.uk/>
- University of Sydney. 2014. Go8 Quality Verification System. Sydney: University of Sydney. <http://sydney.edu.au/ab/qvs> [accessed November 5th 2013]
- University of Tasmania. 2014. "*Faculty of Education: Moderation Guidelines*". University of Tasmania. <http://www.utas.edu.au/education/learning-and-teaching/policies-and-procedures/moderation>
- Watty, K., M. Freeman, B. Howieson, P. Hancock, B. O'Connell, P. de Lange, and A. Abraham. 2014. "Social Moderation, Assessment and Assuring Standards for Accounting Graduates." *Assessment and Evaluation in Higher Education* 39 (4): 461-478.
- Willis, J., and L. Adie. 2014. "Teachers using annotations to engage students in assessment conversations: Recontextualising knowledge." *The Curriculum Journal*. DOI: 10.1080/09585176.2014.968599
- Yorke, M. 2008. *Grading Student Achievement in Higher Education: Signals and Shortcomings*. Abingdon: Routledge.

Table 1. Moderation activities at different stages of the assessment process.

| Stage of the assessment process | Moderation focus | Illustrative moderation activities |
|------------------------------------|---|--|
| 1. Design | Quality of tasks and overall plan (course and program) | * Peer scrutiny (e.g., Informal consideration by course teams or formal Coordinator/Committee approval) Professional accreditation process |
| 2. Calibration | Shared understanding of task specifications, requirements, performance criteria and standards | Informal processes (e.g., socialisation) Formal processes (e.g., mentoring , workshops involving comparison and justification of judgements of student work samples) |
| 3. Judgements | Quality of judgements as evidenced by: <ul style="list-style-type: none"> ● adherence to criteria ● credibility of evidence ● shared standards ● consistency of judgement | * Second, double or collaborative marking Random checking * Consensus moderation discussions * Consideration of grade distributions |
| 4. External validation/ comparison | Comparability/benchmarking of standards | * External examining Peer review/verification Professional accreditation processes |
| 5. Monitoring evaluation | Overall quality of assessment and of components of individual stages | Consideration of: <ul style="list-style-type: none"> ● student work samples ● student satisfaction data ● examiners' reports ● grade distributions ● lecturer and tutor perceptions (adequacy task design and information, criteria and standards, marking guides etc.) |

Activities marked () are those selected for further analysis*