AdvanceHE

MEMBER BENEFIT

Unconscious bias literature review: bias in assessment

Hannah Borkin

:

Contents

1. Introduction

1.1 Background to the literature review

Interest in the prevalence of bias in different settings has increased over the past few years, however literature reviews researching bias within the realm of higher education remain sparse. This includes Equality Challenge Unit's review 'Unconscious Bias in Higher Education' (Cornish and Jones), published in 2013.

While there are numerous publications recognising the importance of understanding how unconscious processes can influence our behaviour and decision-making, Advance HE believes its members would welcome an updated series of concise literature reviews, investigating the prevalence of bias in teaching and learning in higher education. The reviews will cover the following topics:

+ bias in assessment
+ bias in the curriculum and online pedagogy
+ bias in decision-making.

They will also include a review of good practice initiatives currently in place in higher education institutions, to help to create an evidence base of 'what works?' in tackling bias in teaching and learning.

This report is the first within the series and covers academic and grey literature, and examples of 'what works?' relating to bias in student assessment.

1.2 Methodology

The search for existing literature was conducted using two main databases (EBSCO and Scopus), as well as targeting grey literature to find current examples of 'what works' in tackling bias in higher education assessment practices. The latter included a search of university websites, as well as blogposts related to teaching and learning.

Advanced Boolean search terms related to: (i) bias; (ii) marking; (iii) assessment; and (iv) peer assessment were used to find publications that were relevant to the overarching theme of bias in assessment.

As this mini-series of literature reviews are aimed at discussing our current knowledge of bias and what works in mitigating it, documents published in the last five to ten years were prioritised. Article abstracts and summaries were used to assess their relevance then exported using the online software Zotero to develop a longlist of 30 publications.

Upon further examination of the content of each document, the longlist was reduced to a shortlist of 20 articles (15 from academic sources and five from grey sources). The shortlist was organised into the following three themes:

+ anonymous marking and the feedback process
+ multiple choice questions and the use of negative marking
+ peer assessment methods.

Each of these themes posit that bias already exists, and are therefore an exploration of methods that help to reduce or eliminate it (rather than establish its presence).

1.3 Limitations

The practice of searching publications for a literature review is not value neutral and will invite a degree of subjectivity and bias. Advance HE encourages its researchers to reflect on the potential biases they may bring to literature reviews and to put measures in place, such as the process of creating a longlist and shortlist of relevant documents.

Advance HE also recognise that there are gaps within the literature review, such as studies that engage with characteristics other than race and gender (for example, internationalisation, disability including mental health and social class). This is partly due to the concise nature of the review, and we therefore invite members to submit any current examples of 'what works' in mitigating bias in these areas. Examples of good practice can be sent to equalityresearch@advance-he.ac.uk.

2. Summary of findings

2.1 Anonymous marking and the feedback process

2.1.1 Overview

Existing literature on the topic of anonymous marking largely suggests that it continues to be a contentious practice within higher education, with few published studies proving that it helps to eliminate, or reduce, bias. For example, Hinton and Higson (2017) noted that the practice of anonymous marking did little to address already well-documented attainment gaps. Instead, the majority of literature settled on the argument that rather than reducing bias per se, anonymous marking helped to improve students' perceptions of fairness and transparency.

The practice of anonymous marking was often entangled with the feedback process, with studies investigating the potential negative implications anonymising feedback had on the student learning experience. For example, Whitelegg (2002) stated that anonymous feedback 'disrupts the feedback loop by removing the individualisation of feedback comments, increasing the distance between staff and students'. Pitt and Winstone (2018) further concluded that there was a clear 'trade-off to be made between enhancing students' belief in the learning potential of feedback and ensuring perceptions of fairness and transparency on the other'.

While the practice of anonymisation in student assessment has so far not produced a profoundly positive result, it should not necessarily be assumed to be ineffective in all areas. For example, there are a number of additional factors that may reduce or improve the impact of anonymous marking, such as discipline, type of assessment, students' degree level, etc. In 2017, the Irish Research Council (2018) also introduced gender-blind assessment of PhD and post-doctoral applications, and found that STEM awards to women increased from 35% in 2013, to 57% in 2017.

Further research is required to assess the overall impact of anonymous marking in a higher education setting, and to investigate its impact applying an intersectional lens. At present, it appears that students' perceptions of fairness in assessment practices are just as – if not more - important to students than tangible evidence of eliminating, or reducing, bias.

2.1.2 Review of current literature

Research into the practice of anonymous marking in higher education settings began to pick up speed following Higson and Hinton's publication in 2017. They argued that while performance differences in written examinations were relatively well researched, few studies examined the efficacy of anonymous marking, particularly in modern student populations. The research analysed a large archive (N=30,674) of assessment data spanning a 12-year period, and investigated significant differences between assessment marks and factors such as ethnic group, gender and socio-environmental background. When anonymous marking

was introduced, it appeared that this had a negligible effect in reducing any biases between these groups. The research called for a re-examination of the efficacy of anonymous marking in reducing group performance differences, as it could be argued that similar research conducted in alternative settings - such as a small and specialist institution - may yield different results.

Pitt and Winstone (2018) further investigated the role of anonymous marking, and reported similar results to Higson and Hinton (2017) in terms of its negligible effect on eliminating bias. Instead they researched it from a different angle, and explored the perception of 'fairness' students had towards assessment practices, based on the suggestion by National Union of Students (NUS) in 2008 that anonymous marking 'reduces both the fear and likelihood of discrimination'. The NUS recommendation was bolstered by findings in the 2005-2007 National Student Survey (Surridge, 2008), wherein BME students cited that their main cause of dissatisfaction with assessment and feedback was due to perceptions of unfair assessment and a lack of transparency in marking arrangements compared to white students, in turn leading to demands of anonymous marking.

An additional question explored in this study was whether the practice of anonymous marking influenced the nature of feedback given to the student, and the impact of this feedback on the student's future learning and development. It was found that students perceived feedback on non-anonymously marked work to be a more positive experience than receiving feedback on work that had been marked anonymously, highlighting the importance of keeping feedback non-anonymous and echoing Whitelegg's comments on the potential disruption of the feedback loop if it is to be anonymised (2002).

The most recent study on the subject of anonymised feedback was by Winstone and Boud (2020), who further explored its relationship with anonymous marking. Their publication, 'The need to disentangle assessment and feedback in higher education' argued that assessment and feedback tend to get conflated within higher education settings. They concluded that although the practices often coexist, if institutions were to adopt anonymous marking practices, then they must strike a balance between the needs of anonymity with ensuring feedback continues to be meaningful and personal to students.

2.1.3 'What Works?' – Institutional examples of good practice

It is interesting that the majority of UK-based institutions appear to have adopted the practice of anonymous marking despite little evidence within the literature to suggest that it eliminates, or reduces, the prevalence of bias. A number of institutions released a press release announcing the introduction of anonymous marking, to commence in either the 19/20 or 20/21 academic year, which raised the question as to whether a waterfall effect was created once one institution introduced the new measures. The concluding words within Brennan's publication on anonymous assessment of summative work (2008) shed light on the likely reason for this:

"Perhaps it is hinted at by a staff comment about anonymous assessment within UG: '*It is difficult to justify not doing it*'. This suggests some kind of general acceptance that while knowing the identity of the student author of a piece of written assessment is a comfort, it is also a dangerous one that cannot be justified if avoidable."

As an example, in 2019 Middlesex University announced the increased adoption of anonymous marking across the University, to support full implementation in 2020/21. However, supporting existing research detailed in the section above, they specified that it is only used for summative assessments and outlined cases where the practice is not possible, including during the feedback process:

+ formative assessments and information feedback and reflection
+ where the assessor is also the supervisor
+ assessment topics that are individual or group-based
+ assessments that rely on observations.

A second example is from Bournemouth University, which introduced anonymous marking for the 19/20 academic year and alluded to perceptions of fairness within the statement. As mentioned on their website, this was 'introduced to improve the perception of fairness in the assessment process following consultation with BU students and a successful pilot scheme'.

For the institutions that have applied a blanket anonymous marking scheme, for which there are some, they should consider referring to Winstone and Boud's 2020 study, and ensure anonymous marking does not get conflated with the feedback process.

2.2 Multiple Choice Questions (MCQs) and the use of negative marking

2.2.1 Overview

Multiple choice questions (MCQs) are often used as an assessment method when testing large groups of people, due to a relatively quick turnaround for providing feedback. However, a number of studies highlighted concern with MCQs as a result of a possible gender bias that became prevalent when adopting this method. In particular, it was found that women tended to perform less favourably when undertaking examinations that used MCQs alongside negative marking (whereby a point is deducted for an incorrect answer), due to gendered norms relating to risk-aversion. For example, Pekkarinen's (2015) study noted that women performed worse than men in the negatively-marked MCQ entrance examination for Finnish universities, and skipped more questions than was optimal in order to maximise the probability of being accepted.

In an attempt to address this issue, several institutions have since adopted new and alternative marking practices to use in tandem with MCQs. In some cases, this has included the removal of negative marking altogether, and replaced with a process of elimination testing (whereby a student is rewarded for partial knowledge, by indicating whether they think each of the offered answers are incorrect or not).

As this instance of bias within the use of MCQs and negative marking is related to stereotype threat and aversive behaviour, it is possible that the impact could go beyond gender, and could similarly impact students from other underrepresented groups and their intersections (ie. BAME students or students with a disability).

While increased choice of assessment (including the implementation of new and alternative marking practices) could be a way to reduce bias, O'Neill (2017) argued that this should again be approached with caution. Within this study, O'Neill investigated perceptions of fairness on the procedures and outcomes when choice of assessment increased. The results were complex, for example it was found that postgraduate students were more open to assessment choice in comparison to first-year students, possibly due to their evolving competencies. O'Neill stated that while it is important that equity is designed into choice of assessment to ensure fairness, this choice must accommodate diverse circumstances and should be considered when implementing new assessment methods.

2.2.2 Review of current literature

Firstly, a research paper by Espinosa and Gardeazabel (2020) analysed the gender differences in student performance in MCQs. Within their study, it was confirmed that women tended to omit more items, scored fewer correct answers and thus achieved lower grades than men.

A study examining gender effects in different types of undergraduate science assessments yielded similar results to the above (Kacprzyk, Parsons, Maguire and Stewart, 2019). Within this study, the authors conducted research assessing the performance of first-year undergraduate science students from the University College Dublin. It was found that when marking the modules of male and female students using a combination of three types of evaluation, including continuous assessments and MCQs scored with/without negative marking, a significant bias was consistently observed for the cohort of students who undertook the MCQ examinations with negative marking. It should be noted that while the presence of gender identities or expressions outside the gender binary (ie. non-binary students) were acknowledged, they were not included in this study.

The results of Kacprzyk, Parsons, Maguire and Stewart's research discovered a significant bias against women who undertook the MCQ examination with negative marking in place. The authors concluded that this could be partially linked to a gender difference in a willingness to guess, and that preliminary data suggested that this disappeared after the removal of negative marking. This could suggest that it is not the practice of MCQs and

negative marking that are biased, rather the varying student responses to this type of assessment introduce bias into their overall performance.

2.2.3 'What Works' – institutional examples of MCQs and use of negative marking

A couple of institutions in Belgium have thus adopted alternative marking and assessment practices, including the removal of negative marking when using MCQs.

In 2015, KU Leuven presented a research paper which highlighted their concerns with negative marking, as they confirmed it deterred risk-averse students. They investigated whether a process of elimination testing (whereby students were assessed for their partial knowledge by indicating whether each of the offered alternatives were correct or not), would mitigate this gender bias. They found that with the introduction of this alternative method, student performance and satisfaction increased, while anxiety decreased.

Based on the evidence surround MCQs, the University of Ghent also decided to halt the use of negative marking within their assessment practices, and instead replaced it with 'standard setting'. In standard setting, grades are converted afterwards and the student taking the examination has to answer more than half the questions correctly in order to pass. However, a correct answer will always result in a positive score, and wrong unanswered questions result in zero points. They hoped that this change would instil confidence in their students to answer all questions within a multiple choice examination.

These examples chime with Kacprzyk, Parsons, Maguire and Stewart's research (2019), in which they recommended that caution should be exercised while using negative marking, and that regular reviews of assessment strategy should be implemented by higher education institutions to ensure gender-bias free assessment.

2.3 Peer assessment methods

2.3.1 Review of current literature

While peer assessment has been praised for its value in helping to create an effective learning environment, to encourage the development of team management skills among graduates, and to foster a more collaborative culture between students, a number of studies have noted that there is a possibility of bias within its application.

In a 2017 study published by the University of York '*Is Reflective Writing an Effective Peer Assessment Tool for Students in Higher Education?'* the authors argued that peer assessment tools heightened the potential of showing bias to close friends or peers during the marking process. It therefore considered the use of a reflective writing among peers as an alternative means of limiting bias. The reflective writing exercise encouraged students to not only engage with the higher education learning process, but also to learn from each other by giving and receiving critical feedback. It then asked students to justify their scores

by writing out a critical reflection. It concluded that reflective writing within peer assessment was a pivotal team work skill and helped to reduce instances of bias among peers.

Another method used within the practice of peer assessment is Alternative Comparative Judgement (ACJ). A study by Jill Barber in 2018 discussed the findings from a small pilot used in the fourth year of the Manchester Pharmacy programme, whereby 50 students peer-marked their own mock examinations using ACJ. ACJ presents an alternative to conventional marking, in which the assessor compares two answers and chooses a winner. This process continues until the sorting algorithm leads to scripts sorted in order of merit. Within the pilot, it was found to be a very promising technique for peer-assessment, and helped to remove any subjective bias – especially useful for assessment types where there is an element of subjectivity required in the marking and which are incompatible with more rigid marking schemes.

It will be interesting to ascertain whether recent events, such as the issues created by Ofqual's exam results algorithm adopted in 2020, will impact on future student or institutional engagement with automated marking processes. A paper presented by Marcinkowski, Keislich, Stake and Lünich (2020) focused on the fairness concerns that are gaining momentum in both academic and public discourses due the use of algorithmic decision-making (ADM). It concluded that universities aiming to implement ADM systems should, where possible, take students' fairness perceptions into account.

2.3.2 What works – institutional examples of Alternative Comparative Judgement (ACJ)

Both the University of Edinburgh and the University of York have explored the benefits of ACJ within teaching and learning blogposts and forums. The University of Edinburgh have found that the assessment method has been successful when marking a large quantity of examinations; producing more reliable results that are less susceptible to marker variation in comparison to traditional marking methods. The University of York also stated that ACJ, which allowed for an alternative to individual (or double) marking, had a positive impact on the reliability of the assessment.

3. Conclusion and further recommendations

While existing literature on bias highlights little evidence of the prevalence of bias within non-anonymous marking, institutions are nonetheless mindful of its risk and have proceeded with anonymous marking as their chosen method, in turn helping to improve students' perceptions of fairness.

Conversely, clear evidence of gender bias has been discovered within other assessment practices such as Multiple Choice Questions and the use of negative marking, and it is recommended that these practices should be approached with caution.

New and alternative assessment methods such as Alternative Comparative Judgement are therefore gaining interest, and could be carefully applied once consultation with students or a piloting of the approach has taken place. Furthermore, any increase in choice of assessment should take into account the diverse needs of a student population.

Based on the above, there are a number of clear recommendations institutions could adopt to ensure students are receiving the fairest possible assessment experience:

+ Conduct regular departmental reviews of your assessment policies and strategy, and consider whether this matches up with the institutional policy. Ensure any policy and practice changes around assessment undergo an Equality Impact Assessment before implementation.
+ Adopt a transparent approach to assessment, and communicate regularly with students in relation to assessment and feedback practices. Consult directly with students through the use of surveys and focus groups if planning to make any changes.
+ If adopting anonymous marking practices, ensure feedback and assessment do not get conflated by keeping feedback non-anonymous. Where possible, only apply anonymous marking to summative assessments, rather than formative.
+ If using multiple choice questions (MCQs), avoid the use of negative marking as an accompanying assessment approach. Research has found that this creates a gender disparity, in that women are potentially more risk-averse when answering questions.

4. Bibliography

#Bournemouth University (2019) *Anonymous marking launched for all students*. Available at: www.bournemouth.ac.uk/news/2019-11-04/anonymous-marking-launched-all-students

#Ghent University, (2020) *Education and Examination Code*. Available at: www.ugent.be/student/en/class-exam-exchange-intern/class-exam/education-examination-code/multiplechoice.html

#Middlesex University (2019) *Anonymous marking to be extended across the university*. Available at: unihub.mdx.ac.uk/student-life/announcements/2019/anonymous-marking-to-be-extended-across-the-university

#University of York (2019) *Can Adaptive Comparative Judgment achieve more consistency in assessment than traditional academic judgment?* Available at: yorkforum.org/2019/07/02/comparative-judgement/

#Irish Research Council (2018) *Gender Strategy and Actions*. Available at: research.ie/assets/uploads/2018/08/04108-IRC-Gender-flyer-proof03-single.pdf

Baruah, B, Ward, T and Jackson, N (2017) *'Is reflective writing an effective peer assessment tool for students in Higher Education?'*, 16th International Conference on Information Technology Based Higher Education and Training (ITHET), Ohrid, 2017. pp 1-6. DOI: 10.1109/ITHET.2017.8067802

Brennan, D (2008) University student anonymity in the summative assessment of written work, *Higher Education Research & Development*, 27:1, pp 43-54. DOI: 10.1080/07294360701658724

De Laet, T (2015) *Diversity in Engineering Education: "How to remove the gender bias in multiple choice assessments in engineering education?",* Proceedings of the 43rd SEFI Annual Conference 2015, SEFI 2015. Web.

Whitelegg, D (2002) Breaking the feedback loop: problems with anonymous assessment, *Planet*, 5:1, pp 7-8. DOI: 10.11120/plan.2002.00050007

Pitt, E, Winstone, N (2018) The impact of anonymous marking on students' perceptions of fairness, feedback and relationships with lecturers, *Assessment & Evaluation in Higher Education*, 43:7, pp 1183-1193. DOI: 10.1080/02602938.2018.1437594

MP, Espinosa, Gardeazabal, J (2020) 'The Gender-bias Effect of Test Scoring and Framing: A Concern for Personnel Selection and College Admission'. *The B.E. Journals in Economic Analysis & Policy,* Vol.20 (3). Web.

Marcinkowski, F, Kieslich, K, Starke, C, and Lünich, M (2020) *Implications of AI (un-)fairness in higher education admissions: the effects of perceived AI (un-)fairness on exit, voice and organizational reputation*. In Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency. 122–130. DOI: 10.1145/3351095.3372867

O'Neill, G (2017) It's not fair! Students and staff views on the equity of the procedures and outcomes of students' choice of assessment methods, *Irish Educational Studies*, 36:2, 221-236. DOI: 10.1080/03323315.2017.1324805

Hinton DP, Higson H (2017) A large-scale examination of the effectiveness of anonymous marking in reducing group performance differences in higher education assessment. *PLoS ONE,* 12(8): e0182711. DOI: 10.1371/journal.pone.0182711

Barber, J (2018) *Five go marking an exam question: the use of Adaptive Comparative Judgement to manage subjective bias*

Kacprzyk, J, Parsons, M, Patricia, P, Maguire, and Stewart, G (2019) Examining gender effects in different types of undergraduate science assessment, *Irish Educational Studies*, 38:4, pp 467-480. DOI: 10.1080/03323315.2019.1645721

Winstone, N, Boud, D (2020) The need to disentangle assessment and feedback in higher education, *Studies in Higher Education.*
DOI: 10.1080/03075079.2020.1779687

Pekkarinen, T (2015) Gender differences in behaviour under competitive pressure: Evidence on omission patterns in university entrance examinations. *Journal of Economic Behavior & Organization*, 115, pp 94-110. DOI: 10.1016/j.jebo.2014.08.007

Birch, P, Batten, J, and Batey, J (2016) The influence of student gender on the assessment of undergraduate student work, *Assessment & Evaluation in Higher Education*, 41:7, pp 1065-1080. DOI: 10.1080/02602938.2015.1064857

Surridge, P (2008) *The National Student Survey 2005–2007: Findings and Trends*. Bristol, UK: Higher Education Funding Council for England

Contact us

General enquiries
+44 (0) 3300 416201
enquiries@advance-he.ac.uk
www.advance-he.ac.uk

Media enquiries
+44 (0) 1904 717500
communications@advance-he.ac.uk
www.advance-he.ac.uk/contact-us

@AdvanceHE

Advance HE enables excellence in higher education, helping it shape its future. Within the UK and globally, Advance HE supports institutions in the areas of excellence in education, transformative leadership, equity and inclusion and effective governance. This is delivered through membership benefits (including accreditation of teaching, equality charters, research, knowledge and resources), programmes and events, Fellowships, awards, consultancy and enhancement services and student surveys.

Advance HE is a company limited by guarantee registered in England and Wales no. 04931031. Registered as a charity in England and Wales no. 1101607 Registered as a charity in Scotland no. SC043946. The Advance HE logo should not be used without our permission.

To request copies of this report in large print or in a different format,

please contact the Marketing and Communications Team at Advance HE:

+44 (0) 3300 416201 or publications@advance-he.ac.uk