

## Gender stereotyping in student perceptions of teaching excellence: applying the shifting standards theory

Kathryna Kwok & Jacqueline Potter

To cite this article: Kathryna Kwok & Jacqueline Potter (2022) Gender stereotyping in student perceptions of teaching excellence: applying the shifting standards theory, Higher Education Research & Development, 41:7, 2201-2214, DOI: [10.1080/07294360.2021.2014411](https://doi.org/10.1080/07294360.2021.2014411)

To link to this article: <https://doi.org/10.1080/07294360.2021.2014411>



© 2021 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 14 Dec 2021.



[Submit your article to this journal](#)



Article views: 1925





[View related articles](#)



[View Crossmark data](#)

## Gender stereotyping in student perceptions of teaching excellence: applying the shifting standards theory

Kathryna Kwok  and Jacqueline Potter 

Oxford Centre for Staff and Learning Development, Oxford Brookes University, Oxford, UK

### ABSTRACT

Existing characterisations of student conceptions of teaching excellence (TE) implicitly position it as an objective construct. This study investigated gendered differences in student-submitted nominations ( $n = 418$ ) for an excellence award in a mid-sized university in England. Biernat's shifting standards theory, which proposes that evaluative standards can shift due to stereotyping effects, was used to interpret the findings. Chi-square tests revealed significant effects of gender on the distribution and thematic content of nominations. Results suggested that students were more likely to nominate teachers of the same gender, but also that male students were disproportionately less likely to nominate a female teacher. Student conceptions of TE generally conformed to gender biases, particularly for male students. These findings indicate that students' perceptions of high quality teaching are inextricable from sociocultural influence. Future research can continue to engage with the shifting standards theory to investigate the influence of gender and student perceptions of high quality teaching.

### ARTICLE HISTORY

Received 8 June 2021  
Accepted 14 October 2021

### KEYWORDS

Teaching excellence; gender; student evaluations of teaching; shifting standards theory

## Introduction

Teaching excellence (TE) gained prominence as a focus of academic discussion following the introduction of the Teaching Excellence Framework in UK higher education (HE). Critics of TE note its role in neoliberal agendas (Bartram et al., 2019) and suggest that it is so vague as to be 'meaningless' (Wood & Su, 2017, p. 457). On the other hand, there has been a stream of practical work focused on identifying what students regard as TE (e.g., Bradley et al., 2015; Lubicz-Nawrocka & Bunting, 2019; Su & Wood, 2012). This study examined the distribution and content of student-submitted nominations for an excellence award at a mid-sized English university, in order to explore whether there was a relationship between gender and (a) a student's likelihood to nominate a teacher, and (b) what students considered to be excellent about a nominated teacher. The explanatory power of the shifting standards theory (Biernat, 2003)—which posits that evaluative standards can vary due to stereotyping effects—is also

**CONTACT** Kathryna Kwok  [kkwok@brookes.ac.uk](mailto:kkwok@brookes.ac.uk)  Oxford Centre for Staff and Learning Development, Wheatley Campus, Wheatley, Oxford, OX33 1HX, UK

© 2021 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group  
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

tested, in contrast to previous work which has tended to make reference only to the theory's broadest argument (e.g., Sprague & Massoni, 2005).

## Literature review

### *TE as conceptualised in the literature*

Practical TE research has typically attempted to characterise TE by identifying examples of high quality teaching practice, for instance by directly eliciting student responses (Derounian, 2017; Revell & Wainwright, 2009; Shaha et al., 2013), or by examining evaluative data, such as nominations for teaching awards, to make inductive conclusions about TE (Bradley et al., 2015; Lubicz-Nawrocka & Bunting, 2019; Moore & Kuol, 2007; Su & Wood, 2012). Many recurring features can be identified in the characterisations of TE offered in these studies. First, excellent teaching is engaging: students value teachers who present content in an interesting and understandable way (Revell & Wainwright, 2009). Excellent teaching is also inspiring and motivating (Bradley et al., 2015). Sometimes, this stems from a teacher's passion towards their field (Derounian, 2017). Excellent teachers are friendly, charismatic and have a sense of humour (Su & Wood, 2012). They are also approachable and available: just as students feel able to seek support from them, so are they prepared to make time to help (Lubicz-Nawrocka & Bunting, 2019). Lastly, excellent teachers are professionally competent: they are experts in their field and organised in their work (Shaha et al., 2013).

On the surface, these characterisations offer valuable practical insight into what students recognise as TE. Critically however, these characterisations do not account for the inherent subjectivity in how students perceive teaching (Kulik, 2001). For instance, Lubicz-Nawrocka and Bunting (2019, p. 64) assert that 'different forms of teaching excellence can coexist and benefit different students', implying that there are set, objective ways that TE can exist. The assumption of objectivity overlooks that criteria for excellence can vary between and within individuals, as will be explained later. Interestingly, while TE is discursively framed as objective, presuppositions about the identified TE characteristics suggest a recognition that TE is necessarily subjective: nowhere in the abovementioned studies is it suggested that every identified characteristic of TE must be present in order for teaching to be considered excellent, nor indeed that each characteristic is equally important. The ability to teach engagingly, for instance, is presented alongside the quality of being humorous (e.g., Bradley et al., 2015) though the former is arguably a more essential criterion for TE than the latter. As evaluative processes are relational and inextricable from sociocultural contexts (O'Connor et al., 2020), treating TE as objective ignores the critical role of sociocultural biases and expectations. The following section outlines evidence that gender can influence the practices and traits that HE students notice and value in their teachers.

### *How does gender influence student evaluations of teaching?*

Research into gender and student evaluations of HE teaching have found that compared to male teachers, female teachers are rated more lowly or are otherwise not equally considered on measures relating to intelligence and scholarship (Basow, 1995; Basow et al.,



2006; Boring, 2017; Miller & Chamberlin, 2000; Mitchell & Martin, 2018), professionalism (MacNeill et al., 2015) and overall teaching quality (Basow, 1995; Basow et al., 2006; Fan et al., 2019; Mengel et al., 2019). A systematic analysis by Heffernan (2021) of 30 years of research into student evaluations of teaching, which spanned multiple countries and types of HE institutions, concludes that ‘at best [student evaluations of teaching] disadvantage women, and at worst, see women academics placed in untenable positions’ (p. 8). Overall, these findings point to the ubiquity of gender inequality in how teaching is perceived by students, mirroring the historic patterns of women’s marginalisation in HE (Delamont, 2006) and in wider society (Laube et al., 2007).

Basow et al.’s (2006) examination of students’ descriptions of their best and worst HE instructor provides an illustration of the qualitatively different ways that male and female HE teachers are evaluated. They found that best female instructors were more likely than best male instructors to be described as having high interpersonal skill, particularly by male students. Best male instructors were more likely than best female instructors to be described as knowledgeable by students of both genders. Additionally, male students’ evaluations appeared to be influenced by the power dynamic between them and their teachers: male students were likelier than expected to characterise their best female instructors as accommodating and their worst female instructors as ‘inflexible’ (32), and described their worst instructors (particularly if they were also male) as ‘condescending’ (30) more often than expected.

It should be noted that rather than impacting men and women in different but equal ways, it is women who are much more often disadvantaged by gender stereotypes. Within HE, the roles associated with women are undervalued relative to those associated with men. For instance, research is valued over teaching expertise (O’Connor & O’Hagan, 2016), while administrative and professional services roles (disproportionately occupied by women) are marginalised relative to academic roles (Holmes, 2020). The latter pattern can also be seen amongst academics, as female academics often undertake less visible ‘academic housekeeping’ tasks (Burford et al., 2020; Heijstra et al., 2017), leaving less time to engage in more prestigious academic activity (Kandiko Howson et al., 2018). Considering the sociocultural salience of gender, that it has not yet been considered in practical investigations of TE presents an important limitation which this study hopes to address.

### *The shifting standards theory*

One theory that has been used to account for gendered differences in student perceptions of teaching is Biernat’s (e.g., 2003) shifting standards theory (SST). The SST proposes that individuals can be evaluated against different standards depending on the stereotypes associated with the group an individual belongs to. For instance, it would be more difficult for an individual to demonstrate a particular attribute if they were stereotyped as deficient in that attribute. Conversely, this individual would find it easier to demonstrate an attribute for which they were stereotyped to have. While the SST makes specific predictions about how evaluative standards shift between evaluative contexts, research into gender and student evaluations of teaching have tended not to engage with these predictions (e.g., Basow et al., 2006; Sprague & Massoni, 2005); instead, the SST is used in broad support of the argument that evaluative standards can change based on preconceived assumptions about the target(s) of evaluation. To better

understand how and why gender can influence student perceptions of TE, the discussion now turns to what these predictions are, framed as three key distinctions made by the SST.

The first distinction is between common-rule and subjective measures. Common-rule measures require raters to refer to some common standard or scale when making evaluations (e.g., in estimations of highest educational attainment). By contrast, subjective measures are not consistent across contexts, enabling different standards to be used depending on the target of evaluation (e.g., in open-ended descriptions) (Biernat, 2003). Evidence for the SST has shown that judgments elicited with common-rule measures tend to show assimilative effects, meaning they are consistent with stereotyped views. By contrast, judgments elicited using subjective measures can mask stereotyped views, either by producing a null effect where judgments appear equal across groups, or a contrastive effect where judgments are the opposite to the relevant stereotype(s) (Biernat & Eidelman, 2007; Biernat & Vescio, 2002). For example, a student may describe both a male and female HE teacher as experts in their field (a subjective measure) but may nevertheless estimate the female teacher's highest educational attainment (a common-rule measure) to be lower than the male teacher's. Indeed the latter was a finding by Miller and Chamberlin (2000), who noted that HE students tended to underestimate the highest degree attained of female faculty staff but overestimated the same for male staff.

The second distinction pertains to minimum and confirmatory standards. Minimum standards are the threshold at which there is suspicion that a target possesses some characteristic (e.g., being humorous), while confirmatory standards refer to the point at which there is certainty of this (Biernat, 2003). Evidence has shown that when evaluating a group stereotyped as deficient in a characteristic  $x$ , raters apply lower minimum standards but higher confirmatory standards than if there had been no stereotype (Biernat et al., 2010). In other words, someone stereotyped as deficient in  $x$  would find it more difficult to convince raters that they indeed possessed  $x$  (due to higher confirmatory standards), even though relatively little evidence would be needed before a rater begins to suspect that they might have  $x$  (due to low minimum standards). By contrast, for a group stereotyped as proficient in  $x$ , minimum standards are higher and are roughly equivalent to confirmatory standards. Hence, someone stereotyped as proficient in  $x$  would need to demonstrate more evidence before they are suspected to have  $x$  (due to higher minimum standards), though once this initial threshold is reached, raters would also be confident that this person indeed possessed  $x$  (as minimum standards are roughly equivalent to confirmatory standards) (ibid.). Certainly, there is evidence that student perceptions of HE teachers can be influenced by gendered stereotypes: Basow et al. (2006) found that male teachers tend to be noted for male-typed cognitive competencies such as being knowledgeable, whereas female teachers tend to be noted for female-typed interpersonal competencies, such as being caring or approachable.

The last distinction is between zero-sum and non-zero-sum actions. Zero-sum actions occur when resources are finite such that decisions necessarily come at the expense of other options (e.g., selecting someone for a promotion). Non-zero-sum actions, by contrast, can be performed repeatedly at no real cost (e.g., praising employees' work) (Biernat, 2003). Evidence for the SST has shown that zero-sum actions tend to be assimilative to stereotypes, while non-zero-sum actions show null or contrastive effects



(Biernat & Vescio, 2002). Critically, zero-sum and non-zero-sum may be best imagined as points on opposite ends of a continuum, with resources being severely restricted on the zero-sum end, and resources being theoretically unlimited on the non-zero-sum end. In Marchant and Wallace's (2016) analysis of award recipients of an Australian national HE teaching awards scheme, it was found that women were overrepresented in the more numerous lower-level citations (up to 100 awarded), whereas men were overrepresented in more prestigious—and far less numerous—awards (up to eight awarded; see Universities Australia, 2021). This is consistent with the SST's predictions of zero-sum and non-zero-sum actions, as interpreted along a continuum. Marchant and Wallace's (2016) findings illustrate a further point: even in the stereotypically female domain of teaching, women are less readily perceived to be performing to a high standard. As the authors note (citing van den Brink and Benschop [2012]), excellence itself is not gender-neutral.

## Method

This study analysed student-submitted nominations ( $n = 418$ ) for an excellence award at a mid-sized, public research university in England. The university offers a broad-based curriculum across three faculties and gained university status in the 1960s. Nominations for the 2015/16 and 2016/17 rounds of the award were included in the dataset. Nominations were collected via an online form which asked for the names of the nominator (student) and the nominee (teacher), the module/programme on which the nomination was based, and the reason(s) for the nomination (i.e., why the nominee's teaching or support has been excellent). Those submitting nominations were also given the opportunity to consent to their nominations and data being used for research purposes; only those who consented were included in the analysis. Ethical approval was obtained from the University's ethics committee (Ref: ERP1266). Preparation of the initial dataset ( $n = 586$ ) involved four steps. First, all cases were assigned an identification number to facilitate accuracy checks throughout the coding process. Second, nominations were checked so that only student-submitted nominations of individual teachers were included in the analysis. In total, 70 nominations were removed at this stage (e.g., those submitted by teaching staff or which nominated more than one person). Third, repeat nominations were removed such that each student was only represented once in the dataset. This enabled the assumption of independence (i.e., that data are measured on different individuals) for chi-square testing to be met. The procedure was as follows. Nominations from both academic years were first compared: if a student made submissions in both years, only the submission made in the more recent year (2016/17) was included. For repeat nominations within one year, only the earliest submission was included. In total, 70 nominations were removed at this stage. Lastly, in the fourth step, student and teacher gender were coded. Gender was inferred from names provided with the nominations and the content of the nominations themselves (e.g., from personal pronouns). Internet searches were used to clarify ambiguous cases, for instance by conducting searches on first names or online staff profiles. Nominations for which no certain judgments on gender could be made were removed. In total, 28 nominations were removed at this stage, all of which were due to ambiguous student gender. This left 418 nominations in the final sample.

Thematic analysis was conducted using nVivo software. Line-by-line coding was utilised where every statement was coded at all applicable themes. The coding procedure involved two stages. In the first stage, coding was performed using a grounded approach (i.e., themes were not pre-determined) until full coverage was reached. Attaining full coverage ensured that the entire contents of the nominations had been reviewed and considered; not every statement was analysed for content (e.g., functional statements such as 'He should [...] definitely win this prize'). In the second stage, text coded at each theme was individually examined and coded at all other relevant themes missed in the first stage of coding. Additionally, a reviewing process was undertaken whereby themes were either merged or disaggregated into distinct components. This ensured internal consistency in the coding. For instance, an initial judgment of two themes as being qualitatively separate may be reversed after coding is complete and an overview of all the themes in the data becomes possible. This reviewing process allowed the coding to 'settle' into a final set of consistently coded themes. It should also be noted that unlike the thematic approaches typical of previous examinations of TE (e.g., Lubicz-Nawrocka & Bunting, 2019), there was no attempt to identify large, overarching categories; rather, a very conservative approach to merging themes was adopted. This allowed the diversity of the themes to be sustained as they appeared in the nominations.

Chi-square tests were used to investigate the gender distribution of nominations overall, and secondly, if the gender distribution for particular themes significantly differed from that of the overall sample. For the latter, a theme qualified for chi-square testing if it had been mentioned in at least 54 nominations. This ensured that every cell in the contingency table had an expected frequency of at least five, the minimum assumed in a chi-square test.

## Results

Overall, the content of the nominations ( $n = 418$ ) was very similar to that described in previous research (e.g., Lubicz-Nawrocka & Bunting, 2019), with students often describing their excellent teachers as engaging, inspiring, approachable and knowledgeable. However, quantitative analyses revealed unequal gender distributions in the overall sample and in a number of the themes mentioned in the nominations.

### *Gender distribution of overall sample*

Of the nominations, 132 (31.6%) were submitted by female students for a female teacher (FF), 128 (30.6%) by female students for a male teacher (FM), 39 (9.3%) by male students for a female teacher (MF) and lastly, 119 (28.5%) by male students for a male teacher (MM). A chi-square test showed that these differences were significant,  $\chi^2(1) = 27.66$ ,  $p < .001$ . The standardised residuals<sup>1</sup> showed that all four gender combinations substantially deviated from their expected frequencies ( $z_{FF} = 2.49$ ,  $z_{FM} = -2.07$ ,  $z_{MF} = -3.19$ ,  $z_{MM} = 2.65$ ). Same-gender nominations (FF and MM) were overrepresented, while opposite-gender nominations (FM and MF) were underrepresented. Further, while the same-gender nominations were overrepresented to a similar extent, MF nominations were more underrepresented than FM nominations. An odds ratio showed that nominations

by male students were only 0.34 times as likely to be for a teacher of the opposite gender compared to nominations by female students.

These findings suggest that students are more likely to regard teachers of the same gender as excellent, but also that male students are disproportionately unlikely to regard a female teacher as excellent. Publicly available figures from the University showed a roughly even balance of male and female teaching staff across all faculties in 2017/18. These patterns were consistent in the immediately preceding years. Thus, the uneven gender distribution found here cannot be explained by the gender distribution of staff themselves. The disproportionate infrequency of MF nominations is consistent with evidence that female teachers tend to be evaluated less favourably than male teachers, particularly by male students (Basow et al., 2006; Mengel et al., 2019; Nesdoly et al., 2020). In terms of the SST, the disproportionate infrequency of MF nominations suggests that male students hold women to higher confirmatory standards of excellence than they do for men, making excellence less readily recognisable in female teachers. In all, the unequal gender distribution of the nominations suggests that gender can influence the extent to which teachers are regarded as excellent in the first place.

### Gender distribution of themes

Of the identified themes, seven were eligible for chi-square testing: *Supportive* ( $n = 167$ ), *Engaging* ( $n = 133$ ), *Available* ( $n = 114$ ), *Passionate* ( $n = 96$ ), *Inspiring* ( $n = 81$ ), *Approachable* ( $n = 69$ ) and *Teaching techniques (generic)* ( $n = 64$ ). The theme of *Supportive* included descriptions of teachers being supportive and willing to help (e.g., '[She] is always there as a personal tutor for anything that you need help[,] support and guidance with!'). *Engaging* teachers made lessons interesting and effective at keeping students' attentions (e.g., 'I remember [my teacher] as always being especially engaging [...] It could be 9am and run over by a few minutes and people were still happy to sit there.'). *Available* teachers were easily contactable and offered support no matter the time or how busy they were (e.g., 'She will meet you if you need help [...] She makes time out of her regular day, much much more than her drop in hours on her door'). *Passionate* teachers showed enthusiasm and dedication, for instance towards their subject and their role as teachers (e.g., '[...] she is clearly passionate about helping her students achieve their potential.'). *Inspiring* teachers fostered students' interests and passions towards their subject, instilling in them a desire to engage more deeply in their learning, and in some cases, to pursue further study (e.g., 'I have re-found my passion for science and am now looking into doing a PhD because of [name's] influence.'). *Approachable* teachers made students feel at ease and were easy to talk to (e.g., '[She] makes you feel like you can talk freely about your issues, like talking to a friend.'). Finally, *Teaching techniques (generic)* pertained to broad mentions of high quality teaching (e.g., 'Always offers fantastic lectures and seminars').

Chi-square tests showed that the gender distributions of *Supportive*, *Available*, *Passionate* and *Inspiring* significantly differed from that of the overall sample, while *Engaging*, *Approachable* and *Teaching techniques (generic)* did not (see Table 1). The standardised residuals revealed distinct gender distribution profiles of both the significant and non-significant themes.



**Table 1.** The extent of deviation from the overall sample's gender distribution of nominations mentioning the seven themes which qualified for quantitative testing.

Theme	n	$\chi^2$ (df= 1)	Standardised residuals (z)			
			FF	FM	MF	MM
Supportive	167	7.38*	1.69	0.26	0.11	-2.11
Engaging	133	0.62	-0.46	0.36	-0.40	0.35
Available	114	4.32*	1.17	0.18	0.42	-1.66
Passionate	96	3.65 <sup>†</sup>	1.57	-0.81	-0.32	-0.64
Inspiring	81	8.20**	2.26	-0.56	-1.29	-1.05
Approachable	69	2.94	1.12	-0.46	0.62	-1.05
Teaching techniques (generic)	64	0.35	-0.05	-0.14	-0.40	0.42

\*  $p < .05$ . \*\*  $p < .005$ . <sup>†</sup>  $p = .056$ .

### *Available and Supportive: MM underrepresentation*

*Available* and *Supportive* were characterised by MM underrepresentation, and to a lesser extent, FF overrepresentation. In other words, the unequal gender distributions found for *Available* and *Supportive* were mainly due to male students mentioning these themes less frequently than expected when nominating male teachers, and to a lesser extent, female students mentioning these themes more frequently than expected when nominating female teachers. These findings suggest that while availability and supportiveness are important criteria for female excellence, they are less crucial for male excellence. This is consistent with gendered expectations that women engage in communal behaviours and attend to the needs of others before their own (Eagly & Karau, 2002), mirroring traditional but nevertheless persisting social roles of women as mothers and carers (Basow, 1995; Heijstra et al., 2017). The relatively large extent to which MM nominations were underrepresented for *Available* and *Supportive* is consistent with evidence that gendered patterns in evaluations of teaching tend to be more pronounced in male students than in female students, suggesting a greater reliance on gendered stereotypes in male students (Basow et al., 2006; Mengel et al., 2019; Nesdoly et al., 2020). Perhaps counter-intuitively, this may also explain why MF nominations were close to their expected frequencies for *Available* and *Supportive*. A greater reliance on gendered stereotypes would mean greater expectations for women to be available and supportive (Eagly & Karau, 2002). Thus, as the SST predicts, female teachers must meet higher minimum standards of availability and supportiveness before they are perceived by male students to possess these traits. This said, it is also possible that a larger difference was not seen due to the small number of MF nominations in the sample.

### *Inspiring and Passionate: FF overrepresentation*

*Inspiring* and *Passionate* were characterised by FF overrepresentation. This is notable as these themes pertain to stereotypically masculine teacher traits (Leogrande, 2020; Marchant & Wallace, 2016). It is possible that as these traits are associated with a socially powerful group (men), they were particularly salient in female teachers as perceived by female students. This could have created an amplifying effect which, in addition to the low minimum standards (recalling the SST's predictions), resulted in the greater likelihoods that these traits were mentioned in FF nominations. Certainly, the nominations' openness means that the salience of teacher traits would be a key factor in students' descriptions of their teachers. The salience of *Inspiring* and *Passionate* would have

been unique to the FF category due to the particular status dynamics between the student, the teacher, and the traits themselves: for male teacher evaluations, *Passionate* and *Inspiring* would have been relatively unremarkable (in addition to higher minimum standards being used), while gender bias in the MF dynamic could have reduced the likelihood that these valued masculine traits would be recognised (Eagly & Karau, 2002).

#### *Approachable, Engaging and Teaching techniques (generic): a hint of gender bias*

The gender distributions of nominations mentioning *Approachable*, *Engaging* and *Teaching techniques (generic)* did not significantly differ from that of the overall sample. Nevertheless, the standardised residuals indicated a general alignment with gendered norms. *Approachable* patterned very similarly with *Available* and *Supportive*, consistent with gendered expectations that women should be nurturing and warm when interacting with others (Eagly & Karau, 2002). A slight male bias can be seen in the themes *Engaging* and *Teaching techniques (generic)*. This is consistent with evidence showing that cognitive competencies such as academic expertise tend to be associated with male teachers in HE (Basow, 1995; Basow et al., 2006; Miller & Chamberlin, 2000; Mitchell & Martin, 2018). Additionally, the MM overrepresentation in *Teaching techniques (generic)* is similar to Basow et al.'s (2006) finding that globally positive statements (e.g., 'I love the guy', 32) were more often used to describe male teachers than female teachers.

## Overall discussion and conclusion

### *The role of gender in student conceptualisations of TE*

Significant effects of gender were found in the distributions of the overall sample and in nominations mentioning particular themes. This suggests that there are gendered differences in whether TE is recognised in the first place and what it is recognised to be. In the overall sample, same-gender nominations (FF and MM) were overrepresented while opposite-gender nominations (FM and MF) were underrepresented. A discrepancy was noted in the opposite-gender nominations, with nominations by male students being only 0.34 times as likely to be for a teacher of the opposite gender, compared to nominations by female students. This suggests that students may recognise TE more readily in teachers of the same gender as they are, but also that male students are disproportionately unlikely to recognise excellence in a female teacher. With regards to what students recognised as TE, significant effects of gender were found for the themes *Supportive*, *Available*, *Passionate* and *Inspiring*. *Supportive* and *Available* were primarily characterised by MM underrepresentation, suggesting that these are not important criteria for male excellence. The gender distributions of *Passionate* and *Inspiring* were primarily characterised by FF overrepresentation. This was an unexpected finding considering that these are stereotypically masculine traits (Leogrande, 2020). It is possible that these male-typed traits were particularly salient in female teachers, but only as perceived by female students due to the likely presence of strong gender bias in male students. The gender distributions of the remaining three themes—*Engaging*, *Approachable* and *Teaching techniques (generic)*—did not significantly differ from that of the overall sample, though were nevertheless generally aligned with gendered

norms: *Approachable* patterned similarly to *Supportive* and *Available*, while a male teacher bias was noted for *Engaging* and *Teaching techniques (generic)*.

These findings highlight that TE is not gender-neutral, and importantly, that the impact of this gendering is asymmetrical. Women, as the lower status group, experience a unique disadvantage when being evaluated by men. This can be seen in the disproportionate infrequency of MF nominations in the overall sample, a finding mirrored in previous research (Basow et al., 2006; Mengel et al., 2019; Nesdoly et al., 2020). Asymmetrical effects were also seen in the content of the nominations. While male students showed a particular dispreference for stereotypically feminine competencies in their male teachers, female students considered both their male and female teachers along more androgynous terms: they noted both masculine and feminine competencies in their female teachers, and mentioned both masculine and feminine competencies as often as expected for their male teachers. Similarly, the salience of teacher characteristics can also be asymmetrical due to power dynamics in gender. This was seen in the FF overrepresentation of the stereotypically masculine traits of *Inspiring* and *Passionate*, which was argued to be due to (a) the association of these traits with a higher status group (men) relative to the target of evaluation (female teachers) and (b) the relative absence of gender-based prejudice in the FF dynamic as opposed to the MF dynamic. In sum, these findings illustrate that TE is not gender-neutral, and further, is gendered in a way that asymmetrically disadvantages women due to differences in the social values assigned to masculinity and femininity.

### *Examining the explanatory power of the SST*

Regarding the applicability of the SST in interpreting the study's findings, evidence supporting the SST has been based on relatively controlled environments where raters make judgments in the same or similar circumstances (see review in Biernat, 2003). There has also been a strong focus on quantitative rather than qualitative variables, for instance the number of positive comments made rather than their content. By contrast, the present study examined open-ended, uncontrolled elicitations of student judgments. This made it difficult to determine which aspects of the SST were exactly relevant in interpretation. For instance, it was not clear whether submitting a nomination was considered by students to be a zero-sum or non-zero-sum action. While there was a finite number of awards to be won, students could submit multiple nominations. It is also not possible to know what students believed the judging process entailed—would decisions be based purely on the number of nominations received, or would comments also be taken into account? Was there a possibility that the awards scheme would be extended to include more award winners had there been many deserving nominees? These uncertainties made it difficult to predict whether the nominations could be characterised as zero- or non-zero-sum actions. It was also possible that students referred to both minimum and confirmatory standards when determining what to mention on their nominations. It might be expected that confirmatory standards would be the relevant threshold considering the open-endedness of the nominations. However, as discussed earlier, salience could have influenced which standard a student paid attention to: a confirmed but non-salient trait, such as a female-typed competency in a female teacher, may be left off a nomination, whereas a suspected but salient trait, such as a



male-typed competency in a female teacher, might end up being mentioned. Generally, it was interesting that some findings were assimilative (consistent with stereotypes) while others were contrastive (inconsistent with stereotypes). In open-ended elicitations of judgments, as these nominations were, there is a complex mix of factors that could influence an individual's response.

### *Recommendations for future research*

It is important that future investigations take into account the inherent subjectivity of TE and similar notions of high quality teaching. The SST (Biernat, 2003) was useful for interpreting and understanding the gendered patterns found in the data, albeit somewhat limited by the open-endedness of the nominations. Future research can extend the insights gained here by utilising both common-rule and subjective measures. For instance, student participants could be invited to identify an 'excellent' teacher of theirs, then to list the characteristics that make that teacher excellent (a subjective measure), and lastly to rate their teacher's 'performance' on each characteristic as if for a national survey on teaching quality (a common-rule measure). The thematic differences of the identified characteristics (and even their ranks, if students are directed to list characteristics in order of importance) and their associated ratings could be analysed against student and teacher gender. From the findings here and the SST's predictions, it could be expected that (a) male students will select a female teacher disproportionately infrequently (note that the very act of selecting a teacher is a zero-sum activity, and thus is predicted to show assimilative effects); (b) the characteristics listed may show mixed results in that stereotypically masculine and feminine characteristics will be used to describe both male and female teachers; however, (c) ratings for the same characteristics will reveal assimilative effects, where male teachers are rated more highly than female teachers on stereotypically masculine traits, and female teachers are rated more highly than male teachers on stereotypically feminine traits. Being able to contrast common-rule and subjective judgments on individual characteristics known to be associated with TE (or high quality teaching more generally) would allow a more granular understanding of the gendered ways in which these characteristics are perceived by students. Additionally, this would provide insight into how gendered differences in student perceptions of TE would translate into differences in ratings on formal quality and recognition processes. This design could also be expanded to incorporate more (or different) sociocultural grouping variables such as race and age.

### **Conclusion**

This study illustrates that students do not perceive TE in a gender-neutral way. Student perceptions of male and female excellence generally adhered to gendered stereotypes. This was more pronounced in male students than in female students. This study addresses gaps in the literature by exploring TE through a gender-sensitive lens (via the SST), and further, by drawing from the SST's specific predictions about different evaluative contexts. While the study's relatively fine-grained application of SST provided useful insight into the gendered patterns in the data, the open-endedness of the student nominations somewhat limited the SST's explanatory power. Future research

can continue to engage with the SST, utilising common-rule and subjective measures to investigate how gender influences student judgments on what characteristics their teachers possess, and the extent to which they perceive their teachers to embody these characteristics. Demographic categories besides gender can also be incorporated into study designs. It may be tempting to assume that there is objectivity in how teaching is evaluated, whether by students or other stakeholders; however, to ignore the role of sociocultural norms—and their inextricability from standards and benchmarks—would leave unchallenged a hegemony where those devalued in society are also devalued in HE.

## Note


1. The standardised residual is a standardised representation ( $z$ -score) of the distance between an observed frequency and its expected value, with positive values indicating overrepresentation (i.e. more frequent than expected) and a negative value indicating underrepresentation (i.e., less frequent than expected).

## Disclosure statement

No potential conflict of interest was reported by the author(s).

## ORCID

Kathryna Kwok  <http://orcid.org/0000-0002-4738-6866>

Jacqueline Potter  <http://orcid.org/0000-0002-0494-9593>

## References

- Bartram, B., Hathaway, T., & Rao, N. (2019). 'Teaching excellence' in higher education: A comparative study of English and Australian academics' perspectives. *Journal of Further and Higher Education*, 43(9), 1284–1298. <https://doi.org/10.1080/0309877X.2018.1479518>
- Basow, S. A. (1995). Student evaluations of college professors: When gender matters. *Journal of Educational Psychology*, 87(4), 656–665. <https://doi.org/10.1037/0022-0663.87.4.656>
- Basow, S. A., Phelan, J. E., & Capotosto, L. (2006). Gender patterns in college students' choices of their best and worst professors. *Psychology of Women Quarterly*, 30(1), 25–35. <https://doi.org/10.1111/j.1471-6402.2006.00259.x>
- Biernat, M. (2003). Toward a broader view of social stereotyping. *American Psychologist*, 58(12), 1019–1027. <https://doi.org/10.1037/0003-066X.58.12.1019>
- Biernat, M., & Eidelman, S. (2007). Translating subjective language in letters of recommendation: The case of the sexist professor. *European Journal of Social Psychology*, 37, 1149–1175. <https://doi.org/10.1002/ejsp.432>
- Biernat, M., Fuegen, K., & Kobrynowicz, D. (2010). Shifting standards and the inference of incompetence: Effects of formal and informal evaluation tools. *Personality and Social Psychology Bulletin*, 36(7), 855–868. <https://doi.org/10.1177/0146167210369483>
- Biernat, M., & Vescio, T. K. (2002). She swings, she hits, she's great, she's benched: Implications of gender-based shifting standards for judgment and behavior. *Personality and Social Psychology Bulletin*, 28(1), 66–77. <https://doi.org/10.1177/0146167202281006>
- Boring, A. (2017). Gender biases in student evaluations of teaching. *Journal of Public Economics*, 145, 27–41. <https://doi.org/10.1016/j.jpubeco.2016.11.006>

- Bradley, S., Kirby, E., & Madriaga, M. (2015). What students value as inspirational and transformational teaching. *Innovations in Education and Teaching International*, 52(3), 231–242. <https://doi.org/10.1080/14703297.2014.880363>
- Burford, J., Bosanquet, A., & Smith, J. (2020). ‘Homeliness meant having the fucking vacuum cleaner out’: The gendered labour of maintaining conference communities. *Gender and Education*, 32(1), 86–100. <https://doi.org/10.1080/09540253.2019.1680809>
- Delamont, S. (2006). Gender and higher education. In C. Skelton, B. Francis, & L. Smulyan (Eds.), *Chap. 13 in The SAGE handbook of gender And education* (pp. 179–189). SAGE. <https://doi.org/10.4135/9781848607996.n14>.
- Derounian, J. G. (2017). Inspirational teaching in higher education: What does it look, sound and feel like? *International Journal for the Scholarship of Teaching and Learning*, 11(1), 1. <https://doi.org/10.20429/ijstol.2017.110109>
- Eagly, A. H., & Karau, S. J. (2002). Role congruity theory of prejudice toward female leaders. *Psychological Review*, 109(3), 573–598. <https://doi.org/10.1037/0033-295X.109.3.573>
- Fan, Y., Shepherd, L. J., Slavich, E., Water, D., Stone, M., Abel, R., & Johnson, E. L. (2019). Gender and cultural bias in student evaluations: Why representation matters. *PLoS One*, 14(2), e0209749. <https://doi.org/10.1371/journal.pone.0209749>.
- Heffernan, T. (2021). Sexism, racism, prejudice, and bias: A literature review and synthesis of research surrounding student evaluations of courses and teaching. *Assessment & Evaluation in Higher Education*, 1–11. <https://doi.org/10.1080/02602938.2021.1888075>
- Heijstra, T. M., Einarsdóttir, Þ, Pétursdóttir, G. M., & Steinþórsdóttir, F. S. (2017). Testing the concept of academic housework in a European setting: Part of academic career-making or gendered Barrier to the Top? *European Educational Research Journal*, 16(2–3), 200–214. <https://doi.org/10.1177/1474904116668884>
- Holmes, A. (2020). What are the barriers and opportunities for continuing professional development for professional services staff in UK HE? *Perspectives: Policy and Practice in Higher Education*, 24(3), 79–86. <https://doi.org/10.1080/13603108.2020.1750501>
- Kandiko Howson, C. B., Coate, K., & de St Croix, T. (2018). Mid-career academic women and the prestige economy. *Higher Education Research & Development*, 37(3), 533–548. <https://doi.org/10.1080/07294360.2017.1411337>
- Kulik, J. A. (2001). Student ratings: validity, utility, and controversy. *New Directions for Institutional Research*, 2001(109), 9–25. <https://doi.org/10.1002/ir.1>
- Laube, H., Massoni, K., Sprague, J., & Ferber, A. L. (2007). The impact of gender on the evaluation of teaching: What we know and what we can do. *NWSA Journal*, 19(3), 87–104. <http://www.jstor.org/stable/40071230>
- Leogrande, C. (2020). Those who can’t, teach: Representations and challenges of male teachers. In K. Aavik, C. Bland, J. Hoegaerts, & J. Salminen (Eds.) (pp. 53–72). Walter de Gruyter.
- Lubicz-Nawrocka, T., & Bunting, K. (2019). Student perceptions of teaching excellence: An analysis of student-led teaching award nomination data. *Teaching in Higher Education*, 24(1), 63–80. <https://doi.org/10.1080/13562517.2018.1461620>
- MacNell, L., Driscoll, A., & Hunt, A. N. (2015). What’s in a name: Exposing gender bias in student ratings of teaching. *Innovative Higher Education*, 40(4), 291–303. <https://doi.org/10.1007/s10755-014-9313-4>
- Marchant, T., & Wallace, M. (2016). Gender balance in teaching awards: Evidence from 18 years of national data. *Journal of Higher Education Policy and Management*, 38(4), 393–405. <https://doi.org/10.1080/1360080X.2016.1181885>
- Mengel, F., Sauermann, J., & Zöllitz, U. (2019). Gender bias in teaching evaluation. *Journal of the European Economic Association*, 17(2), 535–566. <https://doi.org/10.1093/jeea/jvx057>
- Miller, J., & Chamberlin, M. (2000). Women are teachers, men are professors: A study of student perceptions. *Teaching Sociology*, 28(4), 283–298. <https://doi.org/10.2307/1318580>
- Mitchell, K. M. W., & Martin, J. (2018). Gender bias in student evaluations. *PS: Political Science & Politics*, 51(3), 648–652. <https://doi.org/10.1017/S104909651800001X>



- Moore, S., & Kuol, N. (2007). Retrospective insights on teaching: Exploring teaching excellence through the eyes of alumni. *Journal of Further and Higher Education*, 31(2), 133–143. <https://doi.org/10.1080/03098770701267598>
- Nesdoly, N., Tulk, C., & Mantler, J. (2020). The effects of perceived professor competence, warmth and gender on students' likelihood to register for a course. *Assessment & Evaluation in Higher Education*, 45(5), 666–679. <https://doi.org/10.1080/02602938.2019.1689381>
- O'Connor, P., López, E. M., O'Hagan, C., Wolfram, A., Aye, M., Chizzola, V., Mich, O., Apostolov, G., Topuzova, I., Sağlam, G., & Tan, M. G. (2020). Micro-political practices in higher education: A challenge to excellence as a rationalising myth? *Critical Studies in Education*, 61(2), 195–211. <https://doi.org/10.1080/17508487.2017.1381629>
- O'Connor, P., & O'Hagan, C. (2016). Excellence in university academic staff evaluation: A problematic reality? *Studies in Higher Education*, 41(11), 1943–1957. <https://doi.org/10.1080/03075079.2014.1000292>
- Revell, A., & Wainwright, E. (2009). What makes lectures 'unmissable'? Insights into teaching excellence and active learning. *Journal of Geography in Higher Education*, 33(2), 209–223. <https://doi.org/10.1080/03098260802276771>
- Shaha, M., Berg, J., Wenzel, J., Shaefer, J., Terhaar, M., Klimmek, R., Melvin, K. C., & Belcher, A. E. (2013). Excellence in teaching And learning: Students' perspective. *Journal of Nursing Education and Practice*, 3(12), 35–46. <https://doi.org/10.5430/jnep.v3n12p35>
- Sprague, J., & Massoni, K. (2005). Student evaluations and gendered expectations: What we can't count can hurt us. *Sex Roles*, 53(11/12), 779–793. <https://doi.org/10.1007/s11199-005-8292-4>
- Su, F., & Wood, M. (2012). What makes a good university lecturer? Students' perceptions of teaching excellence. *Journal of Applied Research in Higher Education*, 4(2), 142–155. <https://doi.org/10.1108/17581181211273110>
- Universities Australia. (2021). *Australian Awards for University Teaching*. Accessed March 2, 2021. <https://www.universitiesaustralia.edu.au/policy-submissions/teaching-learning-funding/australian-awards-for-university-teaching/>
- van den Brink, M., & Benschop, Y. (2012). Gender practices in the construction of academic excellence: Sheep with five legs. *Organization*, 19(4), 507–524. <https://doi.org/10.1080/13562517.2017.1301911>
- Wood, M., & Su, F. (2017). What makes an excellent lecturer? Academics' perspectives on the discourse of 'teaching excellence' in higher education. *Teaching in Higher Education*, 22(4), 451–466. <https://doi.org/10.1080/13562517.2017.1301911>