
Article:
Gelber, K., Brennan, K., Duriesmith, D. orcid.org/0000-0003-4210-6113 et al. (1 more author) (2022) Gendered mundanities : gender bias in student evaluations of teaching in political science. Australian Journal of Political Science, 57 (2). pp. 199-220. ISSN 1036-1146

https://doi.org/10.1080/10361146.2022.2043241

Australian Journal of Political Science

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/cajp20

Gendered mundanities: gender bias in student evaluations of teaching in political science

Katharine Gelber, Katie Brennan, David Duriesmith & Ellyse Fenton

To cite this article: Katharine Gelber, Katie Brennan, David Duriesmith & Ellyse Fenton (2022) Gendered mundanities: gender bias in student evaluations of teaching in political science, Australian Journal of Political Science, 57:2, 199-220, DOI: 10.1080/10361146.2022.2043241

To link to this article: https://doi.org/10.1080/10361146.2022.2043241

Published online: 27 Feb 2022.

AUSTRALIAN JOURNAL OF POLITICAL SCIENCE
2022, VOL. 57, NO. 2, 199–220
https://doi.org/10.1080/10361146.2022.2043241

Routledge
Taylor & Francis Group

ORIGINAL ARTICLE

OPEN ACCESS Check for updates

Gendered mundanities: gender bias in student evaluations of teaching in political science

Katharine Gelber, Katie Brennan, David Duriesmith * and Ellyse Fenton

School of Political Science and International Studies, University of Queensland, Brisbane, Australia

ABSTRACT

Much research has been undertaken on gender bias in student evaluations of teaching (SETs) in universities, with inconsistent findings. We undertake a qualitative analysis of the comments in four years of SETs in a school of political science and international relations in a highly regarded Australian university. We ask, can the same evaluations produce different results when analysed qualitatively rather than quantitatively? And do students evaluate male-identified and female-identified teachers differently, and if so what are the differences? We show that qualitative analysis can reveal gender bias that is invisible in quantitative analysis. We find that female-identified staff are evaluated more positively than their male counterparts for undertaking time-intensive, stereotypically feminine, emotional labour. Male-identified staff are evaluated more positively for their technical expertise and teaching style. This suggests SETs evaluate gender-stereotypical behaviour rather than only teaching quality, and has significant implications for their use in universities.

ARTICLE HISTORY
Accepted 9 February 2022

KEYWORDS
Gender; student evaluations; bias; political science; international relations; university; tertiary education

Introduction

The literature on gender and student evaluations of teaching (SETs) has yielded conflicting findings on the extent to which systemic gender bias[1] disadvantages university teachers[2] who identify as female.[3] An analysis of the literature suggests that research design impacts on whether findings show gender bias substantively. Even in contexts where quantitative scores do not appear to be different on the basis of gender, it may be that qualitative approaches can render visible the assumptions underpinning students' evaluations of teaching by male and female-identified teachers. Considering that SETs are routinely used in universities in job applications, confirmation of appointments and promotions, their reliability as a measure of performance may be undermined if they facilitate a gender-bias in how students reward teaching staff. Existing studies have suggested that if SETs are gender-biased, then their use risks incentivising strict adherence to gender-stereotyped behaviour.

In this paper, we investigate two research questions. The first is whether analysis of qualitative comments can provide a different finding from a quantitative analysis of

CONTACT Katharine Gelber k.gelber@uq.edu.au
*Current address: Department of Politics and International Relations, University of Sheffield, Sheffield, UK.

the same SETs regarding the presence or otherwise of gender bias. If it does, this underpins how students may assign quantitative scores in SETs. The second is whether students evaluate male-identified and female-identified teachers differently, and if so what these differences might be. To do this, we engage in an analysis of the qualitative comments provided by students[4] in SETs from 2015 to 2018 in a school of political science and international relations at a prestigious Australian university. After outlining the literature on gender and SETS and justifying our focus on qualitative comments, we set out our approach and findings. We find that a qualitative analysis is able to illuminate gendered expectations of university teachers in ways that quantitative studies do not. We find further that students provide gendered evaluations of staff in two ways. First, students who identify as male and female provide similar evaluations of staff who identify as female, but different evaluations of staff who identify as male. Second, students' evaluations of teaching by staff who identify as female centre more strongly around the nurturing aspects of university teaching, including providing time, being warm, and being approachable for feedback. By contrast, students' comments about teaching by male-identified staff centre more strongly around their technical expertise and lecture structure.

The implications of these findings are manifold. Emotional labour in the academy is often under- or devalued (Rickett and Morris 2021, 90). Expectations on female-identified academics to undertake this work may help to create and enforce unequal expectations, which others have suggested funnel female staff into 'high-service/low-status' patterns of work that hamper their careers (Alter et al. 2020). These expectations are likely to reinforce negative career impacts that result from unrewarded emotional labour, identified by Berry and Cassidy (2013) as including eroded job autonomy and decreasing job satisfaction. They are also likely to reinforce the disadvantage that can be faced by academics who have prioritised teaching in seeking promotion especially at senior levels (Parker 2008; Subbaye 2017), and the existing under-representation of women at senior levels in universities (Curtis 2011; Winchester et al. 2006, 511). Our findings show that when students criticise teaching staff who identify as female they often rely on harmful but easily overlooked forms of gender bias, which we term 'gendered mundanities'. We argue these gendered mundanities risk naturalising unequal expectations for staff based on gender, and that they may have significant implications for the careers of female-identifying teaching staff in universities.

How gender influences student evaluations of teaching

An extensive and complex literature on gender and SETs in universities has been developed since the 1970s, with most studies located in the United States. Here, we provide an overview of trends in that literature, with a focus on the ways in which the methodological approach adopted has influenced findings. The primary questions underlying this research have been whether SETs are susceptible to bias based on the gender of the instructor, and whether teachers who identify as female receive systematically lower ratings than teachers who identify as male. We also highlight the small body of literature on how discipline impacts bias in SETs, and the specific dynamics identified in evaluations of teaching in political science and international relations. We argue that method plays a significant role in determining whether studies are able to locate

gender-bias, and that without qualitative analysis the kind of gendered mundanities we interrogate in this paper are rendered invisible.

The literature on SETs is highly divided over the role of gender. Among early experimental and observational studies, some found differences in ratings of teachers based on gender (e.g. Kaschak 1978; Lombardo and Tocci 1979; Bray and Howard 1980; Basow and Silberg 1987; Sidanius and Crane 1989), whereas others did not (e.g. Elmore and LaPointe 1974, 1975; Harris 1975, 1976; Basow and Suzan Distenfeld 1985; Basow and Howe 1987; Statham, Richardson, and Cook 1991; Basow 1995). Some have reported such mixed results that the authors are hesitant to draw any conclusions about gender (e.g. Hancock, Shannon, and Trentham 1993).

Similarly, those conducting meta-analyses of the literature (e.g. Feldman 1992, 1993; Wright and Jenkins-Guarnieri 2012) have often found it difficult to draw definitive conclusions about the impact of gender on student evaluations due to the sharp divide across studies. Attention has been drawn to methodological limitations such as the aggregation of data which masks important contextual differences, such as which disciplinary areas are covered in each study (Miller and Chamberlin 2000, 284–285). Reviews of the literature have produced conflicting results, and been alleged to suffer from methodological shortcomings (see Laube et al. 2007, 88–89). For example, McKeachie cites one study in support of his claim of a lack of evidence of gender bias (1979), Marsh concludes the research is theoretically confused and methodologically inadequate (1987, 305–328), Aleamoni (1999) relies on limited evidence, and Cashin (1995) and Theall and Franklin (2001) cite only Feldman's (1992, 1993) inconclusive meta-analyses to support their conclusion that gender is not an important influence on evaluation results.

Two arguments help to explain inconsistencies in findings. The first is that research design has been flawed, especially due to studies insufficiently controlling for or taking into account situational variables like class size and discipline (Centra and Gaubatz 2000) or because experimental studies have not adequately reflected students' experiences in the classroom (e.g. Li and Benton 2017; see also Rivera and Tilcsik 2019). Some studies which have attempted to address these kinds of shortcomings have produced more consistent findings of bias. These include one that manipulated an online teaching environment to mislead students as to the gender of their teacher (MacNell, Driscoll, and Hunt 2015, 301), which found significant differences in the way students rated teachers they believed to be male and female. Wagner, Rieger, and Voorvelt (2016) explored mixed-gender teaching teams and found evidence of bias, with female-identified teachers significantly less likely to attain the teaching evaluation score necessary for promotion. Boring (2017) and Mengel, Sauermaan, and Zolitz (2019) used large samples of evaluations generated in real classroom settings, with both finding evidence of gender bias. Other studies have attempted to account for a larger number of situational variables, such as Martin (2016), who found that course size interacts with gender to influence bias, with larger courses more likely to generate bias against female-identified teachers. The geographical remit of studies has expanded to include Spain (Fernandez and Mateo 1997), Canada (Cramer and Alexitch 2000), and Australia (Fan et al. 2019; Worthington 2002). Some of these studies have found significant evidence of gender, and other, bias (e.g. Fan et al. 2019; Fernandez and Mateo 1997; Cramer and Alexitch 2000; Martin 2016), but others have been less conclusive (e.g. Bavishi, Hebel, and Madera 2010; Zabaleta 2007).

The second argument is perhaps more profound; that gender may lead students to reward different behaviour, even if the scores given are similar. This means that statistical evaluations of raw scores cannot capture gender bias. Studies that have sought to address this have adopted creative methods. For example, Statham, Richardson, and Cook (1991) combined a quantitative analysis with classroom observations and teacher interviews in a large cross-disciplinary sample. They identified gendered differences in approaches to teaching, with female-identified teachers more likely than their male counterparts to involve students in teaching, privilege interaction in teaching, and value students' input. They found that teachers were rewarded for adhering to gender-stereotypical behaviour: when female-identified instructors were warm, personable and encouraging they received higher ratings than when they presented material in didactic style, while male-identified instructors received higher ratings for occupying the persona of 'expert', even if this meant interrupting or admonishing students. Other studies have shown that students have gendered expectations of their teachers, and that this informs the behaviour they reward in evaluations. Basow (1995) found that female-identified students gave higher ratings to female-identified faculty on gender-stereotypical behaviour, such as sensitivity, and contact with students, and male-identified students gave lower ratings to female-identified faculty on male stereotypical behaviour, such as fairness and intellectual stimulation.

Experimental studies further demonstrate the importance of understanding why particular scores are given, such as Arbuckle and Williams (2003) analysis of SETs that assessed a stick figure image of a teacher, which students were alternately told was female or male, finding that students consistently rated the perceived male professor higher on expressiveness. They hypothesise that this is because expressiveness in lecturing is more consistent with the masculine stereotype of agency and dynamism than the feminine stereotype of empathy and warmth. Bennett (1982) found evidence that students both expected and reported getting more personal attention from female-identified teachers, yet were more likely to rate them as not sufficiently available. As Laube et al. summarise, 'these students' reference point for "enough availability" clearly shifted to a higher order for women teachers' (2007, 92; see also Burns-Glover and Veith 1995; Bachen, McLoughlin, and Garcia 1999). Overall, these studies suggest that when female-identified teachers fail to exhibit stereotypically feminine qualities associated with care and time, they are criticised by students, while male-identified teachers who do not demonstrate these qualities are not. Female-identified teachers are expected to demonstrate competence on stereotypically feminine, as well as stereotypically masculine, aspects of teaching in order to receive high overall ratings, while male-identified teachers are expected to demonstrate competence on only stereotypically masculine dimensions.

Further studies have endeavoured to unlock the rationale behind scores by focussing on the qualitative comments in SETs. Laube et al. have argued that qualitative comments 'may reveal nuances that cannot be tapped by quantitative indicators that assume raters are using the same criteria and metric across genders' (2007, 96). While there are fewer studies that have engaged with this type of research, their findings appear more consistent than with other methods. For example, Bachen, McLoughlin, and Garcia (1999) found male-identified students were more likely than female-identified students to characterise female-identified teachers as lacking in professionalism or failing to

intellectually challenge students, while female-identified students tended to characterise female-identified teachers as both caring and challenging. Female-identified teachers were described in gender-stereotypical terms as approachable, warm and caring, and when they deviated from these qualities, they received harsher evaluation scores. Analysis of 386,277 qualitative responses led Adams et al. (2021) to conclude that in many cases SETs may be more well suited to assessing whether the instructor is 'doing gender right' than any objective measure of teaching excellence.

Looking specifically at negative comments, Lindahl and Unger (2010) found gender-specific derogatory comments about female-identified teachers, and suggest that student cruelty towards teachers is gendered. Wallace, Lewis, and Allen (2019) found that female-identified teachers were more likely than their male-identified counterparts to receive negative personal comments, a tendency which was more pronounced for Black women. White men in their study received no negative comments of a personal nature. Storage et al. (2016) found male-identified students were nearly twice as likely to describe male-identified professors as 'brilliant' and more than three times as likely to describe them as a 'genius' compared with female-identified professors. Terkik et al. (2016) found that gender influences the language students use to describe their teachers, with female-identified staff more likely to be called 'teachers', and male-identified staff called 'professors'; female-identified staff were more likely to be described as 'amazing', 'wonderful' and 'loved', while male-identified staff were more likely to be described as 'knowledgeable', 'interesting' and 'funny'. Male-identified staff were rated higher by all students on knowledge of the subject matter. Similar findings from Sprague and Massoni (2005, 791) suggest that male-identified staff are rewarded for adopting the role of 'entertainer' while women are expected to 'nurture' students.

What of the literature specifically relating to political science and international relations? Mitchell and Martin (2018) undertook content analysis of the comments students made on evaluations of two professors teaching political science courses, one man and one woman, as well as the comments made about these instructors at RateMyProfessors.[5] They found that female-identified staff are evaluated differently to their male counterparts. Confirming the findings of Terkik et al. (2016) and Miller and Chamberlin (2000), they found the female-identified instructor was more likely to be referred to as a 'teacher' and the male-identified instructor as a 'professor'. They additionally found that the female-identified instructor was more likely to receive comments on her personality, physical appearance and attractiveness. They conclude that students value different things in their male and female professors and use different criteria to evaluate them. Rosen (2018) studied the quantitative ratings on RateMyProfessors and found that while the difference between ratings of male and female teachers is small, political science shows much larger disparities than other disciplines like chemistry (2018, 41). Martin (2016) has theorised these differences as reflecting 'role incongruity', where women are seen to be inappropriate for leading certain courses, such as those with large enrolments or which cover subjects that are the traditional domain of men. Such a bias within political science is likely to punish women when students expect an instructor who is assertive, ambitious or authoritative rather than nurturing or sensitive (Martin 2016, 314). Due to the content taught within political science, this is likely to create additional barriers for women who teach subjects such as on leadership, national security, terrorism, or law. When including consideration for race as well as gender, Chávez

and Mitchell (2019, 273) conclude that within political science, 'SETs might constitute another "weep hole" for women and minorities in academic career pipelines that structurally contribute to higher attrition and lower achievement'. This dynamic has led professional advice papers in political science journals to recommend women 'lean in' and 'make strategic adjustments that may improve their SETs' by being attentive, smiling, and wearing age-appropriate attire, all while showing sufficient authority in the classroom so that they can live up to students' expectations for them to be 'all-knowing and all-nurturing' when teaching (Dion 2008, 855). This literature suggests that gaining precise knowledge on the character of gender bias in student evaluations within political science should be a priority, if we are to understand how women are pressured to become the 'caretakers of the academic family' in order to succeed (Alter et al. 2020).

Despite the contestation in broader literature on whether gender leads women to receive lower SET scores, qualitative studies demonstrate that gender can interact with student evaluations in ways that are not immediately apparent from an examination of statistical results. If the expectations on the part of students from male-identified and female-identified teachers are different, and some types of teaching behaviour are rewarded for male-identified teachers and critiqued for female-identified teachers, and vice versa, this may result in SETs not functioning to measure only the quality of teaching, but also to reinforce gender-stereotypical behaviour in teaching. Further, if qualitative assessments which accompany quantitative scores send gender-stereotyped messages about teachers' behaviour their suitability as a metric of success for promotions and probation is seriously undermined. These studies demonstrate that students' open-ended comments can be a valuable source of insight into the more complex ways that gender influences teaching evaluations, in so far as they render visible the bias that may not be visible using other methods. Considering what existing studies have shown, there is need for further interrogation of how qualitative comments reflect gender-stereotypes, a need which this study addresses.

Our study – justification, data collection, and method

We have chosen to analyse the qualitative comments in student teacher evaluations in all courses offered in the School of Political Science and International Studies at the University of Queensland from 2015 to 2018. The Office of Research Ethics at the University of Queensland (Approval No. 2019001557) granted approval for this study. Pre-anonymised data were obtained from the Institute for Teaching and Learning Innovation (ITaLI). The Institute had already used the data for a large scale, quantitative study, which produced no evidence of gender bias (personal correspondence with the Institute, 2 December, 2019). The National Statement on Ethical Conduct in Human Research recognises that, due to the quantity of data collected by the university, it is 'impracticable' to obtain consent directly from students for use of the evaluation data in research (s2.3.10). It recognises further that consent from students may be expressed by means including 'return of a survey' (s2.2.5). The University of Queensland recognises these consent exceptions as permitting them to consider participation in, and completion of, surveys to constitute consent.[6] Staff participation in student evaluation processes is conducted under the University's Course and Teacher Surveys Policy, which requires compliance with all appropriate ethical standards.

The University of Queensland describes itself as 'one of Australia's leading research institutions', and performs well in global rankings.[7] It is a comprehensive university and a member of the Group of Eight, Australia's leading universities, and the discipline of Political Science is one among many high performing organisational units. This makes these SETs a good focus for study, since the student experience is unlikely to be affected by factors such as specialism by the institution, or an excessively small student cohort. The addition of Australian data makes a specific contribution to the literature, which is dominated by studies in the United States.

The school teaches both undergraduate and postgraduate courses. In undergraduate courses, lecture sizes can be up to 400 students, with tutorials typically of approximately 22 students. Postgraduate courses can be offered for 10–70 students. The surveys analysed here were conducted prior to COVID-19, and therefore predominantly reflect experiences with face to face teaching. Enrolment in every course at the university results in a request to students, two weeks before the end of the teaching semester, to complete a course evaluation and a teacher evaluation (a sample survey is provided in Appendix). The request is received via email and provides a link to complete the surveys online. Students have three weeks in which to respond. The data generated by the survey (both quantitative scores and qualitative comments) are only accessible by the course coordinator, the Head of School, and the staff member's supervisor (if that is not the Head of School). The data are recorded confidentially by the University of Queensland's Institute for Teaching and Learning Innovation (ITaLI). If a course receives fewer than six responses, quantitative results are not recorded. In order to ensure we analysed the most relevant data, and because Fan et al. (2019) argue that bias evident in teaching evaluations may not appear in course evaluations, we focussed on the teaching evaluations.[8]

The teaching evaluation contains two opportunities for students to provide written responses:

Question 9 [Q9]: What aspects of this teacher's approach best helped your learning?

Question 10 [Q10]: What would you have liked this teacher to have done differently?

We obtained the data from ITaLI in the form of a CSV file.[9] The data had already been through the University's internal screening process, which removes egregiously offensive comments before the results are passed on to teachers or recorded by the university. From the data we obtained, 16 comments (0.15% of the total) had been redacted. We ran the larger dataset containing the redacted comments through the same analytical process as the data without the redacted comments. The two produced stable, consistent results, which demonstrates that the redaction of these comments did not alter the validity of our findings. We note that this is a very small proportion of the data, and that as discussed above we are interested in the everyday, mundane ways in which students respond to their perceptions of the gender of their teachers. Egregiously abusive comments are not necessary to examine this phenomenon.

The total number of evaluations completed by students with their gender identity recorded was 10,222. Evaluations were completed by 6860 students identifying as female (67.1%) and 3362 students identifying as male (32.9%). These proportions relatively closely reflect the gender identities of students enrolled in courses in the study

period: the total number of students enrolled was 26,481 students, of whom 16,301 were female-identified (61.6%) and 10,180 were male-identified (38.4%). Across the whole study, 38.6% of students returned completed evaluations, which is a significant proportion and renders the results valid. Female-identified students had a response rate of 42%, and male-identified students had a response rate of 33%.

The majority of courses are taught by tenured members of academic staff ranked from Lecturer to Professor, with a proportion each year (15–20%) taught by casual staff with a PhD qualification where permanent staff members are absent due to leave. Due to the size of the dataset and the need to preserve the anonymity of teaching staff, we did not disaggregate the data according to course size, course level, faculty rank, or whether the teacher was a permanent staff member. We have been unable to disaggregate according to other factors which are highlighted in the literature to be relevant to assessing bias, such as ethnicity, age, or class background because we are analysing data retrospectively collected by the University of Queensland. This means our analysis is limited by the variables which were collected at the time, and we do not have a reliable way to integrate further variables into our analysis. While we recognise the importance of research examining the interplay between these factors and believe this should occupy the attention of further studies, we focussed exclusively on gender. We disaggregated the data according to the gender of the teacher and the gender of the student undertaking the evaluation. The dataset includes a metadata category (gender.map), which labels each response with the associated staff and student gender identities, resulting in four tags:

fstufsta (female student evaluating female staff)	fstumsta (female student evaluating male staff),
mstufsta (male student evaluating female staff)	mstumsta (male student evaluating male staff).

In the SETs we analysed, the quantitative scores achieved by male and female-identified teachers showed no statistically significant difference (personal correspondence from the Institute of Teaching and Learning Innovation, 2 December, 2020). This is consistent with our analysis of the literature, and enables us to determine whether a focus on the qualitative comments produces different results than a quantitative analysis among the same student evaluations.

We subjected the responses to qualitative content analysis to study 'language in use' (Wetherell, Taylor, and Yates 2001, 3; Van Dijk 1997, 2). The goal was to render explicit patterns in utterances (Chilton and Schäffner 2002, 18; Taylor 2001, 6–9) by exposing their components (Van Dijk 1997, 5).

Patterns in the utterances were identified using Leximancer, a text mining software that analyses texts to identify the frequency of words and phrases and their relationship to other words and phrases. Where words co-occur as 'groups of related words … that travel together in the text' (Leximancer 2018) they create a concept. Concept co-occurrence identifies relationships between concepts, and where related concepts occur in close proximity they create themes. While there are several software packages available for analysing medium to large-sized qualitative datasets, we chose Leximancer for three reasons. First, Leximancer is commonly used to analyse qualitative data, especially short texts (Palmer 2013; Sotiriadou, Brouwers, and Le 2014). Second, it is well suited to handling survey data (Scott, Pachana, and Sofronoff 2011; Myers et al. 2012); the

Leximancer User Guide lists qualitative survey data as a core use example (2018). Third, we are sensitive to the issues surrounding researcher bias in data analysis, which can be exacerbated when using software options that require coding as a preliminary step (Sotiriadou, Brouwers, and Le 2014).

Unlike some other tools, Leximancer does not begin with a pre-determined list of concepts, but discovers concepts and themes from the data itself, which a researcher then interprets. This permits valid inferences to be drawn about the ideas contained in the texts without presupposing the results. Leximancer is claimed to produce highly reliable findings (Palmer 2013, 224) because the coding is repeatable, which also renders intercoder reliability tests unnecessary (Chilton and Schäffner 2002, 27; Taylor 2001, 16–20). It is possible for other researchers to analyse the same or similar data without relying on our pre-coding. Our research findings from the conceptual maps rely on human interpretation, but the concept maps themselves are a result of Leximancer's processing algorithm. It is important to note here that a critical step in using Leximancer is to run an analysis multiple times to ensure that the concept maps are relatively stable or consistent. We did this, which gives us confidence in the content and structure of the concept maps we have interpreted in this project.

Results

We chose first to examine all answers to Q9, which asked the students to discuss what most helped their learning. Figure 1 shows how the answers to this question are grouped. We then examined all answers from male-identified students and female-identified students to Q10, which asked them what could be done differently in the course. Figure 2 shows how the answers to this question are grouped. Both Figures 1 and 2 are to be understood as 'concept maps', which are 'indicative and … not a quantitative statement of fact' (Palmer 2013, 226). The results allow us to explore relationships in some depth.

While our data includes 10,248 student survey responses, the number of responses to each question varies. A survey response is valid when a student replies to at least one question. Thus Figure 1 includes 5585 analysable responses to Q9 and Figure 2 includes 4185 analysable responses to Q10.

Leximancer finds concepts in the data and then ranks them by the sum of their co-occurrence with the other concepts it finds. Leximancer's algorithm then goes through these concepts in ranked order to look for themes. The first theme always takes the most highly connected concept as its seed concept. Leximancer then looks at the next concept on the ranked list and determines whether it is close enough to an existing theme to join it, or if it should become the seed concept of a new theme. In the concept map, therefore, themes are ranked by the number of text blocks associated with the concepts in each theme, and this relative level of connectedness is demonstrated on a heat map scale with red being the most connected, orange the second most, and dark blue the least. Leximancer also allows viewing of the themes listed in rank order with the associated number of hits. In Figure 1, the top four themes are of a relatively similar size and thus have a similar level of connectedness to the overall dataset. This is shown in the number of hits for each of the top four themes: 'lectures' (4891), 'students' (4743), 'he' (4728), and 'she' (4292).

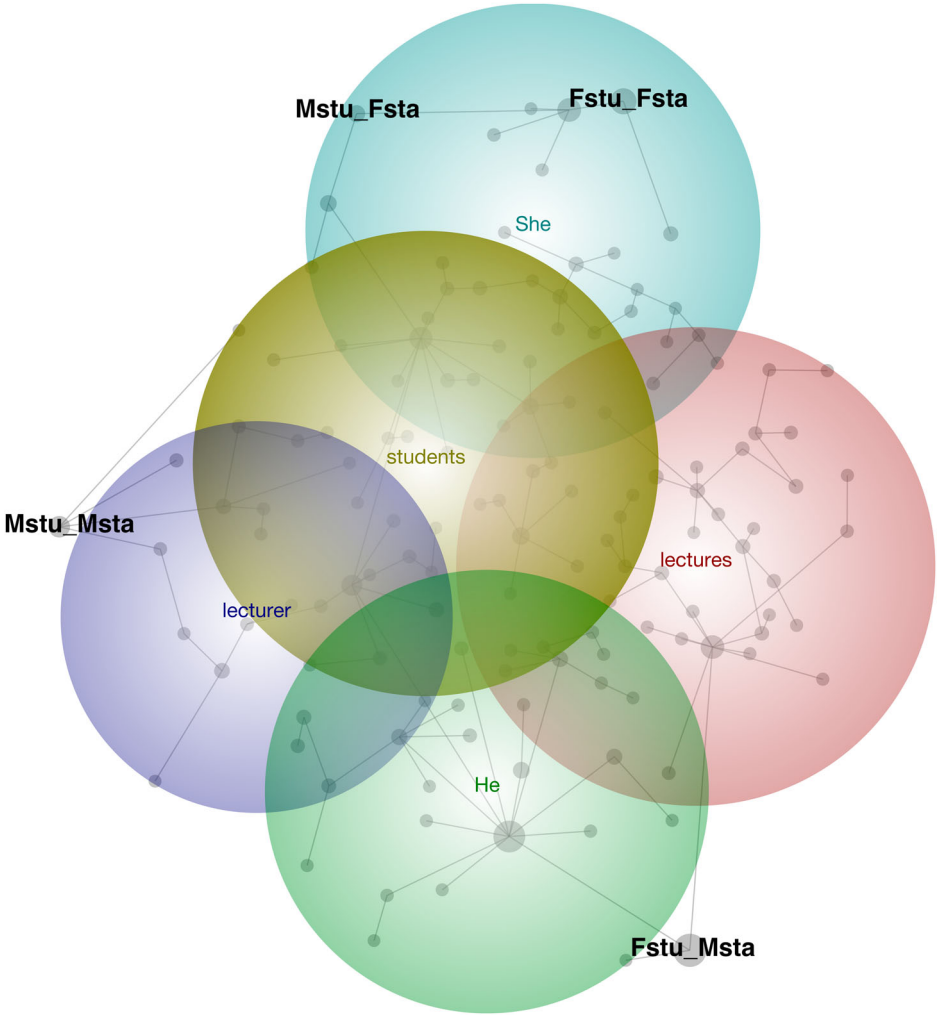


Figure 1. Groupings by student and staff gender identity of answers to Q9.

Our first finding is that there is a close relationship between the use of female gender pronouns (she and her) and students' responses to female-identified staff, compared with a much looser and more diverse relationship between the use of male gender pronouns (he and him) in students' responses to male-identified staff. This implies that male and female-identified students respond to female-identified staff members in similar ways – fstufsta and mstufsta are always close to each other on the concept maps. It implies that there is a comparative diversity in how male and female-identified students respond to male-identified staff – because fstumsta and mstumsta are always relatively further apart on the concept maps.

In analysing the numbers behind the concept maps in Figures 1 and 2 we discovered that the data available for responses to Q10 (What would you like the teacher to have done differently in this course?) show less connectedness or co-occurrence than in the responses to Q9 (What aspects of this teacher's approach best helped your learning?).

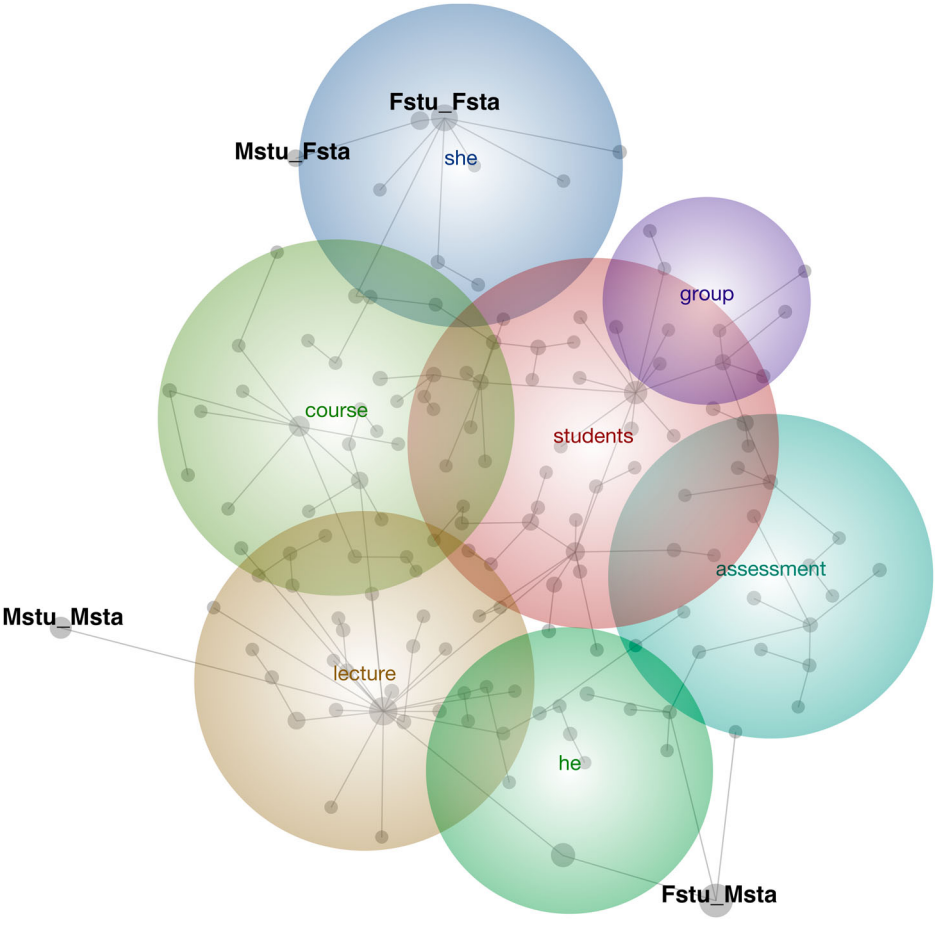


Figure 2. Groupings by student and staff gender identity of answers to Q10.

This lack of coherence has two primary causes. First, as noted above, the dataset for Q10 is smaller than for Q9. Second, the responses to Q10 include many that counter the direction of the question (which asks students to suggest improvements), with about 14% of the responses being a variation of 'N/A' or 'Nothing', 8% beginning with a similar response followed by a positive comment (e.g. 'Nil – I loved this course!'), and 3% being comprised solely of punctuation including emoticons. Of the 4275 responses to Q10, at least 25% did not answer the question that was asked. Based on this we chose to focus our more detailed analysis on the responses to Q9.

We delved more deeply into the responses to Q9 to examine the ways in which students talked about their teachers' strengths. We examined more closely the 'she' theme, to examine the concepts most closely related to the 'she' concept, and the tags associated with female-identified students and male-identified students when they discussed how a female-identified teacher helped their learning. The results are shown in Figure 3.

This produces our second finding. After the seed concept ('she'), the next ten concepts in the 'she' theme are (in the order of their co-occurrence rank in the entire dataset):

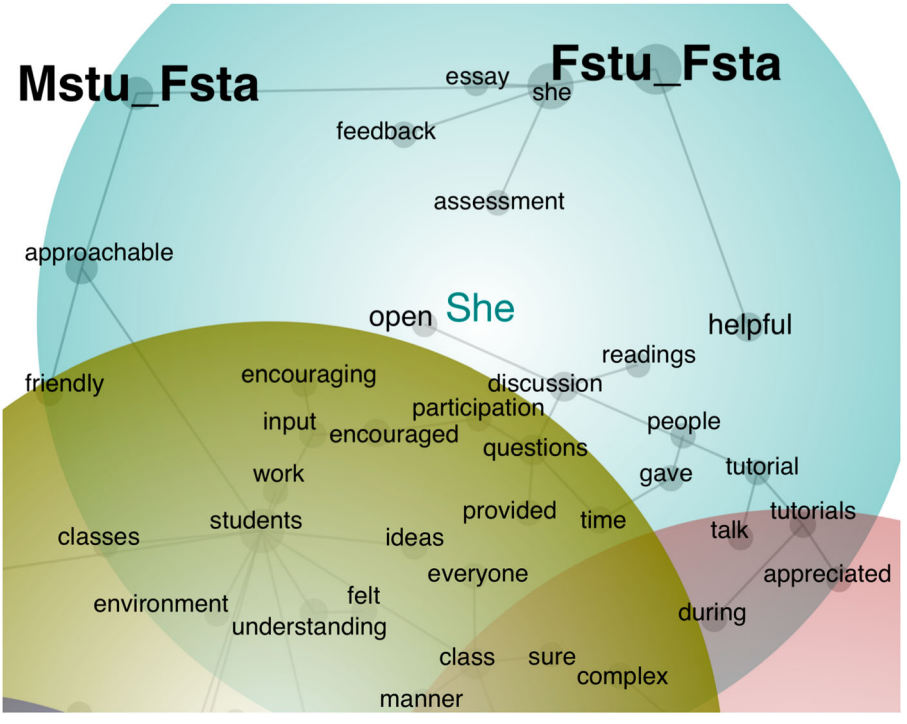


Figure 3. How students discuss how female-identified staff help their learning.

approachable, questions, discussion, helpful, encouraged, input, time, friendly, ideas, feedback, and tutorials. Both male and female-identified students are more likely to use all these concepts in relation to female-identified staff than in relation to male-identified staff. These differences in likelihood are subtle; students of both gender identities use these concepts for staff of both gender identities, but the pattern was consistent.[10] It is also important to remember that while the percentage figures may look low, these are still the most frequently connected concepts with female-identified teachers (Table 1).

Also, we note that female-identified students most directly connect female staff to the concept 'helpful', while male-identified students most directly connect female staff to

Table 1. Likelihood that male and female-identified students used the top concepts from 'She' to discuss female-identified staff.

Concepts	Mstu-Msta likelihood	Mstu-Fsta likelihood	Fstu-Msta likelihood	Fstu-Fsta likelihood
Approachable	3.18%	3.80%	3.74%	3.99%
Questions	1.76%	1.90%	2.23%	2.40%
Discussion	1.47%	2.52%	1.97%	2.22%
Helpful	1.33%	1.57%	2.18%	2.46%
Encouraged	0.87%	0.95%	1.50%	1.55%
Input	0.77%	0.95%	1.00%	1.04%
Time	0.89%	0.95%	1.02%	1.38%
Friendly	0.96%	1.47%	0.93%	1.16%
Ideas	0.94%	1.19%	0.63%	0.83%
Feedback	0.82%	1.04%	0.94%	1.06%
Tutorials	0.36%	0.90%	0.84%	0.92%

'approachable' and 'friendly'. Female-identified staff members are most likely to be praised for being friendly, helpful, encouraging, and giving their students time in an emotionally supportive way.

We also drilled down into the qualitative comments themselves. We have selected those that are representative of the comments in the data. Typical comments in relation to female-identified staff included: 'She was incredibly approachable, friendly, and encouraging of student input/discussion', 'She was very patient and helpful and gave good feedback on our work', 'She definitely stimulates meaningful discussions during the tutorial, by being encouraging and approachable', 'Her approachable, encouraging and respectful communication with the students encouraged learning and participation', 'She was extremely helpful with her feedback on the independent study project and was very encouraging and friendly', 'She was also very approachable, encouraging students to talk privately to her after lectures when she would gladly answer questions', and 'Also she provided soooooo much feedback on my essay which I really appreciated'. Others were even more detailed: 'She was very helpful and approachable, provided me with very insightful feedback on my assignments, and always made it clear that I was welcome to ask questions and speak to her about problems' and 'She was so talented at this; making concepts clear – an extremely approachable, kind, caring, respectful and joyful person to be around (She never made you feel embarrassed to ask for help or ask silly questions)'. These responses emphasise positive associations between female staff members and traits of being approachable, helpful, and encouraging.

We also examined how male and female-identified students talked about male-identified staff members' strengths by examining the 'he' theme. Interestingly, female-identified students are closely related to male pronouns when discussing male-identified staff members, whereas male-identified students are more closely related to the 'lecture' theme when discussing male-identified staff members. We show the detail in this theme in Figure 4, which enables us to see the concepts most closely related to the 'he' theme, and the tags associated with female-identified students and male-identified students when they discussed how a male-identified teacher helped their learning.

This produces our third finding. The responses to male-identified staff are far more diverse than the responses to female-identified staff. While descriptors such as 'engaging' do feature in the 'he' theme, the emphasis is not on the same kinds of qualities as for female-identified staff. After the seed concept ('he'), the next ten concepts are (in the order of their co-occurrence rank in the entire data set): engaging, teaching, passionate, clearly, style, enjoyed, engaged, enthusiasm, world, makes, and humour. Whereas in the 'she' theme we saw a consistent pattern in how male and female-identified students used the top concepts, in the 'he' theme we see greater variability (Table 2).

Both male and female-identified students are more likely to use the following concepts to refer to male-identified staff – teaching, passionate, style, engaged, enthusiasm, world, makes, and humour. Male-identified students are more likely to use 'enjoyed' in discussing male staff while female-identified students are more likely to use it about female staff, and female-identified students are more likely to use the concept 'engaging' in relation to male staff while male-identified students are more likely to use it in relation to female-identified staff.

In looking more broadly at the concept map, male-identified students responding to the question about how male-identified staff helped their learning are most closely tied to the following concepts: knowledge, knowledgeable, inspiring, excellent, theoretical,

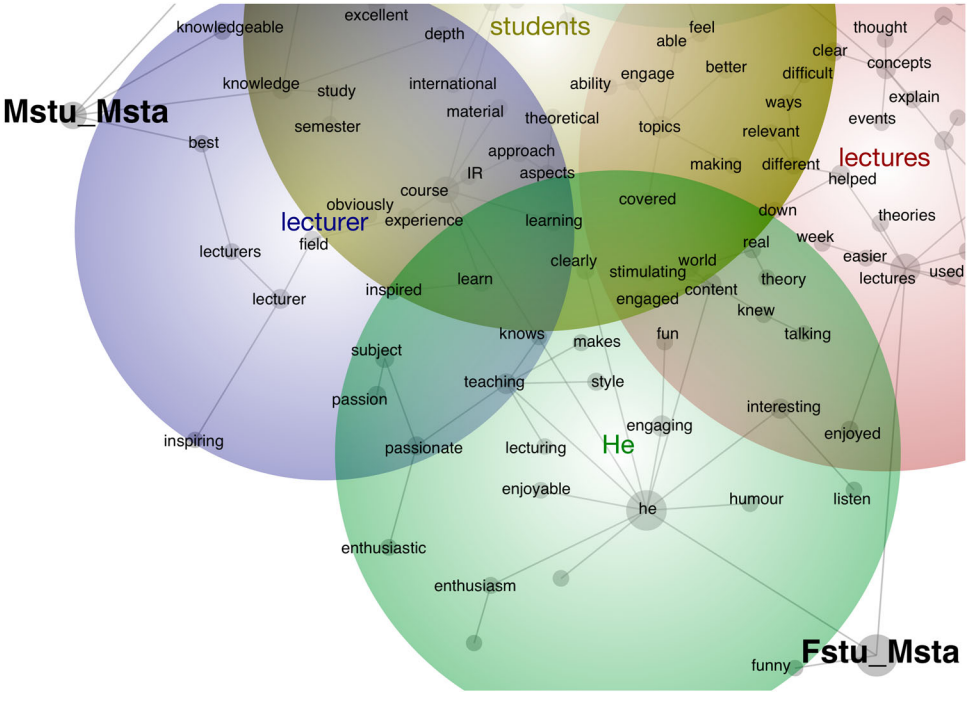


Figure 4. How students discuss how male-identified staff help their learning.

passionate and best. Female-identified students responding to the question about how male-identified staff members helped their learning are most closely tied to the following concepts: funny, knows, fun, and male gender pronouns. Across both male and female-identified students, comments about male-identified staff members highlight their enthusiasm, passion and teaching style.

Again, the data can be illuminated further by examining the comments. Male-identified students' comments about male-identified teachers included: 'His enthusiasm and his love of the topic inspired me to learn and better myself as an IR scholar', 'His passion and enthusiasm about the subject was [*sic*] really felt and definitely inspired me to do well and help grow my interest in the Political science field', and

Table 2. The likelihood that male and female-identified students were to use each of the top concepts from the 'He' theme to discuss male-identified staff.

Concepts	Mstu-Msta likelihood	Mstu-Fsta likelihood	Fstu-Msta likelihood	Fstu-Fsta likelihood
Engaging	2.82%	2.94%	3.47%	2.87%
Teaching	3.18%	2.14%	2.55%	2.18%
Passionate	1.33%	1.04%	2.22%	1.65%
Clearly	1.08%	1.14%	1.29%	1.45%
Style	1.25%	0.90%	1.09%	1.04%
Enjoyed	0.87%	0.71%	1.07%	1.32%
Engaged	0.58%	0.52%	1.00%	0.77%
Enthusiasm	1.06%	0.19%	1.02%	0.45%
World	0.63%	0.52%	0.61%	0.35%
Makes	0.60%	0.43%	0.61%	0.47%
Humour	0.58%	0.28%	0.91%	0.29%

> Basically [*sic*] same comments as before, [name] is by far my favourite lecturer at UQ, he does all the best courses, which actually enourage [*sic*] students to critical [*sic*] analyse, and to challenge their own positions. Other lecturers need to learn from [name], him and [name] are the best.

Female-identified students' comments included: 'The content is really excellent and [name] knows his stuff and is clearly passionate about it – his laid back style was easy to listen to', 'I found his lecturing style clear, engaging and felt his sense of humour made the lectures fun and worthwhile', and 'He has a great teaching style and is able to keep the students' attention on topic, also has a good sense of humour which makes the lectures enjoyable'. We can therefore see a different set of descriptors applied to male-identified teaching staff compared with female-identified teaching staff. When commenting on how male-identified teachers help learning, students emphasise different traits.

Analysis

We have reached three primary findings. The first is that male and female-identified students tend to evaluate female-identified staff in similar ways, but male-identified teachers in different ways. This suggests that gender matters in relation to the staff member teaching, and in relation to the cohort evaluating them, and that gender is doing some work that needs to be further examined. Second, we found that when students comment on what they found helpful about the teaching they received, the traits most associated with female-identified university teachers were related to stereotypically gendered expectations of women. Female-identified teachers were described as helping students' learning when they were approachable, encouraged questions and discussion, were helpful, allowed for student input, gave time, were friendly and gave more feedback including outside of class time. Of all the activities that a university academic may undertake in their teaching role, these are likely to be the most time consuming and emotionally consequential, which means they risk detracting from teachers' time for conducting research, applying for grants, and writing.

In comparison, and finally, the traits most commonly associated with male-identified teachers were more varied than those associated with female-identified teachers and were likely to be related to stereotypically gendered, masculine expectations. Male-identified teachers were described as helping students' learning when they were engaging, passionate, enjoyable, and enthusiastic. Male-identified students in particular, rewarded male-identified teachers for being knowledgeable and theoretically minded. These traits are unlikely to require additional time beyond normal preparation for teaching, or to constitute additional, burdensome, emotional labour (as classified by Berry and Cassidy 2013, cited above). Further, as higher education institutions already tend to privilege research excellence in promotion to senior levels (Parker 2008; Subbaye 2017), such a gendered division is likely to reward men for devoting less time to teaching, at the same time as disadvantaging women who are likely to have less time available for research due to the emotional labour required to achieve good student evaluation results.

Based on these findings we believe that the gender bias in SETs represents a gendered mundanity, a harmful expectation of gendered behaviour which is occluded because of its 'everyday' nature. Our data did not contain direct slurs, hate speech or explicit gendered demands, nor did the comments we analysed explicitly state that female-identified staff

should care more *because they are a woman*. Rather, the patterns we have identified constitute regular, institutionally endorsed reminders about what behaviour is required from men and women to be seen to be good at their teaching role and to achieve equally good SET results.

One might ask if such comments matter, considering that quantitative SET scores tend to be cited by applicants, or required to be provided, in job applications, and probation and promotion committees, rather than qualitative comments. Such an interpretation would misunderstand the function of qualitative comments in SETs. Qualitative comments give insight into *why* students have ascribed a certain score and what the staff must do to achieve it. Yet they show expectations that female and male-identified staff behave in gender-stereotypical ways to achieve good scores. Therefore the institutions that rely on SETs are rewarding female and male staff for behaviours that conform to gendered stereotypes, and that are likely to have differentiated impacts on the amount of time and energy available for research activities. This suggests that it is incorrect to regard SETs as purely a measure to assess the quality of teaching performance.

Conclusion

This study sought to investigate whether a qualitative analysis of the comments in student evaluations might produce different findings from a quantitative analysis of the same dataset. The answer to this question is that it can. While statistical analysis showed no significant difference in the evaluation results for male and female-identified teachers, qualitative analysis does show differentiated results.

The second question we asked was whether this analysis showed evidence of gender bias, and if so how. Our analysis reinforces the findings of earlier studies that suggest that gender bias exists in, and affects students' evaluations of, university teaching. One of the most striking components of our findings is how mundanely these gendered differences present. Students did not provide radically different quantitative assessments of their instructors, nor did gendered language appear in the form of overt slurs, or gendered abuse. Rather, gender operated by producing subtle, but unequal, expectations on male-identified and female-identified staff. These differences are concealed by studies which do not delve into the qualitative comments in SETS, or which only search for the most overt forms of discriminatory language.

Our findings indicate that male and female-identified staff face differing expectations from students. For female-identified staff, the kinds of expectations that exist (being approachable, giving time, providing feedback, listening, and doing so out of class hours) are likely to entail greater demands on their time. In contrast, expectations on male-identified staff (being expert, knowledgeable, funny, enthusiastic, and passionate) are likely to fall within the remit of their normal professional duties. In order to do well in teaching evaluations, female-identified staff are expected to perform more of specific kinds of unrewarding and intensive labour than their male-identified counterparts. To the extent that SETs are used in hiring and promotion decisions, these issues of gender bias are likely playing a part in the persistence of gender inequities in academia (Curtis 2011; Winchester et al. 2006).

While we believe these findings are significant, and contribute to the growing body of work on the issues present in SETs, our study has a number of limitations. First, we focused on one school of political science and international relations. This means that

we could not capture the range of dynamics present across disciplines (indicated by previous studies), or geography. Second, we did not distinguish other factors which are likely to interact with gender. As indicated by Fan et al. (2019), student bias against female-identified staff can be amplified when they are from non-English speaking backgrounds. Examining this would require greater specificity in the data. This exploratory study indicates further evaluation of the comments section of SETs is warranted.

Despite its limitations, this study shows that SETs reflect stereotypically, and mundanely gendered expectations of teaching staff, and that gendered mundanities in students' expectations of their university teachers distort the demands placed on female-identified teachers within higher education. Of course, universities need to evaluate teaching performance. To correct for these risks, universities could consider other measures including altering the emphasis placed on SETs in job application, confirmation and promotion processes; investing appropriately in evaluating teaching in other ways (such as peer review, for example); providing framing statements to students undertaking evaluations that recognise and seek to combat the potential for unconscious bias; and investing in trained evaluators who can go into classrooms to perform moderated, in-person evaluations. Even if such measures are not adopted, great care needs to be taken to ensure that SETs are not regarded purely as a measure of teaching performance.

Notes

1. We define gender bias consistently with the Australian *Sex Discrimination Act 1984* (Cth), s5: treating a person 'less favourably than … the discriminator treats or would treat a person of a different sex,' in a way that 'imposes, or proposes to impose, a condition, requirement or practice that has, or is likely to have, the effect of disadvantaging persons of the same sex as the aggrieved person.'
2. Hereafter, we use the term 'teachers', and note we are exclusively focussed on the tertiary/university sector throughout this article.
3. We recognise the contestations in using binary and essentialising terms like 'woman' and 'man' or 'female' and 'male'. Contributions to the literature use varying terminology. Therefore, we use the terms female identified, and male identified.
4. Our study relied on university records to identify staff and students as male and female identified. Of the 10,248 survey responses, just over 100 had left gender blank or listed 'other'. The university indicated that most of these entries may be data errors, and data on non-binary participants are unreliable.
5. An online site where students anonymously rate teachers on qualities such as clarity, helpfulness, and attractiveness.
6. https://itali.uq.edu.au/advancing-teaching/evaluation-teaching/student-evaluation-course-and-teacher-secat.
7. https://about.uq.edu.au/university-profile.
8. Students are shown a photo of the staff member to ensure they are assessing the correct person.
9. Each row contains an anonymised survey response along with metadata about the student responding, the course in question, and the associated staff member.
10. Our method does not require us to demonstrate statistical significance, merely that the data show a greater tendency to associate female and male identified teachers with specific traits.

Disclosure statement

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

Notes on contributors

Katie Brennan is a Postdoctoral Research Fellow in the School of Political Science and International Studies at the University of Queensland. Her primary research areas are international relations, political theory, digital media, and popular culture.

David Duriesmith is a Lecturer in the Department of Politics and International Relations at the University of Sheffield. His primary research focuses on feminist international relations theory, masculinities, armed conflict and violence prevention. His work has been published in *Security Dialogue, The Review of International Studies, International Theory, The International Feminist Journal of Politics* and his 2017 book *Masculinity and New War* was published with Routledge.

Ellyse Fenton is an independent researcher working.

Katharine Gelber is Head of the School of Political Science and International Studies at the University of Queensland, a Fellow of the Academy of Social Sciences Australia, and a former Australian Research Council Future Fellow (2012-2015). She has recently jointly edited *Free Speech in the Digital Age* (Oxford Uni Press, 2019) with Susan Brison, and has published articles in journals including *Journal of Public Policy, Parliamentary Affairs, Law and Society Review*, *Political Studies*, *Contemporary Political Theory*, *Melbourne University Law Review*, *Review of International Studies*, and the *Australian Journal of Human Rights*.

ORCID

Katharine Gelber http://orcid.org/0000-0002-0239-3552
David Duriesmith http://orcid.org/0000-0003-4210-6113

References

Adams, Sophie, Sheree Bekker, Yanan Fan, Tess Gordon, Laura J. Shepherd, Eve Slavich, and David Waters. 2021. "Gender Bias in Student Evaluations of Teaching: Punish[ing] Those Who Fail To Do Their Gender Right." *Higher Education*. doi:10.1007/s10734-021-00704-9.

Aleamoni, Lawrence M. 1999. "Student Rating Myths Versus Research Facts from 1924 to 1998." *Journal of Personnel Evaluation in Education* 13: 153–166.

Alter, Karen J., Jean Clipperton, Emily Schraudenbach, and Laura Rozier. 2020. "Gender and Status in American Political Science: Who Determines Whether a Scholar is Noteworthy?" *Perspectives on Politics* 18 (4): 1048–1067.

Arbuckle, Julianne, and Benne D. Williams. 2003. "Students' Perceptions of Expressiveness: Age and Gender Effects on Teacher Evaluations." *Sex Roles* 49 (9/10): 507–516.

Bachen, Christine M., Moira M. McLoughlin, and Sarah S. Garcia. 1999. "Assessing the Role of Gender in College Students' Evaluation of Faculty." *Communication Education* 48 (3): 193–210.

Basow, Susan A. 1995. "Student Evaluations of College Professors: When Gender Matters." *Journal of Educational Psychology* 87: 656–665.

Basow, Susan A., and Karen G. Howe. 1987. "Evaluations of College Professors: Effects of Professors' Sex-Type, and Sex, and Students' Sex." *Psychological Reports* 60: 671–678.

Basow, Susan A., and Nancy T. Silberg. 1987. "Student Evaluations of College Professors: Are Female and Male Professors Rated Differently?" *Journal of Educational Psychology* 79 (3): 308–314.

Basow, Susan A., and M. Suzan Distenfeld. 1985. "Teacher Expressiveness: More Important for Male Teachers Than Female Teachers?" *Journal of Educational Psychology* 77 (1): 45–52.

Bavishi, Anish, Michelle R. Hebel, and Juan M. Madera. 2010. "The Effect of Professor Ethnicity and Gender on Student Evaluations: Judged Before Met." *Journal of Diversity in Higher Education* 3 (4): 245–256.

Bennett, Sheila Kishler. 1982. "Student Perceptions of and Expectations for Male and Female Instructors: Evidence Relating to the Question of Gender Bias in Teaching Evaluation." *Journal of Educational Psychology* 74 (2): 170–179.
Berry, K. E., and S. F. Cassidy. 2013. "Emotional Labour in University Lecturers: Considerations for Higher Education Institutions." *Journal of Curriculum and Teaching* 2 (2): 22–36.
Boring, Anne. 2017. "Gender Biases in Student Evaluations of Teaching." *Journal of Public Economics* 145: 27–41.
Bray, James, and George Howard. 1980. "Interaction of Teacher and Student Sex and Sex Role Orientations and Student Evaluations of College Instruction." *Contemporary Educational Psychology* 5: 241–248.
Burns-Glover, Alyson L., and Dale J. Veith. 1995. "Revisiting Gender and Teaching Evaluations: Sex *Still* Makes a Difference." *Journal of Social Behaviour and Personality* 19 (4): 69–80.
Cashin, William E. 1995. "Student Ratings of Teaching: The Research Revisited." IDEA Paper No. 32. https://eric.ed.gov/?id=ED402338.
Centra, John A., and Noreen B. Gaubatz. 2000. "Is There Gender Bias in Student Evaluations of Teaching?" *The Journal of Higher Education* 71 (1): 17–33.
Chávez, Kerry, and Kristina M. W. Mitchell. 2019. "Exploring Bias in Student Evaluations: Gender, Race, and Ethnicity." *PS: Politics* 53 (2): 270–274.
Chilton, Paul, and Claudia Schäffner. 2002. "Introduction: Themes and Principles in the Analysis of Political Discourse." In *Politics as Text and Talk: Analytic Approaches to Political Discourse*, edited by Paul Chilton, and Claudia Schäffner, 1–44. Amsterdam: John Benjamins.
Cramer, Kenneth N., and Louise R. Alexitch. 2000. "Student Evaluations of College Professors: Identifying Sources of Bias." *The Canadian Journal of Higher Education* 30 (2): 143–164.
Curtis, John W. 2011. *Persistent Inequity: Gender and Academic Employment*. Report from the American Association of University Professors. https://www.aaup.org/NR/rdonlyres/08E023AB-E6D8-4DBD-99A0-24E5EB73A760/0/persistent_inequity.pdf.
Dion, Michelle. 2008. "All-Knowing or All-Nurturing? Student Expectations, Gender Roles, and Practical Suggestions for Women in the Classroom." *PS: Political Science and Politics* 41 (4): 853–856.
Elmore, Patricia B., and Karen A. LaPointe. 1974. "Effect of Teacher Sex and Student Sex on the Evaluation of College Instructors." *Journal of Educational Psychology* 66 (3): 386–389.
Elmore, Patricia B., and Karen A. LaPointe. 1975. "Effect of Teacher Sex, Student Sex, and Teacher Warmth on the Evaluation of College Instructors." *Journal of Educational Psychology* 67 (3): 368–374.
Fan, Y., L. J. Shepherd, E. Slavich, D. Waters, M. Stone, R. Abel, and E. L. Johnston. 2019. "Gender and Cultural Bias in Student Evaluations: Why Representation Matters." *PLoS One* 14 (2): e0209749.
Feldman, Kenneth A. 1992. "College Students' Views of Male and Female College Teachers: Part I – Evidence from the Social Laboratory and Experiments." *Research in Higher Education* 33: 317–375.
Feldman, Kenneth A. 1993. "'College Students' Views of Male and Female College Teachers: Part II – Evidence from Students' Evaluations of Their Classroom Teachers." *Research in Higher Education* 34: 151–211.
Fernandez, Juan, and Miguel Angel Mateo. 1997. "Student and Faculty Gender in Ratings of University Teaching Quality." *Sex Roles* 37 (11/12): 997–1003.
Hancock, Gregory R., David M. Shannon, and Landa L. Trentham. 1993. "Student and Teacher Gender in Ratings of University Faculty." *Journal of Personnel Evaluation in Education* 6: 235–248.
Harris, Mary B. 1975. "Sex Role Stereotypes and Teacher Evaluations." *Journal of Educational Psychology* 67: 751–756.
Harris, Mary B. 1976. "The Effects of Sex, Sex-Stereotyped Descriptions and Institution on Evaluations of Teachers." *Sex Roles* 2: 15–21.
Kaschak, Ellyn. 1978. "Sex Bias in Student Evaluations of College Professors." *Psychology of Women Quarterly* 2 (3): 235–243.

Laube, Heather, Kelley Massoni, Joey Sprague, and Abby L. Ferber. 2007. "The Impact of Gender on the Evaluation of Teaching: What We Know and What We Can Do." *NWSA Journal* 19 (3): 87–104.
Leximancer. 2018. *Leximancer User Guide, Release 4.5*. https://doc.leximancer.com/doc/LeximancerManual.pdf.
Li, Dan, and Stephen L. Benton. 2017. "The Effects of Instructor Gender and Discipline Group on Student Ratings of Instruction." *IDEA Research Report #10*. https://eric.ed.gov/?id=ED588355.
Lindahl, Mary W., and Michael L. Unger. 2010. "Cruelty in Student Teaching Evaluations." *College Teaching* 58 (3): 71–76.
Lombardo, John P., and Mary Ellen Tocci. 1979. "Attribution of Positive and Negative Characteristics of Instructors as a Function of Attractiveness and Sex of Instructor and Sex of Subject." *Perceptual and Motor Skills* 48: 491–494.
MacNell, Lillian, Adam Driscoll, and Andrea N. Hunt. 2015. "What's in a Name: Exposing Gender Bias in Student Ratings of Teaching." *Innovative Higher Education* 40: 291–303.
Marsh, Herbert W. 1987. "Students' Evaluations of University Teaching: Research Findings, Methodological Issues, and Directions for Future Research." *International Journal of Educational Research* 11: 253–388.
Martin, Lisa L. 2016. "Gender, Teaching Evaluations, and Professional Success in Political Science." *PS: Political Science and Politics* 49 (2): 313–319.
McKeachie, Wilbert J. 1979. "Student Ratings of Faculty: A Reprise." *Academe* 65 (6): 384–397.
Mengel, Friederike, Jan Sauermaan, and Ulf Zolitz. 2019. "Gender Bias in Teaching Evaluations." *Journal of the European Economic Association* 17 (2): 535–566.
Miller, JoAnn, and Marilyn Chamberlin. 2000. "Women are Teachers, Men are Professors: A Study of Student Perceptions." *Teaching Sociology* 28 (4): 283–298.
Mitchell, Kristina, and Jonathan Martin. 2018. "Gender Bias in Student Evaluations." *PS: Political Science and Politics* 51 (3): 648–652.
Myers, S. A., M. J. Blackmore, T. F. Smith, and R. W. Bill Carter. 2012. "Climate Change and Stewardship: Strategies to Build Community Resilience in the Capricorn Coast." *Australasian Journal of Environmental Management* 19 (3): 164–181.
Palmer, Philip. 2013. "Exploring Attitudes to Financial Reporting in the Australian Not-for-Profit Sector." *Accounting and Finance* 53: 217–241.
Parker, Jonathon. 2008. "Comparing Research and Teaching in University Promotion Criteria." *Higher Education Quarterly* 62 (3): 237–251.
Rickett, Bridgette, and Anna Morris. 2021. "Mopping Up Tears in the Academy – Working-Class Academics, Belonging and the Necessity for Emotional Labour in UK Academia." *Discourse: Studies in the Cultural Politics of Education* 42 (1): 87–101.
Rivera, Lauren A., and Andras Tilcsik. 2019. "Scaling Down Inequality: Ratings Scales, Gender Bias, and the Architecture of Evaluation." *American Sociological Review* 84 (2): 248–274.
Rosen, Andrew S. 2018. "Correlations, Trends, and Potential Biases Among Publicly Accessible Web-Based Student Evaluations of Teaching." *Assessment and Evaluation in Higher Education* 43 (1): 1–14.
Scott, Theresa L., Nancy A. Pachana, and Kate Sofronoff. 2011. "Survey of Current Curriculum Practices Within Australian Postgraduate Clinical Training Programmes: Students' and Programme Directors' Perspectives." *Australian Psychologist* 46 (2): 77–89.
Sidanius, Jim, and Marie Crane. 1989. "Job Evaluation and Gender: The Case of University Faculty." *Journal of Applied Psychology* 19 (2): 174–197.
Sotiriadou, Popi, Jessie Brouwers, and Tuan-Anh Le. 2014. "Choosing a Qualitative Data Analysis Tool: A Comparison of NVivo and Leximancer." *Annals of Leisure Research* 17 (2): 218–234.
Sprague, Joey, and Kelly Massoni. 2005. "Student Evaluations and Gendered Expectations: What We Can't Count Can Hurt Us." *Sex Roles* 53 (11): 779–793.
Statham, Anne, Laurel Richardson, and Judith A. Cook. 1991. *Gender and University Teaching: A Negotiated Difference*. Albany, NY: SUNY Press.
Storage, Daniel, Zachary Horne, Andrei Cimpian, and Sarah-Jane Leslie. 2016. "The Frequency of "Brilliant" and "Genius" in Teaching Evaluations Predicts the Representation of Women and African Americans Across Fields." *PLoS One* 11: 3. doi:10.1371/journal.pone.0150194.

Subbaye, R. 2017. "The Shrinking Professoriate: Academic Promotion and University Teaching." *South African Journal of Higher Education* 31 (3): 249–273.

Taylor, Stephanie. 2001. "Locating and Conducting Discourse Analytic Research." In *Discourse as Data: A Guide for Analysis*, edited by M. Wetherell, S. Taylor, and S. Yates, 5–48. Thousand Oaks: Sage.

Terkik, Andamlak, Emily Prud'hommeaux, Cecilia O. Alm, Christopher M. Homan, and Scott Franklin. 2016. "Analysing Gender Bias in Student Evaluations." *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, 868–876. Osaka, Japan, December 11–17.

Theall, Michael, and Jennifer Franklin. 2001. "Looking for Bias in the All the Wrong Places: A Search for Truth or a Witch Hunt in Student Ratings of Instruction?" *New Directions for Institutional Research* 109: 45–56.

Van Dijk, Teun. 1997. "The Study of Discourse." In *Discourse as Structure and Process*, edited by T. van Dijk, 1–34. Thousand Oaks: Sage.

Wagner, Natascha, Matthias Rieger, and Katherine Voorvelt. 2016. "Gender, Ethnicity, and Teaching Evaluations: Evidence from Mixed Teaching Teams." *Economics of Education Review* 54: 79–94.

Wallace, Sherri L., Angela K. Lewis, and Marcus D. Allen. 2019. "The State of the Literature on Student Evaluations of Teaching and an Exploratory Analysis of Written Comments: Who Benefits Most?" *College Teaching* 67 (1): 1–14.

Wetherell, Margaret, Stephanie Taylor, and Simeon Yates. 2001. *Discourse Theory and Practice: A Reader*. Thousand Oaks: Sage.

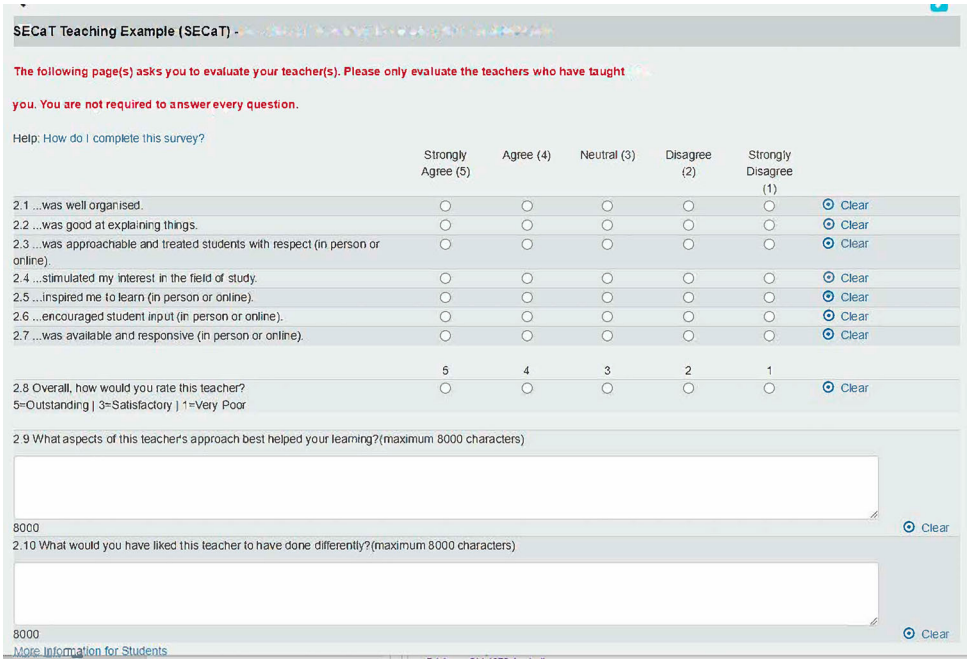
Winchester, H., S. Lorenzo, L. Browning, and C. Chesterman. 2006. "Academic Women's Promotions in Australian Universities." *Employee Relations* 28 (6): 505–522.

Worthington, Andrew C. 2002. "The Impact of Student Perceptions and Characteristics on Teaching Evaluations: A Case Study in Finance Education." *Assessment and Evaluation in Higher Education* 27 (1): 49–64.

Wright, Stephen L., and Michael A. Jenkins-Guarnieri. 2012. "Student Evaluations of Teaching: Combining the Meta-Analyses and Demonstrating Further Evidence for Effective Use." *Assessment and Evaluation in Higher Education* 37 (6): 683–699.

Zabaleta, Francisco. 2007. "The Use and Misuse of Student Evaluations of Teaching." *Teaching in Higher Education* 12 (1): 55–76.

Appendix



SECaT Teaching Example (SECaT) -

The following page(s) asks you to evaluate your teacher(s). Please only evaluate the teachers who have taught

you. You are not required to answer every question.

Help: How do I complete this survey?

	Strongly Agree (5)	Agree (4)	Neutral (3)	Disagree (2)	Strongly Disagree (1)	
2.1 ...was well organised.	○	○	○	○	○	Clear
2.2 ...was good at explaining things.	○	○	○	○	○	Clear
2.3 ...was approachable and treated students with respect (in person or online).	○	○	○	○	○	Clear
2.4 ...stimulated my interest in the field of study.	○	○	○	○	○	Clear
2.5 ...inspired me to learn (in person or online).	○	○	○	○	○	Clear
2.6 ...encouraged student input (in person or online).	○	○	○	○	○	Clear
2.7 ...was available and responsive (in person or online).	○	○	○	○	○	Clear
	5	4	3	2	1	
2.8 Overall, how would you rate this teacher? 5=Outstanding | 3=Satisfactory | 1=Very Poor	○	○	○	○	○	Clear

2.9 What aspects of this teacher's approach best helped your learning?(maximum 8000 characters)

8000 Clear

2.10 What would you have liked this teacher to have done differently?(maximum 8000 characters)

8000 Clear

More Information for Students