

Bond University
Research Repository



Student evaluation of teaching: Reactions of Australian academics to anonymous non-constructive student commentary.

Hutchinson, Marie; Coutts, Rosanne A.; Massey, Deb; Nasrawi, Dima; Fielden, Jann; Lee, Megan; Lakeman, Richard

Published in:
Assessment and Evaluation in Higher Education

DOI:
<https://doi.org/10.1080/02602938.2023.2195598>

Licence:
CC BY-NC-ND

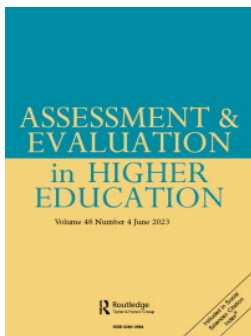
[Link to output in Bond University research repository.](#)

Recommended citation(APA):
Hutchinson, M., Coutts, R. A., Massey, D., Nasrawi, D., Fielden, J., Lee, M., & Lakeman, R. (2023). Student evaluation of teaching: Reactions of Australian academics to anonymous non-constructive student commentary. *Assessment and Evaluation in Higher Education*, 1-11. <https://doi.org/10.1080/02602938.2023.2195598>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

For more information, or if you believe that this document breaches copyright, please contact the Bond University research repository coordinator.



Student evaluation of teaching: reactions of Australian academics to anonymous non-constructive student commentary

Marie Hutchinson, Rosanne Coutts, Debbie Massey, Dima Nasrawi, Jann Fielden, Megan Lee & Richard Lakeman

To cite this article: Marie Hutchinson, Rosanne Coutts, Debbie Massey, Dima Nasrawi, Jann Fielden, Megan Lee & Richard Lakeman (2023): Student evaluation of teaching: reactions of Australian academics to anonymous non-constructive student commentary, *Assessment & Evaluation in Higher Education*, DOI: [10.1080/02602938.2023.2195598](https://doi.org/10.1080/02602938.2023.2195598)

To link to this article: <https://doi.org/10.1080/02602938.2023.2195598>



© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 08 Apr 2023.



Submit your article to this journal [↗](#)



Article views: 816



View related articles [↗](#)



View Crossmark data [↗](#)

Student evaluation of teaching: reactions of Australian academics to anonymous non-constructive student commentary

Marie Hutchinson^a , Rosanne Coutts^a , Debbie Massey^{a,b} , Dima Nasrawi^a ,
Jann Fielden^a , Megan Lee^{a,c}  and Richard Lakeman^a 

^aFaculty of Health, Southern Cross University, Lismore, Australia; ^bSchool of Nursing and Midwifery, Edith Cowan University, Joondalup, Western Australia; ^cFaculty of Society and Design, Bond University, Robina, Australia

ABSTRACT

Within Australian higher education, student evaluation of teaching (SET) is regularly conducted and data are utilised for quality control and staff appraisal. Within current methodologies, students can anonymously provide further feedback as written commentary. There is now growing evidence that, once this narrative becomes derogatory or abusive, it may have the potential to create harm. To investigate staff reactions to receiving anonymous non-constructive commentary, a one group point in time design was constructed, and a survey conducted. Participants ($N=741$) from a broad cross-section of Australian universities responded to Likert questions asking about their reactions. A significant impact was revealed according to age for mental health, stress and professional confidence, with younger and tenured academics indicating the most vulnerability. There were no differences across gender. Non-health disciplines with teaching loads greater than 50% reported an impact of anonymous SET on mental health and professional confidence. Being casually or seasonally employed or from an ethnic background was shown to have a significant effect on professional confidence. Findings suggest that the potential for higher education academics to be harmed *via* this process is a continued risk and highlights the need for review and reform of SET systems and protocols.

KEYWORDS

Student feedback; higher education; teacher evaluation; academia; student evaluation of teaching (SET)

Introduction

In higher education institutions, most measures of success reflect the product of academic labour. Academic teaching staff are held accountable and evaluated against outcomes that include: students' academic success, graduates' job readiness and overall satisfaction measures incorporated into formally conducted anonymous student evaluation of teaching (SET) (Clayson, 2022). Overall, the SET requirement is governed by higher education quality control mechanisms (Australian Qualifications Framework, 2013; Tertiary Education Quality and Standards Agency, 2023), where regular monitoring is required as part of assurance frameworks.

After each teaching period, students are invited to evaluate the teaching performance of academic staff (Cook et al., 2022). Academics are evaluated three, four and sometimes six times a year depending on the structure and delivery at any university. Conducted under conditions of

anonymity, SET typically comprises statements requiring a ranking, usually on a Likert scale. These focus on curricula and assessment and overall satisfaction with the course or unit of study. Additionally, designed with an open-ended approach, students are invited to provide further commentary on satisfaction with teaching or the academic and make suggestions for future improvements. Anonymously provided written feedback may be constructive or non-constructive (Maslova et al., 2022). It may be abusive or potentially harm an academic's wellbeing and/or career prospects (Cunningham et al., 2022; Lakeman et al., 2022b). This article focuses solely on academic staff's reactions to anonymously received non-constructive student commentary.

SET have been utilised in higher education since the 1920s (Carpenter et al., 2020), where they are the most widely applied as a measure of effectiveness (Shao et al., 2007). Currently, in Australian universities, SET is focused on providing: (i) formative feedback to faculty aimed at quality control of teaching and curriculum; (ii) a summary measure of teaching effectiveness for promotion and tenure decisions and (iii) information to students for the selection of courses and teachers (Kember et al., 2002; Chen, 2023). Despite concerns about the reliability of SET it does continue to offer important information about student experience and appears to be a valid measure of teaching performance (Chen, 2023). Research suggests using student feedback to enhance and develop future course offerings promotes professional development.

Student success and SET ratings are widely known to be influenced by many factors outside the control of academics (Cook et al., 2022). Nevertheless, SET is routinely scrutinised in performance reviews and staff management, who may be required to explain negative student commentary or low scores (Heffernan, 2022). This situation places academics at greater risk of workplace-exacerbated stress, known to be associated with SET feedback about teaching (Lee et al., 2021).

While purported to be a quality measure, a sizeable body of research has problematised the assumption that SET measures teaching quality (Stroebe, 2020). SET has been found to favour white male academics and impact adversely and disproportionately on women and those from minority ethnic and racial backgrounds (Adams et al., 2022; Fan et al., 2019). Researchers have identified that particular groups of teaching academics are more vulnerable to receiving non-constructive commentary (Fan et al., 2019). A consistent finding in a recent literature review was that gender makes a difference to SET scores which are highly prejudiced against women (Heffernan, 2022). Boring et al. (2016) and Fan et al. (2019), with SET data from 23,000 and 22,000 students, respectively, found that male students express a significant bias in favour of male academics. Heffernan (2022) concludes that white, able-bodied, heterosexual men who are neither too old nor young (30–50 years) are the least impacted by bias in SET scores and actually benefit from the practice. This calls to question what is the situation for demographic groups that do not fit this image?

Others have argued that SET promotes individualism and competitiveness, which increases stress and negatively impacts the mental health of academics (Lakeman et al., 2022b; Smith et al., 2022). SET has become the prevailing instrument for scrutinising academic performance and the primary measure of teaching quality (Bedggood & Donovan, 2012). Pressure is placed upon academics to maintain above-average SET ratings, to compete with peers and elicit positive commentary that suggests satisfied students (Wang & Williamson, 2022). Otherwise, they may face being penalised in relation to promotion or their workload profile.

A potential consequence of the reliance on SET for decisions about tenure and promotion outcomes are the gaming of SET to ensure high ratings. 'Playing the SET game' (Lakeman et al., 2022b, appendix 1) can entail poor educational practices, grade inflation, the erosion of teaching standards and the oversimplification of teaching content (Carpenter et al., 2020). It has also been suggested that academics who are strict in their grading practices may be penalised and subjected to catastrophic ratings and abusive narrative commentary by students (Stroebe, 2020; Lakeman et al., 2022c).

Among academics, workplace stress has been reported as a predictor of poor emotional wellbeing and mental health (Shen & Slater, 2021). Across the higher education sector in the UK and Northern Ireland, academics reported that, along with sharp increases in occupational

health referrals, stress levels significantly increased from 2009 to 2016 (Morrish, 2019). Further, a little over 70% of academic staff described themselves as stressed and reported greater suicide risk than peers in other professions. The increase in stress was primarily attributed to the increasing pressure on academics to enhance the student experience. A survey with UK academics ($N=158$) using the Kessler Psychological Distress Scale (K10) found that 43% were distressed at levels predictive of at least a mild mental disorder (Gorczyński et al., 2017). However, in recent systematic reviews focusing on factors impacting occupational stress and mental health in academia, SET practices were not specifically identified as a risk factor (Lee et al., 2022; Ohadomere & Ogamba, 2020).

Little research has investigated the impact of receiving non-constructive student commentary. Understanding the nuances of the impact of SET on academics is essential, as the increasing utilisation of SET through anonymous online surveys may be causing disproportionate harm (Heffernan, 2022). In addition, the impact of anonymous commentary, particularly that which becomes non-constructive, has been shown to impact social relationships within the workplace and the general organisational climate (Lakeman et al., 2022a) and job satisfaction (Leung et al., 2000; Lee et al., 2022). Therefore, this research focuses on academic work roles and employment characteristics and how these may relate or even magnify the risk of harm stemming from receiving non-constructive SET commentary.

Methods and materials

Participants and procedure

The aim was to investigate the perspectives of academics currently receiving formal student evaluations of their teaching (SET). The focus was explicitly on reactions to non-constructive student commentary. Questions to be asked of the participants were decided upon using a literature-informed approach followed by discussion and team consensus.

A point-in-time survey was conducted to understand Australian academic viewpoints and perspectives on non-constructive SET. The research was a one-group design, nonparametric and the responses ordinal. The Australian Bureau of Statistics survey sample size calculator (Australian Bureau of Statistics, 2023) was applied to calculate the number of participant responses required for an Australia-wide survey. At the time of the research 137,055 academics were employed in the Australian higher education sector, comprising 95,500 full-time, 17,205 part-time and 24,350 casuals. Applying a 95% confidence interval, a conservative estimate of 50%, and a standard error of 0.03, the recommended sample size for the investigation was 384 participants. The collected sample of 741 Australian academics satisfied the recommended sample size.

Approval for this research was granted by the Southern Cross University Human Research Committee (2021/047). Questions were constructed to allow each participant to self-report information about themselves and then to respond to specific questions focused on retrospectively considered anonymous non-constructive student feedback. Demographic questions included: age, gender, ethnicity, university employment status and level, years in the sector and how much of the current role was evaluated using SET. Participants were asked to indicate whether they were from Australia, New Zealand, Europe, United Kingdom, Asia or Other. The survey was constructed and deployed online using Qualtrics software.

Information about the research was shared through university emails, snowballed through the authors' networks and disseminated with an article published in *The Conversation* (Lee et al., 2021). A link was also provided on social media platforms Facebook and Twitter. A participant information sheet and informed consent were included at the commencement of the survey, along with help numbers specific to academics if the survey brought up negative feelings.

Likert style and open-ended questions were included in the survey to measure academics' perspectives on constructive and non-constructive SET. Qualitative thematic analysis exploring the open-ended questions have been published elsewhere (Lakeman et al., 2022a, 2022b, 2022c). Questions reported on in this paper include: 'Anonymous narrative student feedback causes me stress', 'Anonymous narrative student feedback has impacted my mental health' and 'Anonymous narrative student feedback has undermined my professional confidence'. These questions were measured on a Likert scale (1 'never' to 5 'always'). Such single-item measures have previously demonstrated validity in measuring stress and coping among teachers (Morrish, 2019).

Data analysis

Data from Qualtrics were entered into SPSS version 28 statistical analysis software (SPSS Inc., Chicago, IL). Initially, each question was included as a single variable and tested for skewness, kurtosis and missing data. Descriptive and frequency analysis was conducted. A one-way analysis of variance (ANOVA) was performed to determine whether there was a statistically significant difference between the demographic and employment variables and the impact of anonymously received student commentary. Significance was accepted at $p \leq .05$. Tukey's HSD Test was examined to establish any significantly different means. Given that ANOVA is tolerant of moderate deviations of normality, with robustness increased in large sample sizes, these tests were suited to the data (Harwell et al., 1992).

A two-way ANOVA without repeated measures was conducted to test whether academics were equally impacted by anonymous student commentary. This analysis allowed for exploring the effects of two independent factors and the interaction between these factors. The independent variables tested were demographic and employment characteristics. The dependent variables tested were self-reported stress, undermining self-confidence and impact on mental health. This analysis focused the research question – for academics exposed to non-constructive student commentary, do employment and demographic characteristics interact to influence self-reported effects on stress, mental health and professional self-confidence? As an exploratory two-way ANOVA increases the likelihood of Type 1 error, a Bonferroni correction where one divides α by k was applied and found an adjusted significance of $p = .016$.

Results

The focus was on the reactions of higher education academics receiving non-constructive student commentary within teaching evaluations. The survey was open for 4 weeks. Data were examined for completeness and missing variables. $N = 741$ responses were included for analysis.

Demographics

The highest number of survey responders (63%) was in the 41–60-year age group, with 77% female, 73% tenured or fixed term and 97% less than 5 years in their current employment. Years in the sector ranged between five and greater than 30 years. Despite a lower response rate for the current level that an academic was employed at ($n = 629$), this item was included, with the majority being a lecturer (37%) or senior lecturer (24%). Mean age was 48.6 years and reflected well the profile of the Australian and North American academic workforce (May et al., 2013; McChesney & Bichsel, 2020). Looking at academic discipline, 67% were in health ($n = 473$), and 24% were in humanities and arts, education, law and criminology and business ($n = 172$), 9% were in science and mathematics ($n = 66$). Table 1 presents a summary of the characteristics of the respondents.

Analysis outcomes

Overall workload

Being casual, tenured or sessional was significantly associated with the amount or percentage of workload a staff member was undertaking ($\chi^2 (1, n=31), p<.001$). Casual/seasonal teaching staff were (41%, $n=63$) in the 81–100% teaching load bracket, compared to tenured staff working at a 36–50% teaching load (75%, $n=185$). These figures align with the teaching-focused workloads of casual academics (Evans et al., 2019), ultimately exposing casual and seasonal staff to a higher number of SET each year.

Gender

Analysis of the reactions to receiving non-constructive anonymous commentary revealed no significant differences between male and female academics for mental health ($p=.44$), stress ($p=.51$) or professional confidence ($p=.28$).

Age

Examining the age brackets of the academics providing self-report about their reaction to receiving anonymous commentary *via* SET, a significant impact was revealed for mental health $F(4) = 3.93, p<.001$, stress $F(3)=3.63, p<.001$ and professional confidence $F(3)=3.53, p<.001$. Tukey post hoc test ($\alpha=0.05$) showed that academics in the 20–40 age group reported significantly

Table 1. Demographic characteristics of Australian academics ($n=741$).

	<i>n</i>	%
Age		
20–40 years	181	24.4
41–60 years	465	62.8
61–81 years	95	12.8
Gender		
Male	157	21.2
Female	572	77.2
Non-binary	3	0.4
Prefer not to say	8	1
Self-describe	1	0.1
Employment status		
Tenured	454	61.3
Probation	89	13
Fixed term	86	11.6
Casual/sessional	112	15.1
Academic level		
Associate lecturer	37	5
Lecturer	276	37.2
Senior Lecturer	176	23.8
Associate Professor	76	10.3
Professor	64	8.6
Years in the sector		
<than 5 years	137	18.5
6–10 years	188	25.4
11–20 years	275	37.1
21–30 years	118	15.9
>30 years	23	3.1
Years at current employment		
<5 years	715	96.5
6–10 years	25	3.4
11–20 years	1	0.1

Table 2. Self-reported impact of anonymous SET according to demographic characteristics.

	Mental Health		Stress		Professional confidence	
	Mean (SD)	<i>p</i>	Mean (SD)	<i>p</i>	Mean (SD)	<i>p</i>
Gender		.439		.509		.280
Male	2.88 (1.29)		3.44 (1.17)		2.92 (1.20)	
Female	2.97 (1.22)		3.46 (1.10)		3.06 (1.12)	
Age in years		<.001*		<.001*		<.001*
20–40	3.26 (1.21)		3.72 (1.07)		3.29 (1.07)	
41–60	2.89 (1.24)		3.43 (1.02)		3.00 (1.15)	
61–80	2.71 (1.16)		3.01 (1.05)		2.74 (1.15)	
Ethnicity		.756		.499		.794
Australia/New Zealand	2.97 (1.24)		3.47 (1.09)		3.05 (1.12)	
Europe/United Kingdom	2.88 (1.21)		3.49 (1.11)		3.01 (1.15)	
Asia	3.00 (1.20)		3.57 (1.26)		3.05 (1.18)	
Other	2.74 (1.40)		3.11 (1.32)		2.79 (1.27)	
Employment status		.102		.044*		.081
Tenured	3.04 (1.19)		3.56 (1.05)		3.22 (1.09)	
Probation	2.98 (1.29)		3.40 (1.21)		2.97 (1.27)	
Fixed term contract	2.81 (1.34)		3.40 (1.24)		2.96 (1.18)	
Casual/sessional	2.74 (1.25)		3.22 (1.11)		2.82 (1.10)	
Percent teaching		.748		.384		.404
<25%	2.87 (1.24)		3.33 (1.10)		2.98 (1.16)	
26–50%	2.91 (1.23)		3.44 (1.13)		2.99 (1.17)	
51–80%	3.01 (1.23)		3.56 (1.09)		3.15 (1.08)	
>80%	2.99 (1.27)		3.47 (1.13)		2.99 (1.16)	

**p* < .05

higher stress associated with anonymous non-constructive commentary received *via* SET. A Tukey post hoc test ($\alpha=0.05$) also confirmed that when compared to casual/seasonal academics ($M= 3.26$, $SD= 1.09$), tenured staff reported significantly higher levels of stress ($M= 3.60$ $SD= 1.02$).

Years in the sector

Examining the impact of SET on stress, mental health and professional confidence according to years in the sector confirmed a significant effect on mental health $F(4)=2.39$, $p=.05$. A Tukey post hoc test ($\alpha=0.05$) showed no significant differences in any particular range of years in the sector group.

Differentiating interaction effects

The two-way ANOVA revealed a statistically significant interaction between the percentage of teaching and discipline on mental health and professional confidence $F(3) 3.79$, $p=.01$. Non-health disciplines with a teaching load greater than 50% reported a higher mean score for the impact of anonymous SET on mental health and professional confidence (see Table 3).

A statistically significant interaction between being casual or seasonal and ethnicity on professional confidence $F(3)=2.74$, $p=.004$. Participants who were from Europe, the United Kingdom, or Asia reported higher reduced professional confidence than those from Australia and New Zealand (see Table 4).

Discussion

To our knowledge, this is the first large scale study designed to quantify reactions to receiving non-constructive student responses collected in SET procedures. *Via* self-report, we investigated reactions to this narrative on academics' mental health, stress and professional confidence. Our findings highlight a potential for harm stemming from feedback that may be threatening, abusive or derogatory. Unlike Shen and Slater (2021), who reported no statistically significant

Table 3. Interaction effect of percent teaching and demographic variables on stress, mental health and levels of professional confidence.

	Stress					Mental health					Professional confidence			
	SS	df	MS	F	p	SS	df	MS	F	p	df	MS	F	p
Percent teaching	3.44	3	1.14	0.993	.396	2.63	3	0.879	0.609	.610	3	0.852	0.720	.540
Discipline	1.12	1	1.12	0.976	.324	0.434	1	0.434	0.300	.584	1	0.018	0.015	.902
Interaction	7.37	3	2.45	2.12	.095	12.38	3	4.12	2.85	.036	3	4.48	3.786	.010*
Percent teaching	2.95	3	.985	0.853	.465	5.565	3	1.85	1.27	.281	3	0.740	0.613	.607
Ethnicity	4.02	3	1.34	1.16	.324	5.44	3	1.81	1.25	.291	3	0.932	0.772	.510
Interaction	11.91	9	1.32	1.14	.328	17.47	9	1.94	1.33	.213	9	1.02	0.847	.573
Percent teaching	.049	3	.016	0.014	.998	.110	3	0.037	0.026	.994	3	0.404	0.341	.796
Age	12.78	3	4.26	3.74	.011	19.11	3	6.37	4.43	.004	3	3.64	3.07	.027
Interaction	13.35	6	2.22	1.95	.070	5.62	6	0.937	0.652	.689	6	0.941	0.794	.575
Percent teaching	6.598	3	2.199	1.889	.130	4.624	3	1.541	1.053	.369	3	0.830	0.695	.556
Gender	5.000	4	1.250	1.074	.369	5.66	4	1.416	0.967	.425	4	1.160	0.971	.423
Interaction	6.053	6	1.009	0.866	.519	4.81	6	0.803	0.548	.772	6	1.673	1.400	.212

Table 4. Interaction effect of tenure and demographic variables on stress, mental health and levels of professional confidence.

	Stress					Mental health					Professional confidence				
	SS	df	MS	F	p	SS	df	MS	F	p	SS	df	MS	F	p
Tenure	7.48	3	2.49	2.15	.092	4.95	3	1.65	1.14	.330	8.44	3	2.81	2.37	.069
Discipline	3.69	1	3.69	3.19	.074	0.234	1	0.234	0.163	.687	.289	1	0.289	0.244	.622
Interaction	3.72	3	1.24	1.07	.359	3.93	3	1.31	0.910	.436	2.76	3	0.921	0.778	.507
Tenure	.79	3	0.265	0.235	.872	9.20	3	3.06	2.16	.091	1.56	3	0.522	0.451	.717
Ethnicity	5.07	3	1.69	1.49	.214	18.66	7	2.66	1.88	.070	5.87	3	1.95	1.69	.167
Interaction	25.7	9	2.85	2.52	.007	34.30	16	2.14	1.51	.089	28.5	9	3.17	2.74	.004*
Tenure	5.49	3	1.830	1.61	.185	7.29	3	2.43	1.73	.159	5.52	3	1.84	1.58	.192
Age	16.9	3	5.661	4.98	.002	23.88	3	7.96	5.66	<.001	16.3	3	5.46	4.70	.003
Interaction	8.72	6	1.455	1.28	.264	13.80	6	2.30	1.63	.134	7.84	6	1.30	1.12	.346

relations between academic socio-demographic variables and reported levels of emotional wellbeing or stress, our study found several personal and employment characteristics associated with harmful effects from the receiving of non-constructive commentary within formal student evaluations of teaching.

Age was identified as a significant variable in our study, with younger respondents reporting a higher impact of anonymous SET on stress. This finding may result from early career academics being more likely to be under probation or have SET scores scrutinised to maintain tenure or promotion than older teaching staff (Kreitzer & Sweet-Cushman, 2022).

Teaching allocation was also identified as significant, with respondents with a teaching load greater than 50% reporting increased stress and impact on mental health as an impact of anonymous feedback in SET. As the academic workforce continues to rely more on a casualised workforce (Evans et al., 2019) and the employment of teaching-only academics increases, the level of stress experienced by academics will likely continue to rise (Lee et al., 2022).

Increasing evidence confirms that anonymised SET is biased by discipline, subject area and racist, gendered or homophobic prejudices (Heffernan, 2022). Within our sample, the younger and tenured academics were most impacted by the harmful effects of non-constructive anonymised commentary. Teaching load and tenure interacted with an academic's cultural identity to magnify the harm caused. The negative interaction effect identified between tenure, ethnicity and professional confidence among academics who reported receiving non-constructive anonymised SET provides further evidence that ethnically diverse academics may feel particularly marginalised during the SET process. These findings suggest that, beyond being biased, the student feedback may further marginalise under-represented academics, the same group universities seek to increase in their workforce (Smith et al., 2015).

Although the proportion of female academics in this study who reported receiving non-constructive anonymised SET was three times that of male academics, we found no significant differences in the effect of SET on wellbeing according to gender. Additionally, the higher proportion of female academics who contributed to our research is reflective of the gender demographics of health and allied disciplines (Australian Research Council, 2019).

Anonymised student feedback processes are a source of academic anxiety and highlight a concern for a safe workplace and fairness in promotion and tenure processes (Lakeman et al., 2022b). Ultimately, these findings raise questions about how the continued anonymity of SET can be justified as a quality assurance activity and provides an important platform to explore other means of measuring student satisfaction. Now firmly embedded within routine administrative systems, anonymised SET has become an administrative tool that exerts managerial surveillance and control over academics rather than being a reliable or valid quality measure (Stroebe, 2020; Clayson, 2022; Heffernan, 2022; Lakeman et al., 2022d). While universities may perceive that the rate of abusive commentary is relatively low, it is clear that the harm to wellbeing may not be minimal (Lakeman et al., 2022c). The universities continuing to use anonymised SET may contribute to a psychologically unsafe work environment.

Strengths and limitations

A strength of this research was the large sample size. Many academics were keen to contribute their experiences and representation across multiple Australian universities covering regional and metropolitan areas. There were also limitations to be acknowledged. Firstly, selection bias was inherent in the sampling strategy. Most of the academics surveyed were predominantly female in health-care-related fields, which could lead to unequal distribution and a lack of generalisability in other academic populations. Additionally, academics who had negative experiences with feedback and, therefore, may have felt strongly about what they had received may have been more likely to participate, thus producing a selection bias. Nevertheless, based on our findings, it is now appropriate to challenge both the premise and outcomes of student feedback practices that engage an anonymous narrative to evaluate teaching performance. Future research could extend these findings by focusing on developing valid measures that capture student perceptions and experiences within academic domains.

Anonymous narrative feedback is a well-established formal process across higher education institutions in Australia. The findings from this study highlight the negative impacts of the current use and application of anonymous non-constructive commentary on academics in Australian universities. Albeit using self-report, this research has shown a clear association between stress, mental health and professional confidence of a broad group of academics. Whilst acknowledging that student feedback on teaching is important, we suggest that the primary purpose of student feedback is not being realised, and reform is required.

Acknowledgement

We acknowledge the assistance with statistical analysis and the interpretive contribution provided by Dr Alison Bowling.

Ethical approval

Approval for this research was granted by the Southern Cross University Human Research Committee (2021/047).

Informed consent

Informed consent was received from all participants commenced the online survey.

Disclosure statement

The Authors declare that there is no conflict of interest.

Funding

This research received no grant from any funding agency in the public, commercial or not-for-profit sectors.

Data availability

Data from this study are available upon request from the lead researcher.

ORCID

Marie Hutchinson  <http://orcid.org/0000-0003-0025-2681>
 Rosanne Coutts  <http://orcid.org/0000-0001-5064-8903>
 Debbie Massey  <http://orcid.org/0000-0002-0466-1960>
 Dima Nasrawi  <http://orcid.org/0000-0002-2878-2468>
 Jann Fielden  <http://orcid.org/0000-0003-0084-0838>
 Megan Lee  <http://orcid.org/0000-0001-7531-5312>
 Richard Lakeman  <http://orcid.org/0000-0002-4304-5431>

References

- Adams, S., S. Bekker, Y. Fan, T. Gordon, L. J. Shepherd, E. Slavich, and D. Waters. 2022. "Gender Bias in Student Evaluations of Teaching: 'Punish[Ing] Those Who Fail to Do Their Gender Right.'" *Higher Education* 83 (4): 787–807. doi:10.1007/s10734-021-00704-9.
- Australian Bureau of Statistics. 2023. Sample size calculator. <https://www.abs.gov.au/websitedbs/d3310114.nsf/home/sample+size+calculator>
- Australian Qualifications Framework. 2013. Australian Qualifications Framework (AQF). Australian Qualifications Framework Council. <https://www.aqf.edu.au/framework/australian-qualifications-framework>
- Australian Research Council. 2019. State of Australian University research 2018–19 ERA national report. Australian Research Council. <https://dataportal.arc.gov.au/era/nationalreport/2018/>
- Bedggood, R. E., and J. D. Donovan. 2012. "University Performance Evaluations: What are we Really Measuring?" *Studies in Higher Education* 37 (7): 825–842. doi:10.1080/03075079.2010.549221.
- Boring, A., K. Ottoboni, and P. Stark. 2016. "Student Evaluations of Teaching (Mostly) Do Not Measure Teaching Effectiveness." *ScienceOpen Research*. doi:10.14293/S2199-1006.1.SOR-EDU.AETBZC.v1.
- Carpenter, S. K., A. E. Witherby, and S. K. Tauber. 2020. "On Students' (Mis)Judgments of Learning and Teaching Effectiveness." *Journal of Applied Research in Memory and Cognition* 9 (2): 137–151. doi:10.1016/j.jar-mac.2019.12.009.
- Chen, Y. 2023. "Does Students' Evaluation of Teaching Improve Teaching Quality? Improvement versus the Reversal Effect." *Assessment & Evaluation in Higher Education*. doi:10.1080/02602938.2023.2177252.
- Clayson, D. 2022. "The Student Evaluation of Teaching and Likability: What the Evaluations Actually Measure." *Assessment & Evaluation in Higher Education* 47 (2): 313–326. doi:10.1080/02602938.2021.1909702.
- Cook, C., J. Jones, and A. Al-Twal. 2022. "Validity and Fairness of Utilising Student Evaluation of Teaching (SET) as a Primary Performance Measure." *Journal of Further and Higher Education* 46 (2): 172–184. doi:10.1080/0309877X.2021.1895093.
- Cunningham, S., M. Laundon, A. Cathcart, M. A. Bashar, and R. Nayak. 2022. "First, Do No Harm: Automated Detection of Abusive Comments in Student Evaluation of Teaching Surveys." *Assessment & Evaluation in Higher Education*. doi:10.1080/02602938.2022.2081668.
- Evans, J., S. Ruane, H. Southall, N. Dados, J. Goodman, and K. Yasukawa. 2019. "Counting the Uncounted: Contestations over Casualisation Data in Australian Universities." In *Data in Society: Challenging Statistics in an*

- Age of Globalisation*, edited by J. Evans, S. Ruane, and S. Humphrey, 327–336. Bristol, England: Policy Press. doi:10.51952/9781447348245.
- 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. doi:10.1371/journal.pone.0209749.
- Gorczynski, P., D. Hill, and S. Rathod. 2017. "Examining the Construct Validity of the Transtheoretical Model to Structure Workplace Physical Activity Interventions to Improve Mental Health in Academic Staff." *EMS Community Medicine Journal* 1 (1): 002. <https://researchportal.port.ac.uk/en/publications/examining-the-construct-validity-of-the-transtheoretical-model-to>.
- Harwell, M. R., E. N. Rubinstein, W. S. Hayes, and C. C. Olds. 1992. "Summarizing Monte Carlo Results in Methodological Research: The One-and Two-Factor Fixed Effects ANOVA Cases." *Journal of Educational Statistics* 17 (4): 315–339. doi:10.2307/1165127.
- Heffernan, T. 2022. "Sexism, Racism, Prejudice, and Bias: A Literature Review and Synthesis of Research Surrounding Student Evaluations of Courses and Teaching." *Assessment & Evaluation in Higher Education* 47 (1): 144–154. doi:10.1080/02602938.2021.1888075.
- Kember, D., D. Y. P. Leung, and K. P. Kwan. 2002. "Does the Use of Student Feedback Questionnaires Improve the Overall Quality of Teaching?" *Assessment & Evaluation in Higher Education* 27 (5): 411–425. doi:10.1080/0260293022000009294.
- Kreitzer, R. J., and J. Sweet-Cushman. 2022. "Evaluating Student Evaluations of Teaching: A Review of Measurement and Equity Bias in SETs and Recommendations for Ethical Reform." *Journal of Academic Ethics* 20 (1): 73–84. doi:10.1007/s10805-021-09400-w.
- Lakeman, R., R. Coutts, M. Hutchinson, D. Massey, D. Nasrawi, J. Fielden, and M. Lee. 2022a. "Playing the SET Game: How Teachers View the Impact of Student Evaluation on the Experience of Teaching and Learning." *Assessment & Evaluation in Higher Education*. doi:10.1080/02602938.2022.2126430.
- Lakeman, R., R. Coutts, M. Hutchinson, D. Massey, D. Nasrawi, J. Fielden, and M. Lee. 2022b. "Stress, Distress, Disorder and Coping: The Impact of Anonymous Student Evaluation of Teaching on the Health of Higher Education Teachers." *Assessment & Evaluation in Higher Education* 47 (8): 1489–1500. doi:10.1080/02602938.2022.2060936.
- Lakeman, R., R. Coutts, M. Hutchinson, M. Lee, D. Massey, D. Nasrawi, and J. Fielden. 2022c. "Appearance, Insults, Allegations, Blame and Threats: An Analysis of Anonymous Non-Constructive Student Evaluation of Teaching in Australia." *Assessment & Evaluation in Higher Education* 47 (8): 1245–1258. doi:10.1080/02602938.2021.2012643.
- Lakeman, R., D. Massey, D. Nasrawi, J. Fielden, M. Hutchinson, M. Lee, and R. Coutts. 2022d. 'Lose some weight', 'stupid old hag': Universities should no longer ask students for anonymous feedback on their teachers. <https://theconversation.com/lose-some-weight-stupid-old-hag-universities-should-no-longer-ask-students-for-anonymous-feedback-on-their-teachers-173911>
- Lee, M., R. Coutts, J. Fielden, M. Hutchinson, R. Lakeman, B. Mathisen, D. Nasrawi, and N. Phillips. 2022. "Occupational Stress in University Academics in Australia and New Zealand." *Journal of Higher Education Policy and Management* 44 (1): 57–71. doi:10.1080/1360080X.2021.1934246.
- Lee, M., D. Nasrawi, M. Hutchinson, and R. Lakeman. 2021. Our uni teachers were already among the world's most stressed. COVID and student feedback have just made things worse. <https://theconversation.com/our-uni-teachers-were-already-among-the-worlds-most-stressed-covid-and-student-feedback-have-just-made-things-worse-162612>
- Leung, T., O. Siu, and P. E. Spector. 2000. "Faculty Stressors, Job Satisfaction, and Psychological Distress among University Teachers in Hong Kong: The Role of Locus of Control." *International Journal of Stress Management* 7 (2): 121–138. doi:10.1023/A:1009584202196.
- Maslova, A., O. Koval, V. Kotliarova, M. Tkach, and Y. Nadolska. 2022. "On the Way to Successful Learning and Teaching: Constructive Feedback." *Journal of Higher Education Theory and Practice* 22 (6): 113–122. <https://ezproxy.scu.edu.au/login?url=https://www.proquest.com/scholarly-journals/on-way-successful-learning-teaching-constructive/docview/2689719449/se-2?accountid=16926>.
- May, R., G. Strachan, and D. Peetz. 2013. "Workforce Development and Renewal in Australian Universities and the Management of Casual Academic Staff." *Journal of University Teaching & Learning Practice* 10 (3): 1–26. doi:10.53761/1.10.3.3.
- McChesney, J., and J. Bichsel. 2020. The aging of tenure-track faculty in higher ed: Implications for succession and diversity. <https://www.cupahr.org/surveys/research-briefs/2020-aging-of-tenure-track-faculty-in-higher-ed-implications-for-succession-diversity/>
- Morrish, L. 2019. *Pressure Vessels: The Epidemic of Poor Mental Health among Higher Education Staff*. England: Higher Education Policy Institute Oxford. <https://healthyuniversities.ac.uk/wp-content/uploads/2019/05/HEPI-Pressure-Vessels-Occasional-Paper-20.pdf>.
- Ohadomere, O., and I. K. Ogamba. 2020. "Management-Led Interventions for Workplace Stress and Mental Health of Academic Staff in Higher Education: A Systematic Review." *The Journal of Mental Health Training, Education and Practice* ahead of print. doi:https://doi.org/10.1108/JMHTPE-07-2020-0048.

- Shao, L. P., L. P. Anderson, and M. Newsome. 2007. "Evaluating Teaching Effectiveness: Where we Are and Where we Should Be." *Assessment & Evaluation in Higher Education* 32 (3): 355–371. doi:10.1080/02602930600801886.
- Shen, P., and P. Slater. 2021. "The Effect of Occupational Tress and Coping Strategies on Mental Health and Emotional Well-Being among University Academic Staff during the COVID-19 Outbreak." *International Education Studies* 14 (3): 82–95. <http://files.eric.ed.gov/fulltext/EJ1287926.pdf>. doi:10.5539/ies.v14n3p82.
- Smith, J., S. Trinidad, and S. Larkin, Charles Darwin University. 2015. "Participation in Higher Education in Australia among Under-Represented Groups: What Can we Learn from the Higher Education Participation Program to Better Support Indigenous Learners?" *Learning Communities: International Journal of Learning in Social Contexts* 17: 12–29. doi:10.18793/LCJ2015.17.02.
- Smith, J. M., J. Smith, A. McLuckie, A. C. Szeto, P. Choate, L. K. Birks, V. F. Burns, and K. S. Bright. 2022. "Exploring Mental Health and Well-Being among University Faculty Members: A Qualitative Study." *Journal of Psychosocial Nursing and Mental Health Services* 60: 17–25. doi:10.3928/02793695-20220523-01.
- Stroebe, W. 2020. "Student Evaluations of Teaching Encourages Poor Teaching and Contributes to Grade Inflation: A Theoretical and Empirical Analysis." *Basic and Applied Social Psychology* 42 (4): 276–294. doi:10.1080/01973533.2020.1756817.
- Tertiary Education Quality and Standards Agency. 2023. Tertiary education quality and standards agency (TEQSA). <https://www.teqsa.gov.au/>
- Wang, G., and A. Williamson. 2022. "Course Evaluation Scores: Valid Measures for Teaching Effectiveness or Rewards for Lenient Grading?" *Teaching in Higher Education* 27 (3): 297–318. doi:10.1080/13562517.2020.1722992.