

**Assessment moderation in higher education: Guiding practice with evidence-an integrative review**

Author

Prichard, Roslyn, Peet, Jacqueline, El Haddad, May, Chen, Yingyan, Lin, Frances

Published

2024

Journal Title

Nurse Education Today

Version

Accepted Manuscript (AM)

DOI

[10.1016/j.nedt.2024.106512](https://doi.org/10.1016/j.nedt.2024.106512)

Rights statement

© 2024 The Authors. Published by Elsevier Ltd. This is an open access article distributed under the terms of the Creative Commons CC-BY license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Downloaded from

<https://hdl.handle.net/10072/434239>

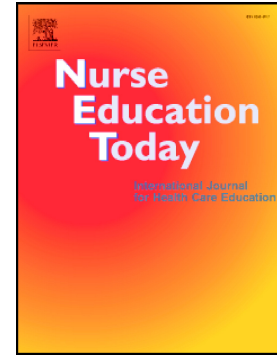
Griffith Research Online

<https://research-repository.griffith.edu.au>

## Journal Pre-proof

Assessment moderation in higher education: Guiding practice with evidence-an integrative review

Roslyn Prichard, Jacqueline Peet, May El Haddad, Yingyan Chen, Frances Lin



PII: S0260-6917(24)00422-2

DOI: <https://doi.org/10.1016/j.nedt.2024.106512>

Reference: YNEDT 106512

To appear in: *Nurse Education Today*

Received date: 20 April 2024

Revised date: 25 October 2024

Accepted date: 23 November 2024

Please cite this article as: R. Prichard, J. Peet, M. El Haddad, et al., Assessment moderation in higher education: Guiding practice with evidence-an integrative review, *Nurse Education Today* (2024), <https://doi.org/10.1016/j.nedt.2024.106512>

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2024 Published by Elsevier Ltd.

# **ASSESSMENT MODERATION IN HIGHER EDUCATION: GUIDING PRACTICE WITH EVIDENCE -AN INTEGRATIVE REVIEW**

## **Author List**

Roslyn Prichard <sup>1</sup>, rprichar@usc.edu.au, <https://orcid.org/0000-0001-8057-6605>

Jacqueline Peet <sup>1</sup>, jpeet2@usc.edu.au, <https://orcid.org/0000-0001-6541-6099>

May El Haddad <sup>1</sup>, melhadda@usc.edu.au, <https://orcid.org/0000-0003-2328-685X>

Yingyan Chen <sup>1</sup>, ychen@usc.edu.au, <https://orcid.org/0000-0001-8452-8176>

Frances Lin<sup>1,2,3,4</sup>, frances.lin@flinders.edu.au, <https://orcid.org/0000-0001-8735-5469>

## **Affiliations:**

1. School of Health, University of the Sunshine Coast, Australia
2. Sunshine Coast Health Institute, Queensland, Australia
3. College of Nursing and Health Sciences, Flinders University, Australia
4. Caring Futures Institute, Flinders University. Australia

**Corresponding author:** Roslyn Prichard, rprichar@usc.edu.au

**Acknowledgements:** Sue Svensen (academic librarian) assisted with the choice of database, as well as the design and management of our keyword search strings and results. Gwen Veltmeyer (sessional academic) assisted with the guideline review.

**Contributions:** Roslyn Prichard: Project administration, Methodology, Data curation, Writing-Original draft, Reviewing and Editing, Jacqueline Peet. Conceptualisation, Writing-Reviewing and Editing, May El Haddad: Conceptualisation, Writing- reviewing and editing. Yingyan Chen: Writing- Reviewing and Editing.: Frances Lin: Supervision, Conceptualisation, Writing-Reviewing and Editing

**Declarations of interest:** none

**Assessment moderation in higher education: Guiding practice with  
evidence-An integrative review**

**Abstract:**

**Background:**

Maintaining the quality and integrity of student assessments, especially in professional fields like nursing, is critical. Managing moderation processes across large teams poses social and logistical challenges, further complicated by varying quality and clarity of institutional guidelines. Systematic reviews on moderation practices in higher education are scarce.

**Objectives:**

This integrative review critically analyses moderation practices in the literature and evaluates the quality of published institutional guidelines to support faculty in enhancing moderation practice.

**Method:**

An integrative review, guided by research questions, was used to facilitate data extraction. Searches in ERIC, Web of Science, and A Plus Education Informit databases focused on higher education papers published in English between 2008 and 2023. Of 552 studies screened, 19 were included in the final review. The quality of the included studies was appraised using the appropriate Critical Appraisal Skills Programme (CASP) checklists, tailored to the specific research designs.

Institutional guidelines were identified via web searches and evaluated using a tool based on elements of a recognised clinical practice guideline assessment tool.

**Results:**

The review included 19 peer-reviewed studies and 27 institutional guidelines. Of the studies, 14 were qualitative, 2 quantitative, 2 mixed-methods, and 1 a guideline review, with most focused on Australian higher education. Key findings indicate the importance

of collaborative moderation processes, in establishing shared standards. The studies highlighted the need for clear marking guides, ongoing training, pre-teaching discussions and adequate resourcing. The guideline evaluation revealed variability in quality, with frequent deficiencies in stakeholder involvement and development rigour. Some guidelines offered practical implementation tools.

**Conclusions:**

Moderation practices in higher education are complex and resource-intensive, particularly for large teams. Effective moderation is crucial for maintaining assessment quality and integrity. Consensus moderation, through collaboration among markers, tutors, and assessors, supports shared standards via dialogue, calibration, and consensus building. Improving moderation practices requires evidence-based guidelines, a focus on consensus-building, sufficient resources, and ongoing professional development.

**Keywords:** Assessment, moderation, higher education, systematic review

## INTRODUCTION:

Assessment moderation processes are quality assurance measures designed to ensure consistency, appropriateness, and fairness in assessment judgments, as well as the validity and reliability of assessment tasks, criteria, and standards (Tertiary Education Quality and Standards Agency [TEQSA], 2021). Engaging in moderation processes presents an opportunity for teaching team members to collaborate, building a shared understanding of assessment criteria, standards, and the evidence that distinguishes different grades achieved (Adie et al., 2013; Sadler, 2013). In contexts where complex assessments involve large marking teams, such as those common in nursing courses, consensus models among multiple markers are often used. In these models, markers are tasked with making subjective grading decisions, often in isolation, which emphasises the importance of establishing a collective understanding of quality standards (Sadler, 2013). However, despite the overarching principles and policies mandating rigorous moderation processes, there is significant variation in how these processes are implemented at both the course and institutional levels (Williams et al., 2023).

The concept of measuring success in moderation practice is also contentious, but higher education institutions are required to implement rigorous and auditable moderation practices to demonstrate consistent and reliable assessment standards (Beutel et al., 2017). Rigorous moderation practices can help identify gaps or weaknesses in assessment processes and quality. Involving staff in moderation processes during the planning phase, including assessment design and alignment with course learning outcomes, and employing traditional quality assurance procedures during and after marking, can help support high standards and assessment integrity (Smith, 2012).

Developing sustainable moderation practices requires active leadership (Beutel et al., 2017; Mason et al., 2022). In higher education, coordinators and faculty leaders are often immersed in day-to-day operations, with little time to step back, and gain a perspective on the reality of moderation practice in their schools and institutions (Mason et al., 2022). Nevertheless, investing time and resources in developing practical and applicable moderation guidelines is essential to provide faculty with the tools and frameworks necessary to effectively navigate and optimise their consensus moderation practices (Beutel et al. 2017). By implementing systematic and auditable moderation practices, nursing faculty can demonstrate commitment to quality assurance and enhance their reputation in higher education. However, tension often arises between the goal of supporting teaching and learning through moderation practices, ensuring accountability, consistency, and fairness, and the difficulty of allocating sufficient resources to achieve these goals (Beutel et al., 2017; Mason et al., 2022). The challenge for nursing faculty and programmes of study is to provide practical and effective models of moderation that meet these expectations. Research evidence on moderation practices in higher education is sparse, and, to our knowledge, no systematic reviews on this topic exist in the literature.

#### **OBJECTIVES:**

The aim of this integrative review was to synthesise literature on moderation practices in higher education and to assess the quality of published institutional moderation guidelines, to provide faculty with insights and guidance for navigating moderation practices and developing effective moderation guidelines. To achieve this, the following research questions were addressed:

1. What higher education moderation practices are described in the peer-reviewed literature?
2. How are faculty and casual staff engaged in moderation processes?
3. How is moderation practice evaluated in higher education?
4. How do prevailing discourses and viewpoints shape the practice of assessment moderation in higher education?
5. What is the quality of publicly available higher education moderation guidelines?

#### **METHODS:**

An integrative literature review was conducted following Whitemore and Knaff's (2005) approach, a method designed to synthesise diverse forms of evidence systematically. Reporting was guided by PRISMA checklist (Moher et al., 2009). This approach was employed to address the first four research questions.

Additionally, a quality assessment of publicly available institutional guidelines was undertaken to address the fifth research question. Currently, there is no published tool specifically designed to assess the quality of institutional guidelines in higher education. However, the AGREE II clinical practice guideline assessment tool (Brouwers et al., 2010), is widely recognised for evaluating the quality and transparency of clinical practice guidelines in healthcare. Drawing from five of the six AGREE II domains, we developed the Education Guideline Evaluation Tool (EduGET), with criteria tailored to the educational context for evaluating the included guidelines. These criteria focus on scope, applicability, clarity, stakeholder involvement, and development rigour, providing a systematic approach to our guideline quality assessment (Table 1). While



formal content validity testing of this tool has not yet been conducted it offers a structured approach for evaluating educational guidelines.

Separate search strategies were employed for published peer-reviewed literature and institutional moderation guidelines, with details outlined in the following sections.

### **Search strategy and quality assessment for published literature**

The literature search was conducted across three databases: ERIC, Web of Science, and A Plus Education Informit, focusing on peer-reviewed publications related to moderation practices in higher education. The search strategy was developed in collaboration with an academic librarian, and targeted papers published in English between 2007 and 2023. This period was selected to capture the significant regulatory changes impacting the management of moderation practices in higher education in Australia and internationally (Beutel et al., 2017; TEQSA, 2017) and an uptick in publication in the area. Search strings were designed to capture relevant studies by combining terms such as "higher education," "assessment," "moderation," and "quality assurance". For example, in ERIC, the search string used was ("higher education" OR "tertiary education" OR university) AND (assessment OR assignment OR grading) AND (moderation OR "quality control" OR "double marking" OR benchmarking) AND (evaluation OR metric). Detailed search strings used are available in Supplementary file 1.

Primary research, case studies, and reviews on higher education moderation processes were included, while studies conducted outside of higher education setting, as well as discussion papers, and non-English publications, were excluded. Two reviewers screened titles and abstracts, resolving any discrepancies with a third author. Final inclusion or exclusion decisions were made after full-text screening by paired authors.

Figure 1 illustrates the search strategy and screening outcomes. Data extraction was similarly performed by paired authors, who independently added details from each study and compared results, with a third author addressing any discrepancies.

Pairs of reviewers independently conducted quality assessments using the Critical Appraisal Skills Programme (CASP, 2018) checklists appropriate to the study design of included papers. The CASP checklists for qualitative research and relevant elements from the randomised controlled trials (RCTs) checklist were used depending on the study design. For mixed-methods studies, we assessed each component using the corresponding CASP checklist items. Papers were categorically graded based on the proportion of “yes” responses on the checklists: those scoring 80-100% were rated as higher quality, 50-80% as moderate quality, and less than 50% as poorer quality. All selected papers were included in the review irrespective of the quality.

### **Data analysis of included papers**

A narrative synthesis approach, focusing on the results, discussion, and conclusion sections of each included study was used. Narrative synthesis, as outlined by Popay et al. (2006), is a well-established evidence-based method for integrating findings from diverse study designs, such as those found in our review, making it especially useful for studies with varying methodologies. Extraction was guided by predetermined research questions, to ensure a structured and consistent analysis of the studies. As described by Whitemore and Knafl (2005), this approach enhances rigour by aligning the data extraction process with the specific areas of interest, ensuring relevant material related to our research questions was systematically captured. This approach allowed for a flexible yet structured synthesis of moderation practices in higher education. Included papers are presented in Table 1.

### **Search strategy and quality assessment for institutional guidelines**

Institutional guideline searches were conducted using Google and Google Scholar to identify published institutional guidelines and policies. Key words used were ‘Assessment, Moderation (guidelines or policy), and (higher or tertiary) education’. The first three pages from each search were reviewed, producing 40 relevant documents for screening. Guidelines not in a higher education context were excluded.

Currently, there is no published tool specifically designed to assess the quality of institutional guidelines in higher education. However, the AGREE II clinical practice guideline assessment tool (Brouwers et al., 2010) is widely recognised for evaluating the quality and transparency of clinical practice guidelines in healthcare. Drawing from five of the six AGREE II domains, we developed the Education Guideline Evaluation Tool (EduGET), with criteria tailored to the educational context for evaluating the included guidelines. These criteria focus on scope, applicability, clarity, stakeholder involvement, and development rigour, providing a systematic approach to our guideline quality assessment (Table 2). While formal content validity testing of this tool has not yet been conducted, it offers a structured approach for evaluating educational guidelines.

The authors worked in pairs to assess each guideline across the five domains of the tool. Guideline quality was analysed categorically, with the percentage of "yes" scores (as opposed to "no" or "can't tell") calculated for each domain based on the review of the included guidelines.

### **RESULTS:**

Our review included 19 peer-reviewed papers and 27 institutional guidelines. The results, presented below, draw on data from the included papers to address the first four

research questions, and on the quality assessment of the institutional guidelines to respond to the fifth question.

A total of 19 research papers were included in the literature review (Table 1). Of these, 14 were qualitative, two were quantitative, two used mixed-methods, and one was a moderation guideline review. Twelve studies were conducted in Australia, three in the United Kingdom (UK), and one in Taiwan. The disciplines represented included economics, accounting, teacher training, art and design, STEM, IT management, communication, humanities, psychology, zoology, and medicine. The quality assessment results and methodologies used in each paper are presented in Table 1. Our analysis of the included papers is outlined below.

### **Moderation Practices in Higher Education**

Moderation processes were described in varying detail in fourteen selected papers (Table 1). Collaborative processes among markers, tutors, or assessors to establish shared standards, ensure consistency, and reach consensus on grading are a central feature of consensus moderation models. This collaboration often involves dialogue, discussions, and meetings among stakeholders, and includes “initial marking” of sample papers (Bausili, 2018), “early moderation” (Broadbent, 2018) or “sample marking” (Buetel et al., 2017; Watty et al., 2014). Grainger et al. (2008) affirm that achieving a shared understanding of the descriptors and rubrics used is a critical element of successful moderation. Conceptually, some authors describe this as a socially moderated process of calibration (Crimmins et al., 2016; Orr, 2007; Watty et al., 2014). Managerial approaches also exist, with Chen et al. (2017) describing rigorous, iterative, and resource-intensive marker training aimed at reducing grading variation in an information technology management department.

Most papers broadly describe a consensus moderation process, with some focussing on the marking phase (Bausili, 2018; Bird & Yucel, 2013; Boyd & Bloxam, 2014; Broadbent, 2018; Chen et al., 2017; Czaplinski et al., 2014; Evans & Gardener, 2018; Grainger et al., 2008; O'Connell et al., 2016), while others adopt a more holistic approach that encompasses the whole assessment development and delivery cycle (Beutel et al., 2017; Crimmins et al., 2016; Hunter & Docherty, 2011; Mason & Roberts, 2023; Watty et al., 2014).

The complexities of moderation for large marking teams are explored in four of the papers (Broadbent, 2018; Crimmins et al., 2016; Czaplinski et al., 2014; Hunter & Docherty, 2011). All four papers emphasise the importance of providing clear marking guides, rubrics, and expectations to markers. Broadbent (2018) suggests using resources like videos to explain criteria and demonstrate performance levels. Crimmins et al. (2016) highlight the need for ongoing training and support to develop marker expertise and ensure consistency. Calibration and consensus building among markers and coordinators can take the form of early and periodic discussions, debates, and negotiations to support shared understanding and consistent judgements (Crimmins et al., 2016; Czaplinski et al., 2014). Hunter and Docherty (2011) highlight the importance of moderation meetings and active engagement among markers in promoting shared understandings of assessment standards. They advocate for pre-teaching discussions as a way to upskill tutors, who can then incorporate their enhanced understanding of the assessment criteria into their teaching practices (Hunter & Docherty, 2011).

### **Faculty engagement in moderation processes**

Investing in bringing teaching team members and markers together to improve assessment practices and foster a collective understanding of tasks, criteria, and work

standards was emphasised in ten of the reviewed papers. Many presented moderation as a community-building or social practice, involving mentorship of new staff and early discussions to develop shared knowledge (Adie et al., 2013; Beutel et al., 2017; Crimmins et al., 2016; Grainger et al., 2016). Bird and Yucel (2013) highlight the value of forming a community of practice for knowledge sharing, but caution that high turnover among casual academics can compromise sustainability. Two studies focussed on using moderation to support casual academics, with one aimed at better understanding how they learn to grade student coursework to maintain academic standards (Boyd & Bloxham, 2014), and the other to support their professional development, and generate a sense of connectedness and collegiality (Crimmins et al., 2016).

Boyd and Bloxham (2014) stress the importance of “practical wisdom” and informal learning through moderation discussions. Mason and Robert (2023) and Mason et al. (2022; 2023) emphasise the need for in-depth discussions before teaching begins to shape learning activities. This early dialogue between coordinators, tutors, and casual staff, helps build positive relationships and fosters academic and professional confidence (Crimmins et al., 2016). In terms of the professional development inherent in these processes, Boyd and Bloxham (2014) explore the vulnerability faced by both markers and coordinators when addressing grade disagreements, which involves ongoing reflection and professional learning. These discussions often touch on participants’ inherent beliefs around moderation (Adie et al., 2013). Descriptions of the moderation process support the assertion that assessment moderation is complex and core academic work requiring recognition and resourcing as such (Czaplinski et al., 2014; Mason & Roberts, 2023).

### **Evaluation of moderation practices in higher education**

Evaluation of moderation practices were reported in three papers (Broadbent, 2018; Chen et al., 2017; O'Connell et al., 2016). Broadbent (2018) conducted a longitudinal study using focus groups and a national student satisfaction survey (Evalu8) to evaluate moderation effectiveness. Results showed improvement in feedback usefulness compared to previous surveys, with focus groups highlighting better understanding of criteria and feedback. Suggestions for improvement included providing timely moderator feedback and holding mid-marking informal meetings (Broadbent, 2018). Chen et al. (2017) focused on enhancing staff skills by trialling a training programme to enhance grade consistency. Using a before-and-after case study, they assessed rating consistency across dimensions like severity, accuracy, and centrality. The programme demonstrated increased consistency, decreased discrepancy, and a narrowing of mark ranges (Chen et al., 2017).

Three papers examined the effects of a social moderation process on marker variation (Bird & Yucel, 2013; Hunter & Docherty, 2011; O'Connell et al., 2016). Hunter and Docherty (2011) used quantitative and qualitative data from a grading moderation process in a large class to explore the presence and nature of underlying grader expectations, their impact on marking variation, and the effectiveness of socialisation in reducing such variations. In this paper, implicit assessment expectations held by academics consistently influenced grading decisions. The authors found that socialisation processes played a significant role in transferring beliefs related to written expression between graders; however, briefings and meetings alone were not sufficient. Tacit knowledge and active engagement were crucial, with negotiation required to establish a shared understanding of standards (Hunter & Docherty, 2011). O'Connell et al. (2016) also explored the problem of grader variability and whether a consensus moderation process can reduce the variability of assessors' judgements. The findings

indicated that the social calibration process enhanced assessors' confidence in grading and resulted in reduced variability in grades (O'Connell et al., 2016). Bird and Yucel (2013) found that structured pre-marking moderation, involving discussions on assessment criteria and standards, significantly improved marking reliability, supporting the idea that collaborative moderation reduces variability. Their study highlights that pre-marking techniques are more effective than post-marking methods, such as double marking, which are resource-intensive but potentially less impactful. Their DUAL programme's focus on shared understanding through discussions and exemplars fostered a community of practice, enhancing consistency in marking outcomes (Bird & Yucel, 2013).

### **Prevailing discourses shaping moderation practices in Higher Education**

Moderation practice is described by Adie et al. (2013) as driven by an idiosyncratic mix of beliefs that broadly fit under one or more discourses. Prevalent moderation discourses presented in the reviewed literature centre on equity, quality assurance, justification and process management, community building, shared understanding, and professional development. A focus on quality assurance, explores marker consistency, fairness, and equity, underpinning concepts of moderation as supporting the justification of, and confidence in, grading, as well as organisational accountability (Adie et al., 2013; Czaplinski et al., 2014). However, this lens may overlook issues related to marker bias, entrenched habits, and the complacency that can set in within stable marking teams (Grainger et al., 2008; Mason & Roberts, 2023; O'Connell et al., 2016).

The process management discourse focusses on methods to achieve efficient grading practices, including the application of "practice wisdom" or tacit knowledge (Boyd & Bloxham, 2014; Mason & Roberts, 2023) and effective communication (Evans &



Gardener, 2018). Whilst high quality marking that reflects student achievement is a priority, concerns about workload and throughput underscore the tension academics face between process efficiency, resource allocation, and maintaining quality (Beutel et al., 2017; Mason et al., 2023).

Boyd and Bloxham (2014) argue that moderation requires consideration of marker identity (whether harsh or lenient), which is shaped by the values held by each marker (Hunter & Docherty, 2011; Orr, 2007). These values often become clear in the collegial discussions integral to consensus moderation (Sadler, 2013). Discourses around professional development highlight the importance of building marker expertise and pedagogical understanding to build confidence and capacity (Crimmins et al., 2016). Boyd and Bloxham (2014) also emphasise the importance of combining subject discipline expertise and pedagogical knowledge in the grading process to uphold academic standards. It is evident that moderation discourse is mostly framed from the perspective of teachers and through an organisational lens. Beutel et al. (2017) question the assumption that current practices, such as sample marking and conference moderation, establish equitable grading. They note that variation in grading standards among markers can persist even after rigorous moderation processes have been undertaken as observed by Hunter and Docherty (2011) and Grainger et al. (2008). Bird and Yucel (2013) investigated the involvement of students in marking practice papers using the assessment rubric. Although the student perspective is not widely represented in the literature, their findings indicate that increasing transparency with students could help mitigate the inherent subjectivity of assessment moderation.

### **Institutional Guideline Quality Assessment**

The quality of 27 institutional moderation guidelines, including seventeen from universities in Australia and ten in the UK, was reviewed to address our fifth research question using the EduGET tool (Table 2). The quality and comprehensiveness of the guidelines varied across the five domains evaluated (Table 3). While the reviewed guidelines often effectively described their scope and purpose, quality was notably more limited in the domains of “stakeholder involvement”, “rigour of development”, and “applicability”. None of the guidelines indicated that a literature review was conducted to inform their development, and only two mentioned stakeholder involvement (Federation University, 2023; University of Wollongong, 2015). This suggests a potentially limited consideration of relevant staff and student perspectives and preferences. The highest-scoring guidelines, such as those from Federation University (2023) in the "applicability" domain, provided practical tools for implementing moderation policy and considered the resource implications of monitoring and audit roles to ensure a sustainable moderation framework.

### **DISCUSSION:**

Moderation primarily serves the purpose of improving the accuracy and reliability of grading and assessment decisions made by markers (Sadler, 2013). However, maintaining consistency in grading can be particularly challenging, especially in large courses that involve multiple assessors with varying levels of professional knowledge, expertise, and understanding of the required standards (Adie et al., 2013; Bird & Yucel, 2013; Mason et al., 2023; Watty et al., 2014). Demonstrating replicable practices is essential to effectively address this challenge, with social consensus and community of practice models of moderation predominating in the literature.

Consensus moderation offers the additional benefit of fostering collaborative professional development within supportive communities of practice, where a shared understanding of the adopted standards and rubrics can be established (Beutel et al., 2017; Bird & Yucel, 2013; Broadbent, 2018). Achieving agreement and consistency in moderation practices for the large cohorts typical in nursing education highlights the need for excellent team communication. Building an assessment community, where critical discussions can occur in a safe environment, may be more effective than an exclusive focus on assessment criteria for improving marker consistency and reliability (Evans & Gardner, 2018). However, facilitating effective consensus moderation practices, including professional development and the creation of safe, collegial moderation spaces, presents considerable complexity (Mason et al., 2023).

Faculty should consider strategies that involve markers early in assessment construction and training (Bird & Yucel, 2013; Chen et al., 2017; Mason & Roberts, 2023), monitor the quality of marking throughout the marking period (Broadbent, 2018), invest in capacity-building efforts (Mason et al., 2022; Mason et al., 2023), and build supportive communities of practice (Adie et al., 2013; Beutel et al., 2017; Bird & Yucel, 2013; Evans & Gardener, 2018; Grainger et al., 2016; Smith, 2012). While these strategies require investment, poor moderation practices can have significant consequences, not only in terms of equitable outcomes for students, but also in terms of academic role satisfaction and university reputation. The loss of embodied knowledge due to staff turnover, whether casual or permanent, remains of particular concern (Bird & Yucel, 2013).

The literature reviewed lacked a comprehensive exploration of success measures and key performance indicators for moderation practices. Identifying these indicators and selecting metrics that align with the interests of all stakeholders would assist in

implementing, monitoring, and sustaining rigorous moderation processes and quality assurance in a meaningful way.

Recognising the influence of individual and culturally adopted moderation discourse on practice is also crucial. Different perspectives highlight distinct aspects of assessment moderation. An equity perspective emphasises fairness and grading consistency among markers (Adie et al., 2013), while a community of practice perspective values collaboration and mentorship (Bird & Yucel, 2013). Prioritising a quality perspective leads stakeholders to focus on moderation's impact on student outcomes, grading validity, and metrics such as grade reliability, student satisfaction surveys, and grade reviews (Adie et al., 2013). While these practices may enhance student confidence and staff ability to respond to grading queries, it is important to promote a sense of “community” among markers, course coordinators and moderators. This focus on community can support the collaborative and mentoring culture, facilitating professional learning and building the tacit knowledge necessary for independent and reliable marking across large cohorts. The importance of supporting this collegial space for moderation is repeatedly emphasised in the literature.

Debate continues regarding the scope, timing and effectiveness of assessment moderation. Processes like double marking aim to ensure reliability across large student groups and teams, yet factors like inconsistent application of criteria, marker fatigue and varying levels of experience can affect outcomes (Smith, 2012). Consensus moderation encourages collaboration and supports newer markers. However, it relies on the development of safe communities of practice (Bird & Yucel, 2013; Sadler, 2013). Smith (2012) suggests that integrating moderation into a “whole-of-course” approach can build communities of practice, enhancing shared understanding and consistency in marking. However, making moderation part of an ongoing quality improvement cycle

requires sustained institutional support, resources, and staff training (Mason et al., 2023; Smith, 2012). Additionally, Mason et al. (2023) argue that moderation often serves organisational interests, driven by economic pressures and neoliberal policies. While rigorous moderation is promoted, university systems, through contract allocation, hours, workflows, and workloads, actively hinder genuine moderation efforts (Mason et al. 2022).

Despite numerous publicly available guidelines for institutional moderation, we found limited evidence of rigour and stakeholder involvement in their development. This paper, along with Poole's (2022) review of UK guidelines, reveals significant inconsistencies in quality. To improve moderation practices, comprehensive approaches to developing and implementing effective consensus moderation guidelines are needed. The **Educational Guideline Evaluation Tool (EduGET)** (Table 2) offers a structured framework for creating and evaluating such guidelines. Future research will focus on assessing and validating the EduGET tool to promote evidence-based moderation practices in teaching and learning.

#### Recommendation for future practice:

Faculty and institutions should adopt a multifaceted approach and consider marker participation across the assessment cycle, from assessment creation, through robust monitoring of marking, to capacity-building initiatives, marker training, and the cultivation of communities of practice. While these strategies necessitate investment and consideration of workload implications for staff, the ramifications of inadequate moderation practices can be substantial.

Building robust processes that prioritise safe communities of practice also necessitates the production of fit for purpose moderation guidelines, informed by input from end-

users, and grounded in evidence. We firmly advocate for the development of structured moderation guidelines that are comprehensive, clear, and evidence-based. Rigorous moderation guideline development requires adequate resourcing to include robust stakeholder consultation, ensuring applicability, transparency, and regular iterative review.

### **Strengths and limitations:**

This review offers valuable insights by synthesising evidence from the literature and conducting a quality assessment of institutional moderation guidelines. Its strength lies in the breadth of resources included, combining both academic research and publicly available institutional guidelines. However, reliance on Google search results to identify institutional guidelines may have led to the exclusion of relevant guidelines that appeared later in the search. Additionally, the guidelines assessed may have been updated since the review was completed. Another limitation is that the content validity of the **EduGET** tool remains untested, indicating a clear area for future research along with an in-depth content analysis of identified guidelines. Despite these limitations, this review and the accompanying guideline assessment tool contribute significant evidence supporting the need for more rigorous and consistent assessment moderation processes, while also calling for enhanced institutional guidelines to support quality in assessment practices.

### **CONCLUSION:**

This integrative review synthesises evidence on current moderation practices, covering key themes such as discourse, metrics, and strategies for staff engagement. It includes a comprehensive quality assessment of published moderation guidelines using a newly developed assessment tool, identifying areas for improvement. The findings emphasise the critical role of effective moderation in maintaining assessment quality. Based on the

evidence, educators and institutions are encouraged to develop and implement clear, evidence-based moderation guidelines, foster a collaborative moderation process among markers, tutors, and assessors to ensure consistency and shared understanding, and invest in ongoing professional development to support the application of moderation criteria. Additionally, allocating sufficient resources to support robust moderation practices is essential for maintaining high standards. These recommendations aim to enhance the effectiveness and efficiency of moderation in higher education, ultimately improving the quality and integrity of assessment outcomes.

#### **Declaration of generative AI and AI-assisted technologies in the writing process**

During the preparation of this work the author used ChatGPT 3.5 to improve the clarity and consistency of the writing style from our authors on different sections, after completion of the draft. The first and last authors reviewed and edited the content after using this AI tool, and the author team takes full responsibility for the content of the final draft for publication.

## REFERENCES

- Adie, L., Lloyd, M., & Beutel, D. (2013). Identifying discourses of moderation in higher education. *Assessment & Evaluation in Higher Education*, 38(8), 968-977. <https://doi.org/10.1080/02602938.2013.769200>
- Bausili, A. V. (2018). From piloting e-submission to electronic management of assessment (EMA): Mapping grading journeys. *British Journal of Educational Technology*, 49(3), 463-478. <https://doi.org/10.1111/bjet.12547>
- Beutel, D., Adie, L., Lloyd, M. (2017). Assessment moderation in an Australian context: processes, practices, and challenges. *Teaching in Higher Education*, 22(1), 1-14. <https://doi.org/10.1080/13562517.2016.1213232>
- Bird, F. L. & Yucel, R. (2013). Improving marking reliability of scientific writing with the Developing Understanding of Assessment for Learning programme. *Assessment & Evaluation in Higher Education*, 38(5), 536-553. <https://doi.org/10.1080/02602938.2012.658155>
- Boyd, P. & Bloxham, S. (2014). A situative metaphor for teacher learning: The case of university tutors learning to grade student coursework. *British Educational Research Journal*, 40(2), 337-352. <https://doi.org/10.1002/berj.3082>
- Broadbent, J. (2018). Large class teaching: How does one go about the task of moderating large volumes of assessment? *Active Learning in Higher Education*, 19(2), 173 –185. <https://doi.org/10.1177/1469787417721360>
- Brouwers, M. C., Kho, M. E., Browman, G. P., Burgers, J. S., Cluzeau, F., Feder, G., Fervers, B., Graham, I. D., Grimshaw, J., Hanna, S. E., Littlejohns, P., Makarski, J., & Zitzelsberger, L. (2010). AGREE II: Advancing guideline development, reporting and evaluation in healthcare. *Canadian Medical Association Journal*, 182(18), E839-842. <https://doi.org/10.1503/cmaj.090449>
- Chen, C.-Y., Chang, H., Hsu, W.-C., & Sheen, G.-J. (2017). Learning, behaviour and reaction framework: a model for training raters to improve assessment quality. *Assessment & Evaluation in Higher Education*, 42(5), 705-723. <https://doi.org/10.1080/02602938.2016.1180663>
- Crimmins, G., Nash, G., Oprescu, F., Alla, K., Brock, G., Hickson-Jamieson, B., & Noakes, C. (2016). Can a systematic assessment moderation process assure the quality and integrity of assessment practice while supporting the professional development of casual academics? *Assessment & Evaluation in Higher Education*, 41(3), 427-441. <http://dx.doi.org/10.1080/02602938.2015.1017754>
- Critical Appraisal Skills Programme. (2018). *CASP-Critical Appraisal Skills Programme*. <https://casp-uk.net/>
- Czaplinski, I., Adie, L., Senadji, B., & Beutel, D. (2014). *Analysis of moderation practices in a large stem-focused faculty*. International Conference of Teaching, Assessment and Learning (TALE), Wellington, New Zealand.



- Evans, C. & Gardener, J. (2018). What universities might learn from professional associations in marking large undergraduate modules. *Quality Assurance in Education*, 26(3), 343-348. <https://doi.org/10.1108/QAE-05-2017-0018>
- Federation University. (2023). *Marking and grading*. <https://federation.edu.au/staff/learning-and-teaching/teaching-practice/feedback/feedback-and-marking>
- Grainger, P., Adie, L., & Weir, K. (2016). Quality assurance of assessment and moderation discourses involving sessional staff. *Assessment & Evaluation in Higher Education*, 41(4), 548–559. <http://dx.doi.org/10.1080/02602938.2015.1030333>
- Grainger, P., Purnell, K., & Zipf, R. (2008). Judging quality through substantive conversations between markers. *Assessment & Evaluation in Higher Education*, 33(2), 133-142. <https://doi.org/10.1080/02602930601125681>
- Hunter, K. & Docherty, P. (2011). Reducing variation in the assessment of student writing. *Assessment & Evaluation in Higher Education*, 36(1), 109–124. <https://doi.org/10.1080/02602930903215842>
- Mason, J. & Roberts, L. D. (2023). Consensus moderation: the voices of expert academics. *Assessment & Evaluation in Higher Education*, 48(7), 926-937. <https://doi.org/10.1080/02602938.2022.2161999>
- Mason, J., Roberts, L., & Flavell, H. (2022). A Foucauldian discourse analysis of unit coordinators' experiences of consensus moderation in an Australian university. *Assessment & Evaluation in Higher Education*, 47(8), 1289-1300. <https://doi.org/10.1080/02602938.2022.2064970>
- Mason, J., Roberts, L. D., & Flavell, H. (2023). Consensus moderation and the sessional academic: valued or powerless and compliant? *International Journal for Academic Development*, 28(4), 468-480. <https://doi.org/10.1080/1360144X.2022.2036156>
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., for the PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *BMJ*, 339(b2535), 2-8. <https://doi.org/10.1136/bmj.b2535>
- O'Connell, B., De Lange, P., Freeman, M., Hancock, P., Abraham, A., Howieson, B., & Watty, K. (2016). Does calibration reduce variability in the assessment of accounting learning outcomes? *Assessment & Evaluation in Higher Education*, 41(3), 331–349. <http://dx.doi.org/10.1080/02602938.2015.1008398>
- Orr, S. (2007). Assessment moderation: constructing the marks and constructing the students. *Assessment & Evaluation in Higher Education*, 32(6), 645–656. <https://doi.org/10.1080/02602930601117068>
- Poole, B. (2022). Moderation: concept and operationalisation in UK universities. *Quality Assurance in Education*, 30(4), 464-476. <http://dx.doi.org/10.1108/QAE-12-2021-0203>

- Popay, J., Roberts, H., Sowden, A., Petticrew, M., Arai, L., Rodgers, M., & Britten, N. (2006). *Guidance on the conduct of narrative synthesis in systematic reviews: A product from the ESRC Methods Programme*. Lancaster University. <https://doi.org/10.13140/2.1.1018.4643>
- Sadler, D. R. (2013). Assuring academic achievement standards: from moderation to calibration. *Assessment in Education: Principles, Policy & Practice*, 20(1), 5–19. <http://dx.doi.org/10.1080/0969594X.2012.714742>
- Smith, C. (2012). Why should we bother with assessment moderation? *Nurse Education Today*, 32, 245-e48. <https://doi.org/10.1016/j.nedt.2011.10.010>
- Tertiary Education Quality and Standards Agency (TEQSA). (2021). *Higher education standards framework - Threshold standards 2021*. Commonwealth of Australia. <https://www.legislation.gov.au/F2021L00488/latest/text>
- Tertiary Education Quality and Standards Agency (TEQSA). (2017). *Review of the impact of the TEQSA Act on the higher education sector*. Commonwealth of Australia. <https://www.education.gov.au/higher-education-reviews-and-consultations/resources/review-impact-teqsa-act-final-report>
- Thomas, J. & Harden, A. (2008). Methods for the thematic synthesis of qualitative research in systematic reviews. *BMC Medical Research Methodology*, 8(45). <http://www.biomedcentral.com/1471-2288/8/45>
- University of Wollongong. (2015). *Faculty consultation forums-moderation of assessment at UOW*. <https://documents.uow.edu.au/content/groups/public/@web/@gov/documents/doc/uow194495.pdf>
- Watty, K., Freeman, M., Howieson, B., Hancock, P., O'Connell, B., de Lange, P., & Abraham, A. (2014). Social moderation, assessment and assuring standards for accounting graduates. *Assessment & Evaluation in Higher Education*, 39(4), 461-478. <https://doi.org/10.1080/02602938.2013.848336>
- Whittemore, R. & Knafl, K. (2005). The integrative review: updated methodology. *Journal of Advanced Nursing*, 52(5), 546-553. <https://doi.org/10.1111/j.1365-2648.2005.03621.x>
- Williams, A. H., Johnston, M. B., & Averill, R. (2023). Theorising a contextual framework for moderation of internal assessment: development and opportunities. *Educational Assessment, Evaluation and Accountability*, 35(3), 449-469. <https://doi.org/10.1007/s11092-022-09391-1>

Table 1: Extracted papers.

<b>Authors (Year) Country</b>	<b>Setting Sample</b>	<b>Design, Methods and CASP quality</b>	<b>Description of moderation discourse or process</b>	<b>Results/conclusion</b>
Adie et al. (2013) Australia	Faculty of Education. Large university.  Sample:25 of 90 faculty.  21 Coordinators.  8 Tutors (2 sessional)	Qualitative:  Semi structured interviews   Moderate Quality	<b>Equity:</b> Ensures consistent, fair assessment.  <b>Justification:</b> Enables confidence in decisions made, allowing academics to support feedback and grades given.  <b>Community Building:</b> Mentoring of new academics and purposeful involvement of the entire teaching team.  <b>Accountability:</b> Ensuring assessment quality, comparability, and coordinator's role as standard setter and expert responsible for the marks awarded to students.	<ul style="list-style-type: none"> <li>Discourses contribute to a complex relationship, reflecting tension between perspectives on learning and accountability within moderation practices.</li> <li>Moderation encompasses beliefs and experiences expressed through identified discourses and shaped by individual experiences and the social context.</li> <li>Despite outward compliance with policies, moderation practice will respond to participants' diverse perceptions and desired outcomes.</li> </ul>
Bausili (2018) UK	Faculty of Humanities, Manchester University  Practices across four schools	Qualitative: Participant observation   Moderate quality	Three types of moderation are observed: <ol style="list-style-type: none"> <li>Moderation by multiple markers,</li> <li>Moderation by a single marker either before or after providing grades/feedback to students, and</li> <li>Two forms of second marking (blind or open to peers)</li> </ol>	Three variables impacted grading variability: <ol style="list-style-type: none"> <li>The marker(s) initially responsible.</li> <li>The scrutiny method in the second marking.</li> <li>The timing of scrutiny in relation to feedback delivery to students</li> </ol>

Authors (Year) Country	Setting Sample	Design, Methods and CASP quality	Description of moderation discourse or process	Results/conclusion
Beutel et al. (2017) Australia	Australian university  Sample: 25 teaching academics,  21 Unit-coordinators  8 tutors	Qualitative:  Semi structured interviews  High quality	Sample marking, blind marking, conference moderation, and coordinator as final arbiter.  Holistic, involving conversations before, during, and at semester end.  Building a shared understanding through dialogue and collaborative development of assessments, criteria, standards, and evidence.  Aligning criteria, learning outcomes, assessment tasks, and pedagogy.	Change in Higher Education assessment culture required, to focus on accountability and enhancing teaching and learning.  <b>Key recommendations:</b>  1. Viewing moderation as a holistic process, 2. Actively engaging teaching team members in building a shared understanding of standards, 3. Linking moderation practices to learning and teaching improvement.  Involvement of students in assessment culture is also emphasised.
Bird & Yucel (2013) Australia	Zoology department  Sample:  Eight laboratory demonstrators	Quantitative:  Pre and post intervention  Moderate quality	Marker training, feedback provision, and sample comment preparation.  Three sample reports marked by all markers. Whole group discussion post-marking to achieve a consensus mark.  Students engage in the moderation process by sample marking in class using the rubric	The DUAL program enhanced marking consistency but did not reduce marker variation. It had no significant impact on marking efficiency but fostered a supportive community of practice. Pre-marking moderation discussions improved marker reliability and facilitated knowledge sharing

Authors (Year) Country	Setting Sample	Design, Methods and CASP quality	Description of moderation discourse or process	Results/conclusion
			to align their understanding with markers' expectations and criteria.	
Boyd & Bloxham (2014) UK	Three universities. Humanities, Art and design, Medicine, and Teacher education  Sample:25 Lecturers.	Qualitative: Semi-structured interview.  Moderate quality	Describes <b>second marking and moderation</b> processes for maintaining academic standards within teams of tutors. Discussion and evaluative feedback important	Collaborative discussions required to maintain academic standards.  Complexity of grading is highlighted. Much work done in isolation, tutors make holistic judgments and norm reference their grades. Subject, discipline, and pedagogical knowledge Important to maintain academic standards.
Broadbent (2018) Australia	Large University Health, Psychology departments  Focus groups and student evaluation survey.	Mixed methods: Survey and focus groups.  Moderate quality	Large marking team moderation described. Extensive marking guides provided. Videos explaining each criterion with concrete graded examples of student work. Marking guides and specific guidance feedback expectations provided.  Written examples of comments for criteria, including language, level of encouragement, and expected level of detail.	Provide clear marking guides, rubrics, and expectations. Use resources like videos to explain criteria and demonstrate performance levels.  Implement blind double marking early and periodically. Develop marker expertise to ensure consistency with ongoing training and support. Engage in feedback loops and calibration exercises.  Facilitate opportunities for marker discussions and feedback sharing

Authors (Year) Country	Setting Sample	Design, Methods and CASP quality	Description of moderation discourse or process	Results/conclusion
Chen et al. (2017) Taiwan	University in Taiwan Department of Information Management Sample: 5	Quantitative: Before and after case study design producing 512 items of training data  Moderate Quality	Training data from past papers.  1. Engage in formal practice exercises. 2. Comparing teacher and rater scores. 3. If accuracy meets threshold, progress; if not, practice more. 4. Mark papers or examinations after training. 5. Provide feedback to raters based on marking performance.	After specific LBR (Learning Behaviour Reaction) training  <ul style="list-style-type: none"> <li>Severity/leniency: Raters' strictness became more consistent.</li> <li>Accuracy/inaccuracy: Scoring discrepancies among raters decreased.</li> <li>Centrality/extremism: Raters' scores became more concentrated.</li> </ul> <p>However narrative responses need more time to assess, and model is less suitable for these more complex assessments.</p>
Crimmins et al. (2016) Australia	Regional university: Australia First year communication course.  Sample: 18 casual tutors, 2 full-time lecturers, 1 part-time lecturer, student cohort of 828	Qualitative: Case study / Audio recorded pre-moderation marking meeting. Focus group transcriptions.  HIGH QUALITY	Systematic approach combining pre-marking moderation, calibration, and post-marking expert review.  1. Tutors calibrate assessment standards before tutorials. 2. Tutors mark a sample assessment item. 3. Tutors discuss feedback, grades, and criteria, reaching consensus. 4. Tutors provide comments, end-of-paper comments, and a grade.	Systematic moderation supports,  <ul style="list-style-type: none"> <li>Quality assurance and self-checking focusing on tutor development, and expert review for standards.</li> <li>Promotes professional development and inclusion improving teaching practices, and , fostering a sense of belonging in team.</li> <li>Reduces anxiety and fosters collegiality despite initial anxiety it can establish respectful collegiality..</li> </ul>

Authors (Year) Country	Setting Sample	Design, Methods and CASP quality	Description of moderation discourse or process	Results/conclusion
			<ol style="list-style-type: none"> <li>5. Coordination team provides expert feedback on assessment standards.</li> <li>6. Marked assignments are checked against standards.</li> <li>7. Coordinators adjust for consistency if needed.</li> <li>8. Final processes: Ensure quality following moderation</li> </ol>	<ul style="list-style-type: none"> <li>• Alignment with quality assurance requirements:</li> </ul>
Czaplinski et al. (2014) Australia	<p>Queensland University</p> <p>Large, STEM-faculty.</p> <p>Course size: 6 to 1060</p> <p>Sample: 47 of 238 (19%) full-time permanent academics</p>	<p>Qualitative: Survey with open text responses</p> <p>Moderate Quality</p>	<ul style="list-style-type: none"> <li>• Reviewing samples individually that have been marked or organising moderation meetings to assess consistency.</li> <li>• Marking collectively during meetings to establish shared standards.</li> <li>• Checking consistency by moderating marks from multiple markers.</li> <li>• Calibrating marking through pre and post marking meetings with exemplars and sample questions.</li> <li>• Providing a clear marking scheme.</li> <li>• Focusing on feedback to students.</li> </ul>	<p>Resource limitations, inadequate training, and misaligned academic perspectives reduce effectiveness of moderation practice.</p> <ul style="list-style-type: none"> <li>• Surveyed STEM academics lacked common understanding of moderation.</li> <li>• Notions of clarity, consistency, fairness, and collaboration were shared among academics.</li> <li>• Moderation policy should consider moderation as an ongoing component of the teaching-learning cycle.</li> </ul>

<b>Authors (Year) Country</b>	<b>Setting Sample</b>	<b>Design, Methods and CASP quality</b>	<b>Description of moderation discourse or process</b>	<b>Results/conclusion</b>
Grainger et al. (2016)  Australia	Queensland  Regional university  Sample: 10 sessional staff	Qualitative phenomenologic al approach.  Structured/ semi-structured interviews  Lower quality	Social moderation- reaching consensus through conversations and professional dialogue,  <ul style="list-style-type: none"> <li>• Consistency of judgments through shared understandings,</li> <li>• Interpreting and applying judgments in a common way,</li> <li>• Sharing and grading representative samples of student work across different standards.</li> </ul>	Discourses identified: <ul style="list-style-type: none"> <li>• Building community,</li> <li>• Equity,</li> <li>• Justification, and</li> <li>• Accountability</li> </ul> Challenges: <ul style="list-style-type: none"> <li>• Staff complacency,</li> <li>• Reluctance to question judgments.</li> <li>• Moderation process structure, and Quality of rubrics.</li> </ul> Sessional staff faced issues related to <ul style="list-style-type: none"> <li>• continuity,</li> <li>• relationships, and</li> <li>• connecting to the community.</li> </ul> Additional barriers: <ul style="list-style-type: none"> <li>• attendance at meetings,</li> <li>• new staff induction,</li> <li>• managing ego and threatened staff,</li> </ul>
Grainger et al. (2008)  Australia	Regional University: Undergrad. Teacher education  Sample: 5 experienced university lecturer markers	Qualitative  Audio recordings of two 1-hour meetings included substantive conversation between markers.  Lower quality	All markers received assessment responses without the original grade.  <ul style="list-style-type: none"> <li>• Each marker had 10 minutes to review responses and make notes on how to assign a specific grade.</li> <li>• Markers engaged in a 30 to 45-minute discourse on levels of achievement and justification</li> </ul>	Tutors and lecturers assess quality against set criteria and standards, but student achievement levels are often influenced by comparison with others.  Content knowledge and technical proficiency are seen as key indicators of quality.  Judgments of quality may vary depending on factors like year level,



Authors (Year) Country	Setting Sample	Design, Methods and CASP quality	Description of moderation discourse or process	Results/conclusion
				<p>technical aspects (e.g., grammar, referencing, word count), and higher-level thinking skills (e.g., evaluation, justification).</p> <p>Consensus among markers on standard descriptors is crucial, but different processes are often used to determine grades despite criterion referencing</p>
<p>Hunter &amp; Docherty (2011) Australia</p>	<p>University of Technology, Sydney.  Large Macroeconomics subject  Sample:5</p>	<p>Mixed Methods  Five graders on a sample of 10 papers at stage 2, and three papers at stage 4  Moderate Quality</p>	<p>Rigorous large class moderation process involving five stages:</p> <p><b>Stage 1:</b> Development of assessment criteria and grading standards.</p> <p><b>Stage 2:</b> Initial grading of a sample of papers by graders.</p> <p><b>Stage 3:</b> Grader meeting to discuss and resolve differences.</p> <p><b>Stage 4:</b> Second grading of a sample of papers by the same graders.</p> <p><b>Stage 5:</b> Final meetings to discuss and resolve any remaining differences in grading standards.</p>	<p>Significant variation in grading standards among markers found, even after moderation process. Variation influenced by beliefs about structure, content, analysis, and written expression.</p> <p>Socialisation (meetings, comments) helped transfer beliefs about written expression.</p> <p>Persisting variation in grading across criterion sub-categories identified</p> <p>Briefings alone not sufficient; tacit knowledge and active engagement among markers is crucial. Even criterion-referenced approaches</p>

Authors (Year) Country	Setting Sample	Design, Methods and CASP quality	Description of moderation discourse or process	Results/conclusion
				<p>still subject to multiple interpretations.</p> <p>Negotiation among markers necessary for establishing shared beliefs about standards</p>
<p>Mason &amp; Roberts (2023)</p> <p>Australia</p>	<p>Australian metropolitan university</p> <p>50,000 students 3000 staff</p>	<p>Qualitative: Semi structured interviews</p> <p>High Quality</p>	<p>Exploration of consensus moderation includes:</p> <ul style="list-style-type: none"> <li>• Pre-teaching discussions to establish shared assessment understanding.</li> <li>• Ongoing dialogue to prepare for marking and moderation.</li> <li>• Calibration and dialogue to establish and uphold standards.</li> <li>• Developing markers' understanding of assessment requirements, expected standards, and the tacit knowledge/expertise used for holistic marking.</li> </ul>	<p>Consensus moderation is complex. Adequate resourcing for calibration activities and dialogue is crucial to developing tacit knowledge and maintaining marking consistency.</p> <p>Challenges include:</p> <ul style="list-style-type: none"> <li>• Time constraints /heavy workloads limit effective dialogue.</li> <li>• Higher education culture emphasising individual achievement and competition, causing tension among markers.</li> <li>• Inherent subjectivity in marking.</li> <li>• Novice markers less vocal about concerns/ tension with experienced markers.</li> </ul>

Authors (Year) Country	Setting Sample	Design, Methods and CASP quality	Description of moderation discourse or process	Results/conclusion
Mason et al. (2023) Australia	As above	Focus groups with experienced <b>sessional academics</b> .  High quality	Consensus moderation described as collaborative approach to quality assurance involving markers coming together to discuss and agree on mark allocation.  Dialogue and engagement among markers create a shared understanding of the marking criteria to ensure that the allocated marks reflect the standard of work submitted.	Challenges for sessional academics included. <ul style="list-style-type: none"> <li>• Power imbalances,</li> <li>• Insufficient training, and</li> <li>• Time constraints.</li> </ul> Practice based professional development required to enhance quality assurance.  Coordinators should dedicate more time to developing the marking skills of sessional academics, through early discussions and calibration activities.  Further research needed to investigate the social factors influencing decision-making processes in consensus moderation.
Mason et al. (2022) Australia	As above	Focus groups with <b>Unit coordinators</b> .  High quality	As above	Consensus moderation identified as: <ol style="list-style-type: none"> <li>1. collaborative.</li> <li>2. Illusory and</li> <li>3. about marker management.</li> </ol> Unit coordinators used varying approaches: from positive collaboration through to more managerial control practices.

Authors (Year) Country	Setting Sample	Design, Methods and CASP quality	Description of moderation discourse or process	Results/conclusion
				Echoing literature concerns regarding variation in consensus moderation practices
O'Connell et al. (2016) Australia	Seventeen Accounting departments.  Australian universities	Quantitative Intervention.  30 assessors  Control 15 assessors  Self and Peer Assessment Resource Kit  Moderate quality	Assessors evaluated non-qualitative work of three students.  Entered their judgments and comments in an online repository called Self and Peer Assessment Resource Kit PLUS (SPARKPLUS).  For reassessment, all previous marks, and comments for samples were removed.  Assessors were instructed to reassess the same three samples as if it was their initial evaluation	Shows effectiveness of workshop calibration in <ul style="list-style-type: none"> <li>Reducing assessment variability.</li> <li>Boosting assessors' confidence, and</li> <li>Fostering shared standards understanding in moderation practices.</li> </ul> Post-workshop reassessment revealed improved consistency.  Advocates for investing in workshop calibration to cultivate shared standards understanding and reduce assessor variability.
Orr (2007) UK	English University Art and design dept	Qualitative epistemological lens  Art and design lecturers/student s  10 hours of non-participant semi-structured observations	Collaborative discussions among lecturers to assess and align marks, in evaluating undergraduate art and design studio work. <ul style="list-style-type: none"> <li><b>Positioning and Evaluation:</b> Lecturers review students' visual artwork and engage in open-</li> </ul>	Two contrasting perspectives on moderation presented: <ul style="list-style-type: none"> <li><b>Positivist approaches</b> emphasise objective criteria and standardised measures for evaluating student performance, relying on</li> </ul>

Authors (Year) Country	Setting Sample	Design, Methods and CASP quality	Description of moderation discourse or process	Results/conclusion
		Moderate quality	<p>ended conversations, evaluating general aspects of each piece.</p> <ul style="list-style-type: none"> <li>• <b>Assessment conversations</b> lasting up to 15 minutes, multiple markers discuss initial marks.</li> <li>• <b>Funnelling Process:</b> Discussion until consensus is reached on the final mark.</li> </ul>	<p>quantitative data and a single, agreed-upon assessment standard.</p> <ul style="list-style-type: none"> <li>• <b>Poststructuralist</b> perspectives challenge singular, objective standards, recognising the social /contextual aspects of assessment.</li> </ul> <p>Highlights importance of discourse in constructing meaning, considering multiple perspectives and power dynamics in assessment practices.</p>
Poole (2022)	Ten UK Universities	Guideline review Moderate quality	<p>The distinction between moderation and marking is sometimes unclearly articulated.</p> <p>The text discusses specific terms for types of marking, such as "first marking," "second marking," "double marking," "blind marking," and "double-blind marking."</p>	<p>There is variability in how moderation is implemented across reviewed guidelines. Some Universities include drafting, redrafting, and approval of assessment artifacts, as part of moderation whereas others do not.</p> <p>Moderation and marking are separate activities, but some documentation discusses both the activities in tandem</p>
Watty et al. (2014) Australia	10 universities: Accounting	Qualitative. Workshops in three locations:	Social moderation in accounting education.	Process promoted active learning and reflection among assessors to ensure

Authors (Year) Country	Setting Sample	Design, Methods and CASP quality	Description of moderation discourse or process	Results/conclusion
	and Accountants	(63 assessors)  Moderate quality	<p>Multiple assessors collaborate to attain a shared understanding of assessment criteria and standards.</p> <p>Assessors actively engage in learning and reflection to align assessment outcomes.</p> <p>Describes calibration intervention, including a literature review on peer learning, valid assessment practices, and social moderation theory.</p> <p>Social moderation to improve assessment task design for valid evaluation of threshold learning standards</p>	<p>shared understanding of threshold learning standards.</p> <p>Qualitative data revealed that this process aided assessors in comprehending and designing tasks for valid assessment.</p> <p>Process mitigated divergence and variations in interpretations and understanding of standards, thereby achieving consistency in judgments and assessment grades across multiple assessors.</p> <p>Assessors identified gaps in knowledge, variances in learning standards, and differences in interpretation across contexts</p>

**Table 2: Guideline Evaluation Tool (EduGET)\***

<b>Domain 1: Scope and Purpose</b>	Yes, No or Can't tell
1.1. The overall objective(s) of the guideline is (are) specifically described.	
1.2. The situations in which the guidelines apply are clearly stated.	
1.3. The staff to whom the guideline is meant to apply are specifically described.	
<b>Domain 2: Stakeholder Involvement</b>	
2.1. The guideline development group includes individuals from all the relevant stakeholders (discipline heads, CCs, Sessional)	
2.2. The relevant staff views and preferences have been sought (i.e., sessional, and faculty).	
2.3. Student views have been sought.	
2.4. The target users of the guideline are clearly specified.	
2.5. The guideline has been piloted among end users.	
<b>Domain 3: Development Rigour</b>	
3.1. Systematic methods were used to search for evidence.	
3.2. The guideline has been extensively reviewed by experts.	
3.3. A procedure for updating the guideline is provided.	
<b>Domain 4: Clarity of Presentation</b>	
4.1. The guidelines are specific and unambiguous.	
4.2. The different options for moderation are clearly presented.	
4.3. The process of moderation is clearly described.	
<b>Domain 5: Applicability</b>	
5.1. The guideline provides advice and/or tools on how the guidelines can be put into practice.	
5.2. The guideline describes potential organisational facilitators and barriers to its application.	
5.3. The potential resource implications of applying the recommendations have been considered.	
5.4. The guideline presents key review criteria for monitoring and/ or audit purposes.	

\*Five domains and some questions reference the Agree II Clinical practice guideline assessment tool (Brouwers et al.,2010)

**Table 3:** Guideline Evaluation Tool overall results

<b>Guidelines (n=27)</b>	Scope and purpose (Three questions)	Stakeholder involvement (Five questions)	Rigour of development (Three questions)	Clarity of presentation (Three questions)	Applicability (Four questions)
Yes	74%	19%	6%	56%	22%
No or can't tell.	26%	81%	94%	44%	78%

*Percentages are the numbers of "yes" scores, over all responses for each domain.*

### **Figure 1: PRISMA diagram**

#### **Highlights**

- Reviewing moderation methods and discourse, staff engagement and measurements of the success in current literature.
- Use of guideline assessment tool to review the quality of publicly available moderation guidelines and enhance development of practical guidance around moderation practice.
- Cultivating supportive communities of practice around assessment moderation
- Encouraging early involvement of markers in the moderation cycle.
- Understanding the impact of moderation discourse and belief on practice
- Investing in professional development to enhance assessment moderation practice among tenured and sessional staff.