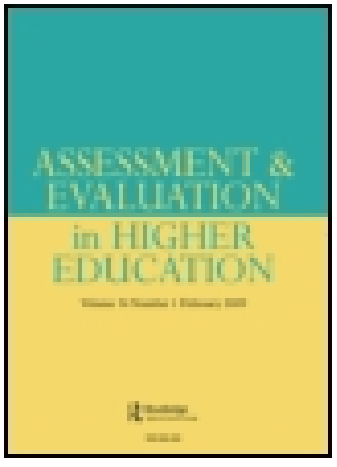

Assessment & Evaluation in Higher Education
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/caeh20

The influence of student gender on the assessment of undergraduate student work

Phil Birch[a], John Batten[b] & Jo Batey[b]
[a] Department of Sport and Exercise Sciences, University of Chichester, Chichester, UK
[b] Department of Sport and Exercise, The University of Winchester, Winchester, UK
Published online: 25 Jul 2015.

To cite this article: Phil Birch, John Batten & Jo Batey (2015): The influence of student gender on the assessment of undergraduate student work, Assessment & Evaluation in Higher Education, DOI: 10.1080/02602938.2015.1064857

To link to this article: http://dx.doi.org/10.1080/02602938.2015.1064857

Assessment & Evaluation in Higher Education, 2015
http://dx.doi.org/10.1080/02602938.2015.1064857

Routledge
Taylor & Francis Group

The influence of student gender on the assessment of undergraduate student work

Phil Birch[a], John Batten[b]* and Jo Batey[b]

[a]*Department of Sport and Exercise Sciences, University of Chichester, Chichester, UK;*
[b]*Department of Sport and Exercise, The University of Winchester, Winchester, UK*

The aim of this study was to investigate the influence of perceived student gender on the feedback given to undergraduate student work. Participants ($n = 12$) were lecturers in higher education and were required to mark two undergraduate student essays. The first student essay that all participants marked was the *control* essay. Participants were informed that the control essay was written by Samuel Jones (a male student). Participants then marked the *target* essay. Although participants marked the same essay, half of the participants ($n = 6$) were informed that the student essay was written by Natasha Brown (a female student), while the remaining participants were informed that it was written by James Smith (a male student). In-text and end-of-text feedback were qualitatively analysed on six dimensions: academic style of writing; criticality; structure, fluency and cohesion; sources used; understanding/knowledge of the subject; and other. Analysis of feedback for both the control and target essay revealed no discernible differences in the number of comments (strengths of the essay, areas for improvement) made and the content and presentation of these comments between the two groups. Pedagogical implications pertaining to the potential impact of anonymous marking on feedback processes are discussed.

Keywords: assessment; student gender bias; feedback

Introduction

Anonymous marking in higher education has been a topic of fractious debate for many years, as reflected by widespread variation in its practice (Brennan 2008; Owen, Stefaniak, and Corrigan 2009; Whitelegg 2002). Yet, since 1999, the National Union of Students (NUS) has campaigned for anonymous marking in the UK, arguing that it provides universities with a remedial method to combat perceived discrimination. Indeed, Wes Streeting, Vice-President (Education) at the NUS stated:

> If we lived in a perfect world, students would be able to put their name on their coursework. Students would not have to fear that their work would be marked any differently based on their gender, sexuality or race. Unfortunately we don't live in that world. (Baty 2007, 1–2)

This suggests the existence of bias in the marking of student work and has resulted in an increasing amount of pressure from across the sector to move towards a uniform method whereby all student work is anonymously assessed (Batten et al. 2013). However, for a number of practical and professional reasons, many

*Corresponding author. Email: john.batten@winchester.ac.uk

academics remain opposed to the widespread implementation of anonymous marking. It is therefore critical that any institutional decisions regarding the use of anonymous marking are both theory and research driven.

One theoretical approach which may be used to explain how biases might manifest themselves in the context of student assessment can be derived from research on expectancy effects. The suggestion here is that marking is a cognitively demanding task and that, for reasons of cognitive efficiency, markers may use certain methods to cope with such high demands. Specifically, in a university setting where a lecturer is marking a student's assignment, schema-driven theorists would argue that a lecturer assigns a student to a specific category based on the cues in the early stages of an interaction (e.g. seeing their name on the cover sheet and assuming knowledge of gender) and then makes a judgement which forms expectancies for the remainder of the interaction (i.e. the marking process). Early proponents of schematic thinking (also termed category activation) considered this process as unavoidable, as illustrated by Allport's statement that 'the human mind must think with the aid of categories … We cannot possibly avoid this process. Orderly living depends on it' (1957, 20).

Support for this viewpoint can also be found in the work of Tversky and Kahneman (1974) and Kahneman and Tversky (1996), who contend that, when individuals are required to make decisions about others in complex situations where processing demand is high, heuristics may be used. Heuristics are thought to be representative tools that serve as shortcuts to provide an individual with simple ways of reasoning to guide decision-making and judgements. Specifically, individuals may achieve this by drawing on prior experiences and knowledge to aid their interpretation of events, as opposed to processing all of the available information. Subsequently, an individual's expectancies may be a central determinant of what information is attended to and how that information is processed within a specific situation. Although there is no direct social interaction between the marker and the student in the assessment of work, it still remains possible that a marker may attempt to fulfil the expectancy they have for a specific student. It is therefore theoretically plausible for the amount and presentation of feedback, as well as the grade, to be influenced by specific cues in the early stages of the interaction.

A wide range of information sources have been found to influence lecturers' judgements of student work, including student ethnicity, socio-economic background, age and physical attractiveness (e.g. Archer and McCarthy 1988; Braun 1976; Meadows and Billington 2005). However, despite the potential impact of student gender bias in assessment and calls for the implementation of anonymous assessment, experimental research in this area is limited. Bradley (1984) examined the difference in marks awarded in final year projects between a student's supervisor and a second independent marker. Bradley hypothesised that student gender bias would occur in the independent marker, whereas the supervisor would be more in touch with the student's true ability. Findings supported this hypothesis, with additional data showing that, when projects were anonymously assessed by the second marker, differences between the two markers were no longer significant. Bradley concluded that blind marking eliminated student gender bias. Spear (1984) found support for this contention in that science scripts were marked higher when they were thought to be written by a male student when compared to a female student. Furthermore, research has identified that female students were awarded higher marks

for composition when compared to their male counterparts (Martin 1972), and higher marks for composition when assessed by a female teacher (Roen 1992).

In contrast, Newstead and Dennis (1990) found no significant differences in the marks awarded by a supervisor and a second independent marker, leading to the conclusion that student gender bias was not present in this instance. Auwarter and Aruguete (2008) identified that gender affected teacher assessments of work only in combination with socioeconomic status (SES), whereby low SES female students were assessed more favourably than high SES female students, and low SES male students were assessed less favourably than high SES male students. Research which has examined the impact of student gender bias in marks awarded has therefore yielded equivocal findings. Although preliminary research has examined the influence of assessor gender on the provision of feedback (Read, Francis, and Robson 2005), to the authors' knowledge, no research has examined the impact of student gender on feedback given.

Research findings have continually supported the contention that effective feedback leads to enhanced student learning and achievement (e.g. Black and Wiliam 1998; Crooks 1988; Hattie, Biggs, and Purdie 1996; Hattie and Jaeger 1998). Feedback plays a key role in learning and development, whereby students learn much faster and more effectively when they are able to identify how they are doing relative to ways in which to improve (Hounsell 2003). According to Nicol and Macfarlane-Dick (2006), good feedback practice is encapsulated by seven principles: helps clarify what good performance is; facilitates the development of self-assessment; delivers high quality information to students about their learning; encourages teacher and peer dialogue around learning; encourages positive motivational beliefs and self-esteem; provides opportunities to close the gap between current and desired performance; and provides information to teachers to inform teaching practice.

Feedback pertaining to student assessment has been emphasised as an increasingly important factor in modern higher education (Bols 2013). This has stemmed from the notion that students have become more like consumers due to the greater financial demands of being in higher education (Cuthbert 2010), which, in turn, has elevated student expectations regarding the perceived quality of feedback received from summative assessments (Laryea 2013). Although National Student Satisfaction scores have increased from 59% in 2005 to 66% in 2010, assessment and feedback generally remain the area of least satisfaction (Bols 2013). In their own survey in August 1999, the NUS reported that 44% of the students unions' believed that discrimination and bias played a part in the way in which students' work was assessed (NUS 2008). In the light of these findings, more research is needed to further understand the factors influencing assessment and feedback to ascertain best practice in higher education assessment.

Although anonymous marking appears to be an appealing method to combat potential assessment bias, there are a number of potential limitations to its implementation. Specifically, anonymous marking is said to decrease the amount of personalised feedback provided throughout the assessment process, which is highly valued by many students (e.g. Bols 2013; Jessop 2007; Laryea 2013). According to Whitelegg (2002), anonymous marking results in a 'disruption to the feedback loop' (7), thus diminishing the necessary lecturer–student relationship for effective feedback provision. Furthermore, weaker students can easily go undetected and be less likely to receive the support they require; thus, anonymous marking could actually

discriminate against those it was designed to protect (Whitelegg 2002), and potentially compromise the student learning experience.

Despite the inherent limitations with anonymous marking, Brennan (2008) posits that it can provide students with the confidence that concerns relating to one assessment can be discussed without fear that the concern will impact future submissions. Furthermore, anonymous marking is thought to shift student expectations so that they actively seek feedback and support after receiving their assessed work (Whitelegg 2002). According to Brennan (2008), anonymous marking 'safeguards' the lecturer and the student by reducing tension and facilitating their relationship, which ultimately promotes greater student learning. Malouff, Emmerton, and Schutte (2013) argue that anonymous marking may protect against halo effects, whereby initial favourable impressions of an individual results in higher subsequent evaluations of assessed work (Nisbett and Wilson 1977).

In line with the benefits of anonymous marking, the NUS argue that 'the script must not include the student's name to remove the possibility of gender bias' (NUS 2008). However, if decisions regarding anonymous assessment were to be informed by substantiated theoretical and research evidence, then it would appear that this call is somewhat premature. Given the ambiguity of the current research literature, it would appear that higher education institutions are not currently in a position to make informed and justified decisions about their marking practices. The distinct lack of empirical evidence which has examined the impact of student gender on feedback provision, coupled with concerns relating to the potential negative impact of anonymous marking on feedback (Whitelegg 2002), suggests the need for further investigation.

In addition, there has been a distinct lack of experimental control in previous research. Most studies (e.g. Bradley 1984; Newstead and Dennis 1990; Spear 1984) have failed to consider the impact of marking stringency amongst the sample and have failed to implement standard marking criteria to guide the assessment of student work. Unless previous research has included a control condition to examine changes in the test variable relative to the norm, it is very difficult to ascertain whether the identified results are attributable to the variable under examination (i.e. student gender) or other confounding variables (i.e. marker stringency). Moreover, although it is possible that each academic will interpret the same marking criteria differently, it is important that the criteria used to assess the quality of student work are consistent across assessments. The aim of this study, therefore, was to examine the influence of perceived student gender on feedback given. Based upon the evidence of previous research, it was predicted that there will be differences in the number of comments made and the content and presentation of these comments between male and female student essays.

Method

Participants

A total of 12 sports academics ($n = 7$ males, $n = 5$ females; M age 28.4, S.D. = 6.8 years) were recruited from six higher education institutions across England. The participant sample (experience in higher education ranging from 6 months to more than 10 years) represented two academic positions ($n = 6$ associate lecturer; $n = 6$ senior lecturer) and reported various marking loads ($n = 5$ less than

50 essays; $n = 2$ 50–100; $n = 3$ 100–200; $n = 2$ 200–500) across the 2011/12 academic year. The protocol was explained to the participants and ethical approval and written consent obtained.

Materials

Student work

The academics were asked to mark the same two first-year undergraduate student essays. The sample essays were approximately 1000 words in length and had previously been submitted for assessment in the Department of Sport and Exercise at the University of Winchester. The original assessors ($n = 2$) confirmed both essays to be of a relatively equal standard (lower second). Consistent with the recommendations of Francis, Robson, and Read (2001, 2002), lower second class (50–59%) essays were used to minimise the potential for gender bias within the assessment process, as they have been found to contain less gender stylistic indicators. Read, Francis, and Robson (2005) also contend that lower second class essays should stimulate more detailed reflections from the prospective markers, as they have both strengths and weaknesses. Written informed consent was obtained from the students to use their original assignments for this research.

Measures

Assessment criteria profiles

The pedagogical research literature (e.g. Francis et al. 2003; Read, Francis, and Robson 2005) has found that academics have a tendency to use different criteria (typically from their own institutions) to help them assess the quality of student work. Furthermore, research has suggested a number of 'models of marking' may be in use (Yorke, Bridges, and Woolf 2000). First, the platonic model proposes that a marker has a predetermined idea of appropriate content, style and presentation to assess work against. Second, the intuitive model proposes that a marker uses experience of being assessed when he/she was a student to assess work. The final model proposes that markers use predetermined marking criteria to assess student work. With this in mind, participants in this study were asked to use the marking scheme (Assessment Criteria Profile; ACP) from the Department of Sport and Exercise at the University of Winchester to help standardise the experimental protocol. The ACP (tick sheet) was derived from academic discourse with external examiners.

Procedure

Participants were asked to carefully read a set of directions prior to marking the essay. These directions were portrayed to the participants in such a way as to emphasise the importance of this information in helping to contextualise the assignment. However, this information also identified the essay as being written by a particular student of a specific gender. The first student essay that all participants marked was the *control* essay. All participants were informed that the control essay was written by Samuel Jones and was submitted for assessment on a first-year Introduction to Research and Study Skills module at the University of Winchester.

Participants then marked the *target* essay. All participants marked the same essay, but the essay was identified as being written by either a female or male student. Participants were randomly and equally ($n = 6$) divided into the two experimental groups. Those that were asked to mark the female student essay were informed that this essay was written by Natasha Brown, and those that were asked to mark the male student essay were informed that this essay was written by James Smith. All participants were informed that the essay was submitted for assessment on a first-year Introduction to Research and Study Skills module at the University of Winchester. The identified names were matched for perceived age and attractiveness and deemed by the International Student Officer and Widening Participation Officer at the University of Winchester to be typical of individuals from the said genders. Participants were required to mark the essays in line with current practice. However, participants were also informed that they should utilise the ACP to help them to assess the quality of the work. This study entailed the exploration of the in-text and end-of-text feedback provided for each essay. Participants were debriefed on the purpose of the investigation following the completion of marking both essays.

Data analysis

The feedback provided for each participant was qualitatively analysed using the template developed by Batten et al. (2013). This template was devised following thematic analysis and was used to facilitate comparisons between the experimental conditions. Both the in-text and end-of-text comments of the control and target essays were assessed on the six dimensions outlined in the marking criteria: academic style of writing; criticality; structure, fluency and cohesion; sources used; understanding/knowledge of the subject; and other. The total number of comments pertaining to the 'strengths of the essay' and 'areas of improvement' was analysed for each of the respective dimensions. Example comments were also provided to substantiate each dimension.

In order to examine between group differences in feedback given, two groups were assigned. Participants who were told that the target essay was written by a female student were assigned to group one, while participants who were told the target essay was written by a male student were assigned to group two. First, visual inspection between group one and group two for the control essay was conducted to ascertain whether differences in feedback were present when student gender was the same. Second, visual inspection between group one and group two of the target essay was conducted to examine whether differences were present in feedback given across female and male students. Findings that show no differences between group one and group two in the control essay, yet show differences between the groups in the target essay, will provide confidence that differences in the target essay were attributable to student gender and not simply differences in marking practice.

Results

Control essay

Upon analysing the feedback given to the control essay, the findings generally indicated very little difference in the number of comments made about the strengths of the essay and areas for improvement between participants in group one and group

Table 1. Qualitative analysis of the control student essay across group one and group two.

		Total number of comments		Example comments	
Theme	Group	Strengths of the essay	Areas of improvement	Strengths of the essay	Areas of improvement
Academic style of writing	1	3	56	Well written Good first sentence	Use the past tense Try and avoid asking questions Writing style is too colloquial Grammar This long sentence needs attention
	2	1	79	The student writes fairly well	Poor writing style Avoid rhetorical questions Avoid colloquial language Check grammar Long sentence – be short and concise
Criticality	1	4	12	Balanced argument Good foundation to build critical knowledge	Need more critiquing Try to develop more of a critical standpoint
	2	1	8	You make some good points for both sides of the argument	Lack of extension of arguments Better critical analysis of arguments needed
Structure, fluency and cohesion	1	0	16	No comments made	Maybe put this earlier Not the best way to open an essay Focus on developing a clear structure
	2	3	24	Some form of structure Good introduction	Lack of flow Contradiction to previous argument Better link needed

(*Continued*)

Table 1. (*Continued*).

Theme	Group	Total number of comments		Example comments	
		Strengths of the essay	Areas of improvement	Strengths of the essay	Areas of improvement
Sources used	1	5	34	Good reference Good list of references Effective use of references	Need a reference to support this Alphabetised references Some of these references are not in your text
	2	2	48	Good referencing style Good attempt at a proper referencing style	Haven't provided sufficient evidence base Alphabetical order All references should be cited in the text
Understanding/knowledge of the subject	1	1	17	Demonstrate understanding of their argument	You often make unsubstantiated bold statements Not sure this is relevant or necessarily true
	2	3	10	You have raised some relevant points Good point	Limited academic knowledge Wider reading needed
Other	1	0	1	No comments made	Page numbers
	2	2	1	Some good points made The essay addresses the main points	A generally weak essay

Note: Group 1 = Female student essay; Group 2 = Male student essay.

two (see Table 1). Findings also indicate that there was no discernible differences in the content of the feedback and how the feedback was presented (e.g. how emotive and animated it appeared) between the two groups. The following quotations from the 'criticality theme' (areas for improvement) demonstrate that feedback was not different between the two groups assessed. For example, one participant from group one commented 'need more critiquing' and a participant from group two commented 'better critical analysis of arguments needed'. Additionally, quotations from the 'academic style theme' (strengths of the essay) supported this finding, in that one participant from group one commented 'well written' and one participant from group two commented 'the student writes fairly well'. Exemplar comments for each dimension across the two groups are presented in Table 1.

Target essay

Findings for the target essay revealed that, in general, the number of comments made and the content and presentation of the feedback given by participants were not specific to the perceived gender of the student (see Table 2). Quotations from the 'structure, fluency and cohesion' theme (strengths of the essay) supported this finding, in that one participant from group one commented 'good structure', and a participant from group two commented 'well structured'. Furthermore, comments from the 'understanding/knowledge of the subject' theme (areas for improvement) provided support for this finding, in that one participant from group one commented 'weak application of literature', and one participant from group two commented 'provide more scientific information about the effects of drugs'. Collectively, these quotations indicate that student gender did not appear to influence the content and presentation of feedback given.

The findings also show that, in general, there were more comments made relating to areas of improvement than strengths of the essay for both the control and target essays. Although some differences were present regarding the focus of the feedback for both essays, with more comments being targeted at 'academic style of writing' and 'sources used', the focus of feedback was consistent between group one and group two.

Discussion

The purpose of this study was to investigate the influence of perceived student gender on the feedback given to undergraduate student work. In contrast to the prediction that there will be differences in the number of comments made and the content and presentation of these comments between male and female student essays, the findings of this study suggest that student gender has little or no influence on feedback given. The findings pertaining to the control essay provided evidence to suggest that, when lecturers assessed an essay thought to be written by the same student, the feedback given was similar regarding the number of comments made and the content and presentation of those comments. This finding provided the necessary prerequisite (manipulation check) to assess the influence of student gender on feedback provided. Similarly, the findings of the target essay indicated that feedback given was not dependent upon the perceived gender of the student. Collectively, the findings suggest that no gender bias in feedback given was observed in this study.

Table 2. Qualitative analysis of the target student essay across group one and group two.

		Total number of comments		Example comments	
Theme	Group	Strengths of the essay	Areas of improvement	Strengths of the essay	Areas of improvement
Academic style of writing	1	4	64	Well written Very well written essay Read very well with good grammar	Do not use 1st person Grammar Try not to make sentences too long
	2	6	43	Good writing style Well written	Avoid first person Try to be more concise with sentences
Criticality	1	8	4	Good effort with evidence of arguing the issue from a balanced perspective Provided a critical/logical exposition of the drugs used in sport	You summarised the two viewpoints but then failed to reach a conclusion Try and develop a critical stance in relation to the literature
	2	9	6	Good counter argument Good example and critique	In-depth analysis needed You could have selected stronger arguments
Structure, fluency and cohesion	1	12	2	Good structure Reasonable introduction Good organisation and coherence Good introduction explaining your plan	Should start with more neutral argument Introduction could be more substantial
	2	14	12	Well structured Succinct introduction Coherent essay	Flow to the essay needed Slight tangent here This does not appear in your essay
Sources used	1	4	30	You provide evidence to support your arguments Wide range of literature highlighted	References needed Format of citation requires attention More quality sources needed No need for bullet points
	2	10	56	Good use of references to support argument Use of a range of relevant sources	References needed Some issues with referencing style Use more research not just examples Do not use bullet points
Understanding/knowledge of the subject	1	0	7	No comments made	Weak application of literature Provide definition

					Theory needed
	2	5	21	Good points raised Well defined	Provide more scientific information about the effects of drugs Use more theory and research Could explain more
Other	1	6	2	Good Overall well done	Page numbers
	2	5	0	Good attempt Generally a good essay	No comments made

Note: Group 1 = Female student essay; Group 2 = Male student essay.

The findings therefore appear to provide little support for the theoretical contentions of Tversky and Kahneman (1974) and Kahneman and Tversky (1996), whereby expectancies are argued to bias how people process and interpret information. From a theoretical standpoint, the findings suggest that participants in this study appear to have used the content of the essay to inform the provision of feedback, as opposed to being biased in their judgements by student gender. However, the information (i.e. student gender) presented to the participants in this study was based on artificial information, which could have influenced the extent to which expectancies were formed. According to White, Jones, and Sherman (1998), expectancies can be formed by indirect and direct information sources, whereby expectancies derived from indirect information are determined by the degree of credibility the perceiver places on such information. Given that no gender bias in feedback was observed in this study, it could be argued that participants did not deem the information received regarding student gender as a salient cue to inform their judgements. Consequently, future research may need to consider the extent to which lecturers use such information to inform the provision of feedback.

Such findings appear to lend support to researchers who believe individuals can either control category activation (Bargh 1994; Blair 2002; van Ryn and Fu 2003; Wegner and Bargh 1998), or once a category is activated can engage in category inhibition to suppress the application of their expectancy (e.g. Bodenhausen 2005). Alternatively, category activation might only occur under specific triggering conditions. Indeed, Blair (2002) found that levels of automatic category activation were dependent upon (a) self and social motives, (b) specific strategies (e.g. stereotype suppression), (c) the perceivers focus of attention (e.g. attentional load), and (d) the configuration of stimulus cues (e.g. the context within which cues are received). Blair (2002) reported that individual characteristics influenced the extent of category activation and stereotyping. Therefore, future research is needed to examine the specific triggering conditions which may lead to the occurrence of expectancy effects in student assessment.

The findings of this study are inconsistent with research which has found student gender bias in marks awarded to undergraduate work (e.g. Bradley 1984; Spear 1984). However, the findings are consistent with previous research (e.g. Newstead and Dennis 1990) where no differences were found in the marks awarded across student gender. Given the lack of experimental control adopted in previous research, this study not only extends understanding of student gender bias in assessment, but also provides a promising methodological platform for this line of research to develop.

However, one possible limitation of this study is that participants were aware that they were assessing student work as part of the study objectives. This may have led participants to present themselves in a desirable way, and potentially alter their typical assessment and feedback practices. Research examining self-presentational effects has consistently supported the contention that individuals are more likely to change their behaviour when placed in an evaluative environment (e.g. Fernald et al. 2012). For example, the participants in this study may have felt the need to artificially increase the amount of feedback given in order to reduce the likelihood of being evaluated as a lecturer who provides insufficient feedback. Holmes and Papageorgiou (2009) have shown that students consider the quantity of feedback when making an overall assessment of the quality of feedback they receive. Although this methodological limitation is inherent with examining bias in

experimental research concerning assessment and feedback, researchers are encouraged to consider the impact of such effects on the conclusions drawn from research projects.

While there are a number of pedagogical implications of this study, the most pronounced impact could be directed at the use of anonymous marking procedures in higher education. The findings of this study indicate the absence of student gender bias in the provision of feedback. Given that the primary role of anonymous marking is to eliminate the possibility of assessment bias and subsequent student discrimination, the preliminary findings of this study tentatively suggest that anonymous marking may not be needed in higher education. Institutions which consider moving away from anonymous marking may reduce the likelihood of disrupting the feedback loop (Whitelegg 2002) and reap the rewards of more personalised feedback. Nicol and Macfarlane-Dick's (2006) seven key principles for effective feedback emphasise the importance of personalised feedback, which is thought to 'encourage teacher and peer dialogue around learning'.

In order for external feedback from lecturers to be effective, it must be understood and internalised by the student to facilitate any future improvements (Nicol and Macfarlane-Dick 2006). Research (e.g. Chanock 2000; Hyland 2000) has shown that students who do not understand feedback are less likely to be able to take the necessary action to reduce the gap between their goals and their desired outcomes for current and future assessments. Laurillard (2002) further argues that feedback dialogue must afford the opportunity for the student to engage the lecturer in follow-up discussions regarding assessment feedback. Given the importance placed on feedback facilitating student learning (e.g. Black and Wiliam 1998; Crooks 1988; Hattie, Biggs, and Purdie 1996; Hattie and Jaeger 1998) and the value placed on personalised feedback by students (e.g. Bols 2013; Jessop 2007; Laryea 2013), it could be argued that the impersonal nature of anonymous feedback may be comprising the mechanisms outlined by Nicol and Macfarlane-Dick (2006) to underpin effective feedback.

Compromising the means by which feedback operates may not only influence student achievement, it may also result in negative motivational beliefs and perceptions of ability (Nicol and Macfarlane-Dick 2006). For instance, research by Harlen and Crick (2003) has shown that student assessments which only offer grades, as opposed to accompanying feedback, resulted in negative motivational beliefs for learning that mitigates against lifelong learning attitudes. Research by Butler (1988) showed that, when students are given feedback only, they were more motivated to learn when compared to students who were given marks only. Furthermore, feedback given in the form of grades has been shown to have negative influences on self-esteem, with low ability students most at risk (Craven, Marsh, and Debus 1991). The impact of feedback (especially personalised) on student beliefs is therefore a poignant factor for higher education institutions to consider in their quest to develop motivated and knowledgeable graduates.

Although this study does not call for a complete shift in marking practices in higher education, the preliminary evidence gleaned does suggest that institutions should strongly consider the potential negative impacts of anonymous marking on feedback and the student learning experience. Further research is therefore warranted to examine the potential impact of anonymous marking upon feedback given and subsequent student cognition (e.g. motivation) and behaviour (achievement). Research also needs to move beyond merely testing for the existence of bias in

assessment and towards an examination of the processes that underpin assessment bias. Such an understanding will help higher education institutions to generate the means to counteract and prevent discrimination in situations where anonymous marking is not practically feasible (i.e. student presentations).

Acknowledgements

The authors would like to acknowledge the participation of those lecturers who took the time to assess student essays in this study.

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes on contributors

Phil Birch is a senior lecturer in the Department of Sport and Exercise Sciences at the University of Chichester. His present research interests have examined measurement issues in mental toughness.

John Batten is a senior lecturer in sport and exercise psychology at the University of Winchester. His present research interests have examined expectancy effects in education, health and sport.

Jo Batey is a senior lecturer in the Department of Sport and Exercise at the University of Winchester. She has published in a diverse number of areas including applied sport psychology, the experiences of women in sport and career-ending injury.

References

Allport, G. W. 1957. "European and American Theories of Personality." In *Perspectives in Personality Theories*, edited by H. P. David and H. von Bracken, 3–26. New York: Basic Books.

Archer, J., and B. McCarthy. 1988. "Personal Biases in Student Assessment." *Educational Research* 30: 142–145.

Auwarter, A. E., and M. A. Aruguete. 2008. "Effects of Student Gender and Socioeconomic Status on Teacher Perceptions." *Journal of Educational Research* 4: 243–246.

Bargh, J. A. 1994. "The Four Horsemen of Automaticity: Awareness, Efficiency, Intention, and Control in Social Cognition." In *Handbook of Social Cognition*. 2nd ed, edited by R. S. Wyer Jr. and T. K. Srull, 1–40. Hillsdale, NJ: Erlbaum.

Batten, J., J. Batey, L. Shafe, L. Gubby, and P. D. J. Birch. 2013. "The Influence of Reputation Information on the Assessment of Undergraduate Student Work." *Assessment & Evaluation in Higher Education* 4: 417–435.

Baty, P. 2007, May 25. *Trust Eroded by Blind Marking*. http://www.timeshighereducation.co.uk/209075.article.

Black, P., and D. Wiliam. 1998. "Assessment and Classroom Learning." *Assessment in Education: Principles, Policy & Practice* 5: 7–74.

Blair, I. V. 2002. "The Malleability of Automatic Stereotypes and Prejudice." *Personality and Social Psychology Review* 6: 242–261.

Bodenhausen, G. V. 2005. "The Role of Stereotypes in Decision-making Processes." *Medical Decision Making* 25: 112–118.

Bols, A. 2013. "Student Views on Assessment." In *Improving Student Engagement and Development through Assessment: Theory and Practice in Higher Education*, edited by L. Clouder, C. Broughan, S. Jewell, and G. Steventon, 4–18. London: Routledge.

Bradley, C. 1984. "Sex Bias in the Evaluation of Students." *British Journal of Social Psychology* 23: 147–153.
Braun, C. 1976. "Teacher Expectation: Sociopsychological Dynamics." *Review of Educational Research* 46: 185–213.
Brennan, D. J. 2008. "University Student Anonymity in the Summative Assessment of Written Work." *Higher Education Research and Development* 27: 43–54.
Butler, R. 1988. "Enhancing and Undermining Intrinsic Motivation: The Effects of Task-involving and Ego-involving Evaluation on Interest and Performance." *British Journal of Educational Psychology* 58: 1–14.
Chanock, K. 2000. "Comments on Essays: Do Students Understand What Tutors Write?" *Teaching in Higher Education* 5: 95–105.
Craven, R. G., H. W. Marsh, and R. L. Debus. 1991. "Effects of Internally Focused Feedback and Attributional Feedback on Enhancement of Academic Self-concept." *Journal of Educational Psychology* 83: 17–27.
Crooks, T. J. 1988. "The Impact of Classroom Evaluation Practices on Students." *Review of Educational Research* 58: 438–481.
Cuthbert, R. 2010. "Students as Customers?" *Higher Education Review* 3: 3–25.
Fernald, D. H., L. Coombs, L. DeAlleaume, D. West, and B. Parnes. 2012. "An Assessment of the Hawthorne Effect in Practice-based Research." *The Journal of the American Board of Family Medicine* 25: 83–86.
Francis, B., J. Robson, and B. Read. 2001. "An Analysis of Undergraduate Writing Styles in the Context of Gender and Achievement." *Studies in Higher Education* 26: 313–326.
Francis, B., J. Robson, and B. Read. 2002. "Gendered Patterns in Writing and Degree Award." In *Gender, Teaching and Research in Higher Education: Challenges for the Twenty-first Century*, edited by G. Howie and A. Tauchert, 167–180. Aldershot: Ashgate Press.
Francis, B., B. Read, L. Melling, and J. Robson. 2003. "University Lecturers' Perceptions of Gender and Undergraduate Writing." *British Journal of Sociology of Education* 24: 357–373.
Harlen, W., and R. D. Crick. 2003. "Testing and Motivation for Learning." *Assessment in Education: Principles, Policy & Practice* 10: 169–207.
Hattie, J., J. Biggs, and N. Purdie. 1996. "Effects of Learning Skills Interventions on Student Learning: A Meta-Analysis." *Review of Educational Research* 66: 99–136.
Hattie, J., and Jaeger, R. 1998. "Assessment and Classroom Learning: A Deductive Approach." *Assessment in Education: Principles, Policy & Practice* 5: 111–122.
Holmes, K., and G. Papageorgiou. 2009. "Good, Bad and Insufficient: Student's Expectations, Perceptions and Uses of Feedback." *Journal of Hospitality, Leisure, Sport & Tourism Education* 1: 85–96.
Hounsell, D. 2003. "Student Feedback, Learning and Development." In *Higher Education and the Lifecourse*, edited by M. Slowey and D. Watson, 67–78. Maidenhead: Open University Press.
Hyland, P. 2000. Learning from Feedback on Assessment. In *The Practice of University History Teaching*, edited by A. Booth and P. Hyland, 233–247. Manchester: Manchester University Press.
Jessop, T. 2007. *Student Perceptions of Assessment and Feedback*. Unpublished Report to the University Learning and Teaching Committee, Winchester: The University of Winchester.
Kahneman, D., and A. Tversky. 1996. "On the Reality of Cognitive Illusions." *Psychological Review* 103: 582–591.
Laryea, S. 2013. "Feedback Provision and Use in Teaching and Learning: A Case Study." *Education and Training* 7: 665–680.
Laurillard, D. 2002. *Rethinking University Teaching*. 2nd ed. London: Routledge Falmer.
Malouff, J. M., A. J. Emmerton, and N. S. Schutte. 2013. "The Risk of a Halo Bias as a Reason to Keep Students Anonymous During Grading." *Teaching of Psychology* 40: 233–237.
Martin, W. D. 1972. "The Sex Factor in Grading Composition." *Research in Teaching of English* 6: 36–47.

Meadows, M., and L. Billington. 2005. "A Review of the Literature on Marking Reliability." Paper Commissioned by the National Assessment Agency. Accessed May 19, 2015. https://cerp.aqa.org.uk/sites/default/files/pdf_upload/CERP_RP_MM_01052005.pdf

Newstead, S. E., and I. Dennis. 1990. "Blind Marking and Sex Bias in Student Assessment." *Assessment & Evaluation in Higher Education* 2: 132–139.

Nicol, D., and D. Macfarlane-Dick. 2006. "Formative Assessment and Self-regulated Learning: A Model and Seven Principles of Good Feedback Practice." *Studies in Higher Education* 31: 199–218.

Nisbett, R. E., and T. D. Wilson. 1977. "The Halo Effect: Evidence for Unconscious Alteration of Judgments." *Journal of Personality and Social Psychology* 35: 250–256.

NUS. 2008. "Higher Education Campaign: Mark my Words, Not My Name" [online]. Accessed February 23, 2012. http://www.nus.org.uk/en/Campaigns/Higher-Education/Mark-my-words-not-my-name/

Owen, C., J. Stefaniak, and G. Corrigan. 2009. "Marking Identifiable Scripts: Following Up Student Concerns." *Assessment & Evaluation in Higher Education* 1: 33–40.

Read, B., B. Francis, and J. Robson. 2005. "Gender, 'Bias', Assessment and Feedback: Analysing the Written Assessment of Undergraduate History Essays." *Assessment & Evaluation in Higher Education* 3: 241–260.

Roen, D. 1992. "Gender and Teacher Response to Student Writing." In *Gender Issues in the Teaching of English*, edited by N. McCracken and B. Appleby, 126–141. Portsmouth, NH: Heinemann.

van Ryn, M., and S. S. Fu. 2003. "Paved With Good Intentions: Do Public Health and Human Service Providers Contribute to Racial/Ethnic Disparities in Health?" *American Journal of Public Health* 93: 248–255.

Spear, M. 1984. "Sex Bias in Science Teachers' Ratings of Work and Pupil Characteristics." *European Journal of Science Education* 6: 369–377.

Tversky, A., and D. Kahneman. 1974. "Judgment under Uncertainty: Heuristics and Biases." *Science* 185: 1124–1131.

Wegner, D. M., and J. A. Bargh. 1998. "Control and Automaticity in Social Life." In *Handbook of Social Psychology*. 4th ed, edited by D. T. Gilbert, S. T. Fiske, and G. Lindzey, 446–496. Boston, MA: McGraw-Hill.

White, K. J., K. Jones, and M. D. Sherman. 1998. "Reputation Information and Teacher Feedback: Their Influences on Children's Perceptions of Behaviour Problem Peers." *Journal of Social & Clinical Psychology* 17: 11–37.

Whitelegg, D. 2002. "Breaking the Feedback Loop: Problems with Anonymous Assessment." *Planet* 3: 7–8.

Yorke, M., P. Bridges, and H. Woolf. 2000. "Mark Distributions and Marking Practices in UK Higher Education: Some Challenging Issues." *Active Learning in Higher Education* 1: 7–27.