Scientific Research Publishing

Advances in Applied Sociology, 2026, 16(11), 909-922
https://www.scirp.org/journal/aasoci
ISSN Online: 2165-4336
ISSN Print: 2165-4328

Understanding Fairness in Marking Postgraduate Assessments: Towards an Equitable Framework University of Cambridge

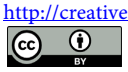
Moamer Gashoot[1*], Tahani Mohamed[2]

[1]Department of Design and Engineering, Faculty of Science and Technology, Bournemouth University, Bournemouth, UK
[2]Accounting, Finance and Economics Department, Business School, Bournemouth University, Bournemouth, UK
Email: *mgashoot@bournemouth.ac.uk

How to cite this paper: Gashoot, M., & Mohamed, T. (2026). Understanding Fairness in Marking Postgraduate Assessments: Towards an Equitable Framework University of Cambridge. *Advances in Applied Sociology, 16,* 909-922.
https://doi.org/10.4236/aasoci.2026.1611053

Received: September 15, 2026
Accepted: November 21, 2026
Published: November 24, 2026

Copyright © 2026 by author(s) and Scientific Research Publishing Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
http://creativecommons.org/licenses/by/4.0/



Open Access

Abstract

This study was undertaken as part of postgraduate work within the **University of Cambridge Faculty of Education** and examines fairness in marking postgraduate assessments, with particular attention to Design and Engineering education. Fairness is fundamental to the credibility, validity, and integrity of higher education assessment; however, achieving equitable marking can be particularly challenging in disciplines where assessment combines creative judgement, technical competence, analytical reasoning, and professional decision-making. The study therefore explores how postgraduate marking practices can be strengthened through a more equitable, transparent, and consistent approach. Drawing on assessment theory and literature concerning fairness, validity, reliability, constructive alignment, marker judgement, rubrics, moderation, and inter-rater reliability, the study critically reflects on assessment practice within a Level 7 Competitive Product Development unit. Particular attention is given to subjective professional judgement, variations in assessors' interpretations of criteria, potential marker bias, and the complexities associated with evaluating creative and technical work. The literature review indicates that fairness cannot be achieved through standardised marking criteria alone. Instead, equitable assessment requires a shared understanding of academic standards supported by transparent criteria, constructive alignment between intended learning outcomes and assessment, clearly defined rubrics, inclusive assessment practices, marker training, calibration exercises, double-marking, and structured moderation. Greater student engagement with assessment criteria, rubrics, and exemplars is also identified as important for improving transparency and understanding of expected standards. The study con-

cludes that fairness in postgraduate marking should be conceptualised as a multidimensional principle encompassing validity, reliability, transparency, consistency, inclusion, accessibility, and equitable opportunities for students to demonstrate achievement. It proposes that a collaborative and systematically moderated approach to marking can reduce subjective bias, strengthen assessment reliability, and contribute towards a more equitable framework for postgraduate assessment practice.

Keywords

Assessment Fairness, Postgraduate Education, Design and Engineering, Marking Reliability, Constructive Alignment, Assessor Bias, Moderation, Educational Equity

1. Introduction

Arguably, higher education institutions derive their legitimacy and authority, in part, from fair and equitable assessment practices. In particular, this essay will seek to explore postgraduate Design and Engineering programs. Assessment fairness transcends mere numerical scoring; it encompasses the fundamental principles of transparency, consistency, and construct validity in evaluating student achievement.

This study investigates the challenges inherent in ensuring equitable marking practices at the postgraduate level, where complex project-based assessments and diverse student cohorts are found. Shaw and Nisbet's (2021) work emphasises that inequitable assessment practices can significantly impact students' educational trajectories and career opportunities. However, the discourse has mostly focused on undergraduate assessment, leaving a notable gap in understanding the unique dynamics of postgraduate evaluation, particularly in technical disciplines.

The research aims to develop an innovative framework for enhanced assessment fairness through three primary objectives:

1) Establishing transparent and consistent marking criteria that align with learning outcomes.

2) Implementing inclusive assessment practices that accommodate diverse learning styles and backgrounds.

3) Developing robust moderation processes to ensure marking reliability.

This study's impact and significance lie in its potential to transform assessment practices, moving beyond traditional grading paradigms to create a more nuanced, equitable, and effective evaluation system. The proposed framework seeks to balance academic rigour with fairness, ensuring that assessment practices support both student success and maintain the high standards expected in postgraduate education. In particular, to address one issue in depth and as hinted by the title, there will be a focus on marking rather than other aspects of assessment.

Research Question

How can marking practices within the Level 7 Competitive Project Development (CPD) unit be improved to enhance fairness, transparency and consistency while preserving the validity of professional academic judgement?

This research question provides the central focus for the study and guides the subsequent literature review, critical analysis of current marking practices and discussion of potential improvements. The conclusion will consider how the findings can inform a practical framework for fair and reliable postgraduate marking while maintaining academic standards.

Methodological approach: This paper adopts a literature-informed reflective case analysis rather than a primary empirical research design. Its practice evidence is limited to the CPD unit characteristics and assessment arrangements described here: a 20-credit Level 7 unit taught in 2023/24, a 3000-word project, four equally weighted marking domains, student-selected product or company cases, and the reported involvement of two staff members in marking. These features are considered alongside published scholarship to identify plausible sources of inconsistency and to formulate improvements. No student scripts, cohort size, individual marker scores, systematic observations or interviews are analysed. The discussion therefore cannot establish whether the proposed changes reduce bias or increase reliability; those claims require prospective evaluation.

1.1. Purpose of This Report

The purpose of this work is to understand the issue of fairness in marking assessments (Camilli, 2006, 2013; Kane, 2010; Nisbet & Shaw, 2019, 2020; Shaw & Nisbet, 2021; Tierney, 2013, 2016; Worrell, 2016). This will help me as an assessment developer to design and deliver sound, fair, and useful assessments for the learner technically. It will be of benefit to policymakers and university administrators too. Reliability and validity are sometimes conflated. In the context of assessment marking, particularly for Design and Engineering, the former could refer to the demonstration that the operations and procedures can be repeated by other educators who then achieve similar findings with a similar cohort of students. In the literature on reliability, it is well known that this is more of an issue in creative undertakings like this CPD. We will discuss this shortly.

Construct validity usually refers to the degree to which a test or other measure assesses the underlying theoretical construct it is supposed to measure (i.e., the test is measuring what it is purported to measure) (Brown & Dowling, 2009). Assessors, particularly in my context of design and engineering, usually try to measure learning outcomes and evidence of applied and original thinking. Fairness is a part of, yet distinct from assessment validity. It aligns with the key questions proposed by Messick (1980) and Elwood (2013): (i) "Does the test accurately measure the characteristics it claims to assess?" and (ii) "Is the test appropriate for the intended purpose and context?" Both questions are vital in ensuring that assessments not only measure accurately but also do so equitably across diverse student groups, for example concerning neurodivergent or hard-of-seeing or hearing

students (as well as other needs that department administrators can access). In my role, this perspective shapes my approach to developing assessments that are both fair and valid (Messick, 1980; Elwood, 2013).

In our previous discussions, it has been discussed that while fairness is necessary for validity, it alone does not guarantee valid results. Sound construct validity in assessment alone does not guarantee fairness. Fairness is something more, and it is to this that the discussion now turns.

1.2. Definition of Fairness and Its Importance in Marking

According to ETS' Standards for quality and fairness (ETS, 2014: p. 50), a fair test is one that "reflects the same construct(s) for all test takers, and scores from it have the same meaning for all individuals in the intended population". I have suggested that transparency, consistency, and construct validity are all part of fairness in postgraduate assessments. Fairness in marking is specifically important because it ensures that all students are evaluated equitably, reflecting a closer fidelity to their abilities and efforts (Nisbet & Shaw, 2019). A fair assessment process builds trust between students and educators, fostering a positive learning environment where students feel their work is judged on merit rather than subjective bias. For example, while AI-based assessment methodologies represent an interesting and exciting avenue of educational advancement in assessment, even outsourcing judgement on assessments, in part or in whole, to artificial intelligence technologies and adjacent technologies would not be without human subjective bias (see Colonna (2024) for discussion; see Rahm (2021) for discussion of dangers).

Nevertheless, fairness is also essential for maintaining academic integrity, as it promotes consistency, transparency, and accountability in the grading process. For me, maintaining fairness in assessments goes beyond upholding academic integrity; it is about providing every student with an equal opportunity to showcase their abilities and reach their full potential.

1.3. Overview of the Unit

The Competitive Product Development (CPD) module is a 20-credit unit for master's-level design students. Taught over the second semester of 2023/2024, students critically evaluated an existing design business and explored new market opportunities for retail products. The module aimed to deepen understanding of product design development, considering factors like political, social, and environmental pressures. Delivered through lectures, seminars, 3D design activities, and videos, the course culminated in a 3000-word project submission, providing a practical assessment of students' analytical and creative skills in design.

Table 1 below sets out a proposed constructive-alignment map for the CPD project. These are working formulations, not verified quotations from the approved unit guide; the learning outcomes and criteria must be checked against the official documentation before implementation.

Table 1. The proposed constructive-alignment map for the CPD project.

Proposed CPD learning outcome	Marking criterion (weight)	Required student evidence
Critically evaluate an existing design business	Business and competitive analysis (25%)	Evidence-based analysis of the business, market position and competitors
Investigate market opportunities and external pressures	Market and contextual research (25%)	Credible market sources; political, social and environmental analysis
Develop and justify a competitive product direction	Design strategy and creative problem-solving (25%)	Product opportunity, alternatives considered and justified design decisions
Integrate findings into a defensible development proposal	Critical synthesis and professional justification (25%)	Integrated recommendation, constraints and reflective evaluation

2. Literature Review: Theories and Perspectives on Marking

There are several theories and perspectives on marking that are worth highlighting. The first is what we might refer to as "Marker Bias Theories". The halo effect refers to how initial impressions influence subsequent judgments (Dennis, 2007) specifically examines the halo effect in educational grading. In this work, Dennis shows that even experienced assessors showed a significant halo effect when grading. This is notable because one might expect experts to be more objective.

There is also the priming effect which highlights the impact of previous marks on the current assessment (Spear, 1997); central tendency bias, which is the tendency on the part of assessors to avoid extreme ranges in marks across a student cohort (Landy & Farr, 1980); severity and leniency biases, which involve consistent over- or under-marking (Ecclestone, 2001) and order effects, which refer to how the sequence of marking affects judgement (Hughes, Keeling, & Tuck, 1983).

2.1. Fairness, Reliability, and Validity as Distinct yet Related in Marking

The implementation of a high-quality assessment system is essential to ensure fairness and validity (Nisbet & Shaw, 2019; Tierney, 2013). Although fairness in assessment is increasingly recognised as significant, there is a pressing need for greater clarity in the definition of the term 'fair' (Nisbet & Shaw, 2019; Tierney, 2013).

Furthermore, fairness has gained prominence in educational assessment to the extent that some scholars suggested that it has succeeded the "big two" (validity and reliability) to become part of the "big three" in assessment theory (Messick, 1989, 2000; Näsström, 2008; Worrell, 2016). I understand validity and reliability as vital and inter-related components of fairness though including other components too. I will elaborate on key types of fairness below. Key aspects of fairness include ensuring that different assessors assign similar marks to identical work (inter-rater reliability) and that the marks effectively differentiate among varying levels of student achievement (Daw, 2023; Tierney, 2013).

So, while it can be seen that validity and fairness are undeniably interconnected, they represent distinct concepts. In practice, however, when designing and mark-

ing an assessment, it's unlikely that one would focus solely on fairness or validity in isolation. Both elements must be considered together to ensure a well-rounded and effective evaluation (Elwood, 2013; Scott et al., 2014).

Drawing on Nisbet and Shaw's (2019) fairness frameworks, we can understand marking practices as several distinct types of frameworks: procedural fairness (ensuring consistent processes), substantive fairness (acknowledging different approaches to design problems), and transparent fairness (clear communication of assessment criteria). These frameworks help illuminate the inherent tensions in assessing creative and technical work.

2.2. Application and Relevance to a Design and Engineering Context

Assessment in this kind of market research CPD presents unique challenges due to the interplay between technical precision and creative expression. In creative disciplines, there is often no single "correct" solution, making consistent assessment more complex (Orr, 2007). To do so, it would be worth unpacking specific challenges in assessing creative/technical work. To address these challenges, higher education institutions have set about actioning various quality assurance mechanisms. Price et al. (2011) describe how mark review processes and appeal procedures provide important safeguards for assessment fairness. External moderation requirements help maintain standards across different institutions and provide independent verification of marking practices (Bloxham, 2009).

Standardisation practices typically include pre-marking standardisation meetings, sample marking exercises, and post-marking reviews to identify and address any systematic issues (Sadler, 2013). Solutions to these challenges require structured approaches to reduce bias. Bloxham et al. (2016) emphasise the importance of regular marker training and calibration exercises. The implementation of robust moderation systems requires regular sampling of marked work across grade boundaries and cross-checking between markers (Price et al., 2011). Documentation and transparency measures should include clear communication of assessment criteria to students which I discuss at 3.1.

2.3. The Literature on Quality in Marking

In my view, Biggs's (2003) work on constructive alignment in curriculum theory is a critical part of orienting towards a fair and valid approach in marking because it minimises the disconnect between what is taught, what is designed from taught or research modules, and assessment to connect it back to the teaching. As part of this, the scoring or marking criteria in higher education ought to align with the structure of the content being assessed, ensuring that the weights assigned to different criteria are appropriately adjusted, and aligned with ILOs, rubrics, and marking criteria. This is especially important in this particular CPD, where subjective interpretations of creative decisions in market research may seem unconvincing or hard to justify to markers and can lead to inconsistencies in marking.

To elaborate on the mitigation of bias and the promotion of transparency, the

literature on quality marking emphasises several key approaches to ensuring assessment reliability and fairness. Inter-rater reliability theories focus on the consistency of marks between different assessors, with studies showing that structured training and clear assessment criteria can significantly improve marking consistency (Bloxham et al., 2016; Sadler, 2013). Standardisation approaches have evolved to include detailed marking schemes, exemplar answers, and calibration exercises, which help establish a common understanding among markers before assessment begins.

Rubric development and implementation have emerged as vital components of fair assessment. Well-designed rubrics provide clear performance descriptors across different grade levels, helping to reduce subjective interpretation and increase marking consistency (Jonsson & Svingby, 2007; Reddy & Andrade, 2010). The effectiveness of rubrics is particularly evident in design and engineering education, where complex projects often require evaluation across multiple dimensions.

Double-marking systems, while resource-intensive, offer additional quality assurance. Research indicates that independent double-marking followed by reconciliation discussions can identify and resolve marking discrepancies effectively. Yorke (2011) discusses various assessment approaches including double marking. This approach is particularly valuable for high-stakes assessments or when evaluating creative work where professional judgment plays a significant role.

Moderation processes serve as a final quality control mechanism, ensuring consistency across different markers and student cohorts. These processes typically involve reviewing samples of marked work, analysing mark distributions, and addressing any systematic differences between markers.

2.4. Theories of Fairness in Marking

In previous literature, several models have been proposed to conceptualise fair assessment. Here, I critically review three key models. One of the most comprehensive models was introduced by Kunnan (2004) (Association of Language Testers in Europe, 2004), outlining five key features: validity, absence of bias, access, administration, and social consequences (Moghadam & Nasirzadeh, 2020).

In my experience as an assessor, Kunnan's model resonates, particularly in design and engineering education, where ensuring fairness across varied skill sets is critical. To unpack this further, some might excel at technical drawing, others at creative conceptualisation, while some might be stronger in analytical market research.

Considering validity in light of how I have written the CPD assignment brief this would mean ensuring the assessment (3000-word project) measures what it claims to measure. As such, when students critically evaluate an existing design business, the assessment should capture both their analytical abilities and design thinking, not just their writing skills. I believe the assessment does this by using

the module's four marking criteria and learning outcomes. The 3000-word format allows students to demonstrate their ability to evaluate products and develop competitive frameworks. Each marking criterion is weighted equally (25%), ensuring a balanced assessment of different competencies.

Secondly, students can choose their own product/company to analyse, allowing them to work with familiar contexts. Assessment criteria focus on analytical ability rather than prior industry knowledge which the majority of students do not have. This is particularly crucial when students approach design problems differently. For instance, one student might focus on environmental sustainability while another emphasises market viability—both approaches could be equally valid if well-justified. This is potential absence of bias.

A third application of Kunnan's fairness model is access. Our students should have equal access to resources needed for the assessment. In our unit, this might mean ensuring all students can access the necessary design software, research materials, and support for both theoretical and practical components—something which is not being comprehensively done.

The fourth feature relates to administration. The assessment process should be consistent and clear which I believe it is. Although academic, several colleagues have suggested ways for the language of the assignment brief to be clearer and the rubric establishes ways for students to understand what is guiding markers' conception of analytical abilities and design thinking with clarity. In other words, there are relatively transparent criteria for how I evaluate both the business analysis and design development aspects.

The final feature in the model is social consequences meaning the assessment should consider its impact on students' future development. In design education, this might mean ensuring the assessment helps prepare students for real-world product development challenges.

These five key features from Kunnan's model also resonate with Nisbet and Shaw's (2019) fairness frameworks (mentioned at the beginning of 2.2), where we can understand marking practices as several distinct types of frameworks. For instance, procedural fairness (ensuring consistent processes) may be found in the administration feature where the assessment process should be consistent and clear; likewise in the equal weighting of marking criteria ensuring balanced assessment of different competencies, and in the consistent use of four key domains in the rubric for evaluating all students.

There is substantive fairness (acknowledging different approaches to design problems) by acknowledging one student might focus on different subjects in their CPD—both approaches could be equally valid if well-justified. This is further supported by allowing students to choose their own product/company to analyse. Overall, I have aimed for a recognition of varied skill sets.

Transparent fairness (clear communication of assessment criteria) is potentially located in the rubric which clearly defines performance levels across the four key domains. Clear marking criteria are linked to learning outcomes using the module's four marking criteria and learning outcomes.

I have seen that emotional responses to perceived unfairness can be particularly intense, something that resonates with anyone who has experienced the strong reactions often seen in such situations, including in educational settings. This view has been further supported and expanded by other scholars (Camilli, 2006; Kane, 2010; Kane, 2013; Willingham et al., 1997), who argued that assessments must be designed and administered in a manner that guarantees equitable treatment for different subgroups of candidates across different vectors of difference including gender, ethnic, bodies, long-term health conditions and so on. In the past, I have been made aware of students who are hard-of-hearing. There are various assistive technologies made available by the university and supportive mechanisms I can offer as an educator. These include speaking to students after class or during office hours about the speed and clarity with which I speak and encouraging them to sit near the front during lectures and seminars. This is part of a broader conception of fairness.

As I have discussed above, fairness in this assessment process is acknowledging individual differences between learners, ensuring that assessments are appropriate for the entire student population in terms of their interests, though not necessarily learning styles and pace. That is to say, there is still a 3000-word essay with a set deadline that applies to all students, although some can apply for extenuating circumstances extension through our department.

To mark work in the Level 7 Competitive Product Development unit, I am committed to ensuring fairness through the use of detailed rubrics and up-to-date marking criteria as I have described with Kunnan's model and Shaw and Nesbit's works in this second section. These rubrics provide clear, consistent guidelines that reflect current industry standards and academic expectations. By regularly reviewing and updating these criteria, I ensured that the work produced was assessed not only according to the latest developments in the field but also in a way that was consistent with previous years (Shaw & Nisbet, 2021). This approach guarantees that students' performance is judged fairly, with transparent criteria applied consistently across cohorts, protecting the integrity of both the marks awarded and the value of their qualifications.

3. Practice

3.1. How Can I Make Assessment Fairer as a Senior Lecturer?

Ensuring Fairness for Competitive Product Development (CPD): It is important to design, develop, administer, and evaluate assessments to ensure they accurately measure the intended skills or knowledge while minimising the influence of irrelevant factors. For any new or significantly revised assignment, there is a strategy that addresses fairness throughout the design, development, administration, and application of the assessment or service (ETS, 2014).

Evaluations of fairness of the Unit plan: ETS guidelines recommended in the operation that the assignment design should be intended to eliminate symbols, language, and content that generally regarded as racist or offensive, in addition to

the above I have to review the teaching material, assignment brief, this will ensure the learner receive a fair assessment the meet the slandered required in higher education (ETS, 2014).

3.2. Recommendations

Another challenge I encounter is the lack of clarity in marking criteria. We have endeavoured to make our rubric less ambiguous in several ways. Jonsson and Svingby (2007) suggest a few key strategies to improve the clarity and transparency of such rubrics, which I believe we have followed for this CPD. Firstly, breaking down broad criteria into more granular, separate aspects like 'Innovation in problem-solving'. Next, providing clear, detailed descriptors for each performance level that explicitly define the expected standards of achievement. This gives students a much clearer understanding of how their work will be evaluated.

Third, involving students in the development and refinement of the rubric. Getting student feedback can help ensure the criteria and descriptors are meaningful and understandable from the learner's perspective. Using exemplars or sample work with students to illustrate the different performance levels too. This helps calibrate students' (and markers') understanding of what meets the criteria. Finally, training our peers, markers, extensively on using the rubric consistently which I will discuss below.

4. Implications for Policy and Practice

Perceptions of fairness in different types of assessments

Students' perspective on fairness in marking:

One of the key issues that I consider is inconsistency in marking, which can lead to perceptions of unfairness among students. This inconsistency often arises when academic staff interpret marking criteria differently, leading to varying grades for similar work (La Marca, 2001).

Enhancing fair assessment practices benefits students by helping them understand how evaluations contribute to developing their design skills and how to effectively self-assess, both of which are crucial for lifelong learning. These skills have been shown to positively impact performance, enabling students to adopt a learning approach aligned with the intended learning outcomes (Bhola et al., 2005). Students' perceptions of assessments strongly influence their learning behaviour, making clarity and fairness in assessments essential. When students fully comprehend the task requirements, they are more likely to perform at their best, fostering positive learning experiences (Chai et al., 2024).

Assessors' perspective on fairness in marking:

As part of my plan for ensuring fairness in assessments, I will attempt to implement several key strategies in marking among my department. At present, two members of staff mark each student's work. Members of staff come from fairly different backgrounds: some have worked in industry, for large, commercial builders. Others have exclusively spent their lives in academia, or some years in

industry too, and have more of a predisposition to the artistic and theoretical. Nevertheless, peer discussions about students' work, as lively as they sometimes can be, are for the most part, at least in my experience, grounded in fact and not value-based judgements. As a department, I would suggest that we need more marker training and a formally sound marker monitoring process that goes beyond weekly meetings and advisory documents.

This CPD, like the wider discipline in which it is situated, combines creativity and technical precision; such discrepancies can undermine the fairness of assessments (Orr & Bloxham, 2012). To address this, I believe regular calibration sessions among staff are essential to align our understanding of marking criteria. This would promote a more transparent and reliable assessment process that accurately reflects students' efforts and abilities.

Calibration sessions might look like structured marking exercises where markers assess the same pieces of work independently and then discuss their decisions. This is what Price et al. (2011) recommend. These sessions should include a discussion of exemplars representing different grade boundaries. Regular comparison of marking standards through blind double-marking of samples. Documentation of decisions made during calibration to build a shared understanding over time.

Calibration sessions promote transparency in several ways. For students they can be shown examples of discussions from calibration sessions that illustrate how different quality levels are distinguished; shown how markers reach decisions potentially helps them better understand grade criteria and they can see how different aspects of their work (technical vs creative) are weighted in assessment.

For staff, regular discussions make assessment criteria more explicit. Different interpretations of standards are openly debated rather than remaining implicit or confined to comments in a document window. These decisions and rationales can be documented for future reference and updated too.

Next, other formal monitoring processes like cross-checking of marks across different marker pairs (Sadler, 2013); more sophisticated, statistical analysis of marking patterns to identify systematic differences between markers and regular monthly or twice monthly moderation meetings to discuss borderline cases and resolve discrepancies. External examiners from other higher education institutions, especially if they have been recognised by their peers for excellence in marking, could also visit to review marking standards.

This leads to another aspect which is the development of shared understanding which should be made explicit. Rust et al. (2003) emphasise the importance of 'socialising' markers into a shared assessment community. Again, there is an emphasis on regular workshops to discuss the interpretation of criteria and the development of annotated exemplars. Building a repository of case studies showing different quality levels is worth having.

5. Conclusion

In conclusion, this project aimed to enhance the fairness of assessments by taking

a consistent, shared social approach to marking among academic peers—one that not only takes actions but also scrutinises the theoretical perspectives and approaches laid out. By implementing the recommendations outlined in this essay, including clear communication of criteria, clear, well-defined rubrics, structured marker training, and inclusive assessment practices, my peers and I can work towards a more equitable and effective evaluation system. This, in turn, will foster a positive learning environment where students feel their work is judged on merit rather than subjective bias.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

Association of Language Testers in Europe (2004). *European Language Testing in a Global Context: Proceedings of the ALTE Barcelona Conference July 2001.* Cambridge University Press.

Bhola, D. S., Impara, J. C., & Buckendahl, C. W. (2005). Aligning Tests with States' Content Standards: Methods and Issues. *Educational Measurement: Issues and Practice, 22,* 21-29. https://doi.org/10.1111/j.1745-3992.2003.tb00134.x

Biggs, J. (2003). *Aligning Teaching for Constructing Learning*. Higher Education Academy. https://s3.eu-west-2.amazonaws.com/assets.creode.advancehe-document-manager/documents/hea/private/resources/id477_aligning_teaching_for_constructing_learning_1568036613.pdf?utm_source=chatgpt.com

Bloxham, S. (2009). Marking and Moderation in the UK: False Assumptions and Wasted Resources. *Assessment & Evaluation in Higher Education, 34,* 209-220. https://doi.org/10.1080/02602930801955978

Bloxham, S., den-Outer, B., Hudson, J., & Price, M. (2016). Let's Stop the Pretence of Consistent Marking: Exploring the Multiple Limitations of Assessment Criteria. *Assessment & Evaluation in Higher Education, 41,* 466-481. https://doi.org/10.1080/02602938.2015.1024607

Brown, A., & Dowling, P. (2009). *Doing Research/Reading Research: A Mode of Interrogation for Education* (2nd ed.). Routledge.

Camilli, G. (2006). Test Fairness. In R. L. Brennan (Ed.), *Educational Measurement* (4th ed, pp. 221-256). American Council on Education/Praeger.

Camilli, G. (2013). Ongoing Issues in Test Fairness. *Educational Research and Evaluation, 19,* 104-120. https://doi.org/10.1080/13803611.2013.767602

Chai, F., Ma, J., Wang, Y., Zhu, J., & Han, T. (2024). Grading by AI Makes Me Feel Fairer? How Different Evaluators Affect College Students' Perception of Fairness. *Frontiers in Psychology, 15,* Article 1221177. https://doi.org/10.3389/fpsyg.2024.1221177

Colonna, L. (2024). Teachers in the Loop? An Analysis of Automatic Assessment Systems under Article 22 GDPR. *International Data Privacy Law, 14,* 3-18. https://doi.org/10.1093/idpl/ipad024

Daw, M. (2023). Mark Distribution Is Affected by the Type of Assignment but Not by Features of the Marking Scheme in a Biomedical Sciences Department of a UK University. *Assessment & Evaluation in Higher Education, 48,* 806-819. https://doi.org/10.1080/02602938.2022.2134552

Dennis, I. (2007). Halo Effects in Grading Student Projects. *Journal of Applied Psychology, 92,* 1169-1176. https://doi.org/10.1037/0021-9010.92.4.1169

Ecclestone, K. (2001). 'I Know a 2:1 When I See It': Understanding Criteria for Degree Classifications in Franchised University Programmes. *Journal of Further and Higher Education, 25,* 301-313. https://doi.org/10.1080/03098770126527

Educational Testing Service (ETS) (2014). *ETS Standards for Quality and Fairness.* Educational Testing Service. https://www.ets.org/pdfs/about/standards-quality-fairness.pdf

Elwood, J. (2013). Educational Assessment Policy and Practice: A Matter of Ethics. *Assessment in Education: Principles, Policy & Practice, 20,* 205-220. https://doi.org/10.1080/0969594x.2013.765384

Hughes, D. C., Keeling, B., & Tuck, B. F. (1983). Effects of Achievement Expectations and Handwriting Quality on Scoring Essays. *Journal of Educational Measurement, 20,* 65-70. https://doi.org/10.1111/j.1745-3984.1983.tb00190.x

Jonsson, A., & Svingby, G. (2007). The Use of Scoring Rubrics: Reliability, Validity and Educational Consequences. *Educational Research Review, 2,* 130-144. https://doi.org/10.1016/j.edurev.2007.05.002

Kane, M. (2010). Validity and Fairness. *Language Testing, 27,* 177-182. https://doi.org/10.1177/0265532209349467

Kane, M. T. (2013). Validating the Interpretations and Uses of Test Scores. *Journal of Educational Measurement, 50,* 1-73. https://doi.org/10.1111/jedm.12000

Kunnan, A. J. (2004). Test Fairness. In M. Milanovic, & C. Weir (Eds.), *European Language Testing in a Global Context* (pp. 27-48). CUP.

La Marca, P. M. (2001). Alignment of Standards and Assessments as an Accountability Criterion. *Practical Assessment, Research, and Evaluation, 7,* Article 21. https://doi.org/10.7275/ahcr-wg84

Landy, F. J., & Farr, J. L. (1980). Performance Rating. *Psychological Bulletin, 87*, 72-107. https://doi.org/10.1037/0033-2909.87.1.72

Messick, S. (1980). Test Validity and the Ethics of Assessment. *American Psychologist, 35,* 1012-1027. https://doi.org/10.1037/0003-066x.35.11.1012

Messick, S. (1989). Meaning and Values in Test Validation: The Science and Ethics of Assessment. *Educational Researcher, 18,* 5-11. https://doi.org/10.3102/0013189x018002005

Messick, S. (2000). Consequences of Test Interpretation and Use: The Fusion of Validity and Values in Psychological Assessment. In R. D. Goffin, & E. Helmes (Eds.), *Problems and Solutions in Human Assessment: Honoring Douglas N. Jackson at Seventy* (pp. 3-20). Springer. https://doi.org/10.1007/978-1-4615-4397-8_1

Moghadam, M., & Nasirzadeh, F. (2020). The Application of Kunnan's Test Fairness Framework (TFF) on a Reading Comprehension Test. *Language Testing in Asia, 10,* Article No. 7. https://doi.org/10.1186/s40468-020-00105-2

Näsström, G. (2008). *Measurement of Alignment between Standards and Assessment.* Department of Educational Measurement, Umeå University.

Nisbet, I., & Shaw, S. (2020). *Is Assessment Fair?* SAGE Publications Ltd. https://doi.org/10.4135/9781529739480

Nisbet, I., & Shaw, S. D. (2019). Fair Assessment Viewed through the Lenses of Measurement Theory. *Assessment in Education: Principles, Policy & Practice, 26,* 612-629. https://doi.org/10.1080/0969594x.2019.1586643

Orr, S. (2007). Assessment Moderation: Constructing the Marks and Constructing the Students. *Assessment & Evaluation in Higher Education, 32,* 645-656.

https://doi.org/10.1080/02602930601117068

Orr, S., & Bloxham, S. (2012). Making Judgements about Students Making Work: Lecturers' Assessment Practices in Art and Design. *Arts and Humanities in Higher Education, 12,* 234-253. https://doi.org/10.1177/1474022212467605

Price, M., Carroll, J., O'Donovan, B., & Rust, C. (2011). If I Was Going There I Wouldn't Start from Here: A Critical Commentary on Current Assessment Practice. *Assessment & Evaluation in Higher Education, 36,* 479-492. https://doi.org/10.1080/02602930903512883

Rahm, L. (2021). Education, Automation and AI: A Genealogy of Alternative Futures. *Learning, Media and Technology, 48,* 6-24. https://doi.org/10.1080/17439884.2021.1977948

Reddy, Y. M., & Andrade, H. (2010). A Review of Rubric Use in Higher Education. *Assessment & Evaluation in Higher Education, 35,* 435-448. https://doi.org/10.1080/02602930902862859

Rust, C., Price, M., & O'Donovan, B. (2003). Improving Students' Learning by Developing Their Understanding of Assessment Criteria and Processes. *Assessment & Evaluation in Higher Education, 28,* 147-164. https://doi.org/10.1080/02602930301671

Sadler, D. R. (2013). Assuring Academic Achievement Standards: From Moderation to Calibration. *Assessment in Education: Principles, Policy & Practice, 20,* 5-19. https://doi.org/10.1080/0969594x.2012.714742

Scott, S., Webber, C. F., Lupart, J. L., Aitken, N., & Scott, D. E. (2014). Fair and Equitable Assessment Practices for All Students. *Assessment in Education: Principles, Policy & Practice, 21,* 52-70. https://doi.org/10.1080/0969594x.2013.776943

Shaw, S., & Nisbet, I. (2021). Attitudes to Fair Assessment in the Light of COVID-19. *Research Matters: A Cambridge Assessment Publication, 31,* 6-21.

Spear, M. (1997). The Influence of Contrast Effects upon Teachers' Marks. *Educational Research, 39,* 229-233. https://doi.org/10.1080/0013188970390209

Tierney, R. D. (2013). Fairness in Classroom Assessment. In J. H. McMillan (Ed.), *SAGE Handbook of Research on Classroom Assessment* (pp. 124-144). SAGE Publications, Inc. https://doi.org/10.4135/9781452218649.n8

Tierney, R. D. (2016). Fairness in Educational Assessment. In M. Peters (Ed.), *Encyclopedia of Educational Philosophy and Theory* (pp. 1-6). Springer. https://doi.org/10.1007/978-981-287-532-7_400-1

Willingham, W. W., Cole, N. S., Bridgeman, B., Dwyer, C. A., Johnson, L. M., Leung, S. W., Lewis, C., Morgan, R. L., Pollack, J. M., & Schmitt, A. (1997). Fairness Issues in Test Design and Use. In W. W. Willingham, & N. S. Cole (Eds.), *Gender and Fair Assessment*. Routledge.

Worrell, F. C. (2016). Commentary on Perspectives on Fair Assessment. In N. J. Dorans, & L. L. Cook (Eds.), *Fairness in Educational Assessment and Measurement* (pp. 283-293). Routledge. https://doi.org/10.4324/9781315774527-18

Yorke, M. (2011). Summative Assessment: Dealing with the 'Measurement Fallacy'. *Studies in Higher Education, 36,* 251-273. https://doi.org/10.1080/03075070903545082