Check for
updates

Abusive comments in student evaluations of courses and teaching: the attacks women and marginalised academics endure

Troy Heffernan[1] (ID)

Abstract

This paper examines the volume and type of anonymous comments academics receive in student evaluations of courses and teaching (SETs) at the 16,000 higher education institutions that collect this data at the end of each teaching period. Existing research has increasingly pointed to the negative issues of student surveys, but very little research has focused on the volume, type, and impact of anonymous student comments on academics. This paper analyses the survey results of 674 academics to inform higher education leaders and the sector more widely of the amount and type of abusive comments academics are receiving. The work also demonstrates that the highest volume, most derogatory, and most threatening abuse is directed towards women academics and those academics from marginalised groups. The paper finds that previous estimates of the rate and severity of abusive comments that academics receive, and the impact to academics' wellbeing, mental health, and career progression, have underestimated what is taking place. The paper argues that many universities are failing to protect their staff from this abuse, and the prejudice nature of SET results, which will continue to have a negative impact on the career progression of marginalised academics - a major flaw in a sector that prides itself on diversity and inclusion.

Keywords Higher education · Abuse · Social justice · Equity · Diversity · Marginalised

Introduction

This paper evaluates abusive student comments towards academics in student evaluations of courses and teaching (abbreviated to the standard term SETs within this paper). In part, the paper highlights a problem that is known to exist by some universities and has caused some procedural changes to be made. For example, Tucker (2014) is one of the few researchers to question an institution as to why they allowed anonymous student evaluations to continue when they knew academics were receiving abusive comments. The

✉ Troy Heffernan
    t.heffernan@latrobe.edu.au

1   La Trobe University, Plenty Rd & Kingsbury Dr, Bundoora, VIC 3086, Australia

🌀 Springer

university's response was that they believed there was a low percentage of abuse, and the abuse was worth the [perceived] value of the data student evaluations generated (Tucker, 2014). Conversely, some of this study's participants (142 of 674 participants, or 21%) stated that their institutions or faculties filter or censor student evaluation comments before they received them. Other participants noted that their university/faculty had trialled filtering student evaluations (which indicates that they knew there was a problem), but have since abandoned the process due to the labour and/or expense involved. This is evidence that some universities accept that abusive comments are a problem, and that some are actively trying to prevent staff from their negative impacts.

These examples (and this study's data) suggest universities know a problem exists. However, most institutions make decisions that clearly prioritise the perceived value of data gained from SETs above their academics' wellbeing and their right to work in an abuse-free environment. This paper approaches these topics from a perspective that hopes that in most cases universities do not appreciate how common the abuse in SET comments is, and what types and the volume of comments to which academics are being subjected. It is hoped these institutions will be receptive in calls for SETs to be removed. In other cases, however, existing research and this study make clear that some universities know what is occurring but do not care. They likely hope the findings of papers such as this do not gain traction because of the customer satisfaction data student evaluations generate and the extra time and labour it would require to gain this data via methods such as focus groups, student interviews, or peer reviews of teaching (Enache, 2011; Osoian et al., 2010; Sunindijo, 2016; Vivanti et al., 2014).

This paper is thus not apologetic in what it discusses or what it aims to do. This paper's findings make it clear that academics are routinely subjected to abuse in student evaluations. If any institution, and indeed the university sector, wish to declare themselves as being inclusive or interested in diversity, anonymous comments in student evaluations must be removed. The paper highlights that we already know that SET results are prejudice towards marginalised academics (Fan et al., 2019; Heffernan, 2021a), and that the data produced even when not abusive is likely to portray men as professions and experts in the field while women are caregivers who a nice but less likely viewed as knowledgeable (Schmidt, 2015; 2022). Additionally, this work confirms that the abusive SET comments are directed towards the marginalised groups within the sector based on gender, sexual identity, ethnicity, language, appearance, or disability.

SETs are causing academics stress and anxiety in the workplace, are preventing some academics, usually women and those from marginalised backgrounds, from seeking promotion, and this is occurring in a sector that prides itself on diverse and inclusive practices. The paper highlights to university administrators and policymakers which demographics of academics are receiving the most abuse, and provides examples of what some of the abusive comments have included. It should be noted that the examples within this paper **are not censored**. This decision was made to highlight how severe some comments are because it is these statements that will either prompt universities to remove anonymous SET comments, or else it is these comments to which universities will have to justify knowingly exposing their staff.

This paper is aimed at improving the academy for every academic in the 16,000 higher education institutions that collect SET data at the end of each teaching period (Cunningham-Nelson et al., 2019). It is unacceptable to continue practices that allow these groups to face abuse which causes anxiety, stress, other psychological discomfort, and impacts on employment, promotion, and leadership prospects. This is particularly the case when those facing the abuse are those already in marginalised and underrepresented groups. Finally, if

the sector is not motivated by the wellbeing of their staff, it is worth noting that researchers such as Jones et al. (2014) and Mitchell and Martin (2018) have suggested universities may be open to legal challenges for allowing their staff to knowingly be subjected to abuse.

Literature review

The existing literature around SETs is rarely favourable, and while this paper is primarily concerned with abusive comments, it is necessary to note that this research contributes to existing work outlining the problematic nature of student evaluations (which is part of this literature review). It is also necessary to define the term 'abuse', or rather, point out that no clear definition as it stands in SET research is present. This paper follows the trends set by the literature below. That is, when it comes to negative comments directed towards the teaching academic, they are either unhelpful/unprofessional, or abusive. Abusive comments are thus defined as those which include offensive words or language, or else are directed at the academic's demographics or characteristics such as gender, race, sexuality (Tucker, 2014).

Why universities collect SET data

A starting point when discussing SETs and the data they generate is an acknowledgement that universities place ample faith in SET data because they believe the data is sound (Marsh, 2007; Osoian et al., 2010; Stark & Freishtat, 2014). This is because methodologically the system appears to be one that would generate valuable data. SET data is generated by students providing anonymous feedback (to remove concerns of negative repercussions from staff if the student's feedback is not positive), and this information provides the university and faculty with an assessment of class content and an academic's teaching performance. It is then often expected that academics use this information to help shape their classes by better providing what students feel the class needs or how they could facilitate learning better as a teacher (Osoian et al., 2010; Tucker, 2014). Universities, and indeed the sector as a whole, are so confident of this system that SET results are then often used as a component of hiring and promotional decisions, but also in times of restructuring and ceasing an academic's employment (Arthur, 2009; Boring et al., 2016; Jones et al., 2014; Shah & Nair, 2012).

The problem these researchers highlight is that the methodology SETs use is inherently flawed because the information input into the evaluation that results in the data is influenced by student biases and prejudices that are invisible in the data (Marsh, 2007; Osoian et al., 2010; Stark & Freishtat, 2014). For decades, researchers have thus found that statistically significant biases impacting on women and academics from marginalised groups are present in the SET results which universities have assumed were providing a somewhat objective assessment of courses and teaching (Tucker, 2014).

Research about abuse

Research around abuse in SETs is often similar in its findings. For example, Uttl and Smibert's (2017) study of 325,000 SETs, and Jones et al's (2014) analysis of the risks of treating

SET data as objective data, concluded that abusive comments were present in SETs, and that the comments were primarily directed towards women and marginalised groups.

One of the few major studies to focus strongly on abusive comments is Tucker's (2014) analysis of 43,000 SETs where she found that abusive comments were mostly directed towards women and marginalised groups, but it was Tucker's approach to the university in which the study was conducted that is most relevant to this paper. The university's response to why they were satisfied to allow SET comments to continue was because they deemed the rate of abusive comments (that they claimed was as low as 1% using keyword/phrase searches), as being too low to change procedures and lose the data generated from student evaluations (Tucker, 2014). Tucker nonetheless concluded from her research across several other studies, and anecdotal evidence as she conducted her primary study, that abuse was significantly more common, growing at a rapid pace, and noted that abusive comments can be made without abusive words picked up by keywords or phrase searches.

Fan et al.'s (2019) study of 22,000 SETs concluded that the mass-market higher education system has led to upwards of 30–40% of school-aged children now going on to further education in several major geographic areas such as Australia, North America, and the United Kingdom. Thus, a growing number of society's less progressive views are now entering higher education. The result is that as abuse in SETs rise, it is also being directed towards conservative criticisms of issues including but not limited to gender, race, and sexuality.

The impact of SET result prejudices and bias

The largest group negatively impacted in SET results is almost universally accepted to be women, with many studies finding that women are disadvantaged by a statistically significant amount (Adams et al., 2021; Valencia, 2021). Boring et al. (2016) found that higher performing women academics were often graded harsher in SETs to such an extent that they were placed in lower grading brackets (excellent, very good, good etc.) than lower performing males. Fan et al. (2019) determined that for women in subjects usually dominated by men in terms of teaching staff and students (e.g. science and maths), women were at risk of receiving SET results up to 37 percentage points lower than their male counterparts. Several researchers have also determined that correlation can often be found between SET results and an academics' age, and even perceived attractiveness (Boring et al., 2016: Felton et al., 2004, 2008).

SET results have been found to be significantly worse for any academic from an ethnically diverse group or those from other marginalised backgrounds (Hendrix, 1998). Several studies have also found that SET scores, and particularly negative scores, are accumulative in their damage. Thus, a woman will receive lower scores than a man, but a woman from a diverse background will receive lower scores than a white woman (Fan et al., 2019). Researchers examining SET prejudice towards academics who are visibly disabled, or whose students perceive to not be heterosexual or are not a binary gender, have found similar results with significant biases evident in the SET results. However, these researchers also note that it is difficult to conduct research surrounding these academic groups as they are so underrepresented in the academy (DiPietro & Faye, 2005; Hamermesh & Parker, 2005; Rubin, 1998).

Studies have also found high correlation between student grades (both perceived due to GPAs or mid-term results received) and SET results which makes it clear that student evaluations are being influenced by factors exceeding the intended purpose of SET questions (Boring et al., 2016, Short et al., 2008; Stark & Freishtat, 2014; Worthington, 2002).

Studies have found that factors including classroom design, cleanliness of the university, quality of course websites, library services, and food options available on campus have all been found to have a larger influence in SET results than practices concerning courses and teaching (Benton et al. 2012; Osoian et al., 2010).

The value of SET data

Many researchers argue that SETs are clearly not measuring course quality or teacher effectiveness (as is made clear by the above sections of this literature review). Researchers argue that what SETs are measuring is customer satisfaction (Brandl et al., 2017; Enache, 2011; Osoian et al., 2010). Researchers note that this is valuable information in the current higher education climate because student enrolments do matter. However, Sunindijo (2016) and Vivanti et al. (2014) argue that SETs are not needed because universities already know what students desire in a class. They suggest students want engaging lectures, having assessment be explained clearly and graded fairly, and to be taught by interested lecturers who provide guidance in the course. In both of these instances, the researchers conclude that SETs are an ineffective measure because of the prejudices involved, and suggest focus groups or student interviews would be a more equitable way of gaining information and judging potential biases and prejudice. It can also not be ignored that the growth in student voice, and students as partners initiatives in recent years, would be well placed to provide discussions to improve learning opportunities rather than relying on the current prejudice practices.

Methodology

This paper has selected qualitative data collection and analysis methods as it allows the paper to, as Creswell (2013) suggests, take a systematic approach to the human element and lived experiences, in this case, of academics on the receiving end of abusive comments in SETs. As Shulman (1988) noted several decades ago, it is the focus of the research questions needing to be answered that dictates the methods that should be used. This suggestion led to Denzin and Lincoln's (2011) conclusion that it is the researcher's duty to ensure they have considered the full array of research methods available to them, and Miles et al.'s (2013) suggestion that the complexity of qualitative data design should not be evident in the research design, questions, interactions with participants, or a study's findings.

The paper's research focus and design is thus based on three primary research questions:

RQ1: Which groups receive the highest rate of abusive SET comments?
RQ2: What types of comments are academics receiving?
RQ3: What are the impacts of abusive comments on the academics who receive them?

After the study was designed and methodologies selected, institutional ethical approval was gained before participant recruitment began. Participants were sourced through open calls for participation via my social media accounts, and recruitment then relied on snowball sampling. While I acknowledge the potential issues of sourcing data via social media, as researchers have noted, it is a method that can source a number of participants that would otherwise be difficult to achieve without significant funds, time, and resources (Kosinski 2015; Kosinski et al., 2015). In this study, the use of online

recruitment and snowball sampling led to 674 participants completing the survey, and resulted in over 60,000 words of raw data being generated.

The survey consisted of twenty questions in total, and was divided into subsections including:

- Five career demographic questions such as What academic level are you?, What is your field area? What country is your institution location within?.
- Six individual demographic questions such as What gender do you identify with?, What is your sexual identity?, How old are you?.
- Five questions relating to the institutional SET practices such as Are SETs collected?, Are SETs used for hiring/promotion purposes?, Do SETs impact on your decision to apply for promotion or leadership roles?.
- Four questions relating to individual experiences such as Please list some examples of negative student feedback. What changes would you make to the evaluation process?

Despite the potential issues of working with open-ended questions such as participants not understanding the questions, or attempting to predict the answers the researcher is seeking, I have chosen this method as it is often noted (Punch, 2013) as an effective way of gaining knowledge relating to lived experiences. As the survey was formed to source information relating to specific aspects of a dedicated topic, Punch (2013) notes that in these situations, open-ended questions in surveys are a succinct method of sourcing data relating to lived experiences.

Though this paper is dedicated to examining the types of abuse academics are receiving, the researcher acknowledges that bias is inherent in self-reporting data. It is more likely that the survey would have been of higher interest to someone with pre-conceived thoughts about SETs and thus, would be more likely to complete the survey. However, this limitation does not alter the examples of abuse academics are receiving, or the impact of bias SET data, to their wellbeing, mental health, and career prospects.

A significant portion of the *Findings* and *Discussion* sections are dedicated to exploring the demographics of the participants. Therefore, while it is customary to provide some overall data in the methodology section, in this paper, much of this information is included and discussed in later sections. The one exception to this may be in terms of the location of the surveyed academics; Table 1 therefore indicates where the participants institutions are located.

Table 1 Participants' institutional location

Location	Number	Percent
Australia/New Zealand	458	68
United Kingdom	49	8
North America	94	14
South America	9	1
Africa	9	1
Europe	40	6
Asia	15	2
Total	674	100

The study includes insights from several locations, however, this data is provided as supplementary information. This paper seeks to understand the volume and type of abuse academics are receiving; it is not concerned with locations for comparative purposes.

Braun and Clarke's (2006) multi-stage thematic analysis was used to analyse the data. This method was selected as when analysing qualitative data, Braun and Clarke's method provides a tool by which patterns and statistical information can be sourced from participants' experiences and perspectives (Clarke & Braun, 2017). In this study, this meant examining the open-ended question responses for words and phrasing relating to positive/ negative experiences with SET comments. As the data began to be coded, further rounds of reading and coding resulted in codes being sub-divided as they become more nuanced, and it is these final themes that are discussed in the following Findings and Discussion section.

Positionality

This paper's author identifies as a white, middle-class, heterosexual male. They are also blind. They are aware they only faced one 'closed gate' to acceptance into the academy, and they now research to ensure all the once closed gates to higher education are open for marginalised academics to enter and succeed in the sector.

Findings

Who receives abusive comments?

The study of 674 academics shows that in the early-2020s, 59 per cent of participants report receiving abusive comments in their student evaluations. Tables 2, 3, 4 and 5 demonstrate the rates of abuse across several demographic areas of academics working in the sector.

Tables 2, 3, 4 and 5 makes several points clear. The first is that abusive comments are much more common than previous studies have found or suggested. This on its own should not come as a surprise as Tucker (2014) highlights that large-scale studies of SET comments tend to rely on data-mining techniques usually built around keyword or phrase searches. It is, however, possible to be abusive without using words likely to be picked up via an automated system; an aspect made clear in this study where participants have provided examples of comments that they felt were abusive. It should also be noted that existing studies tend to focus on set time periods such as data collected over one teaching period or one year (Fan et al., 2019; Uttl & Smibert, 2017). In this study, I asked

Table 2 Abusive comments by gender

Gender	Participants indicating abuse	Rate of abuse
Men	93 of 178	52%
Women	301 of 476	63%
Non-binary	9 of 15	60%
Prefer not to say	3 of 3	100%
Not listed	2 of 2	100%
Total	408 of 674	

Table 3 Abusive comments by sexual identity

Sexual identity	Participants indicating abuse	Rate of abuse
Straight Men	70 of 127	55% yes
Gay	24 of 37	64% yes
Straight women	227 of 379	60% yes
Lesbian	17 of 21	83% yes
BTQIA + Men	0 of 0	0% yes
BTQIA + Women	12 of 17	71% yes
Women Sexual Identity Other	39 of 65	60% yes
Men Sexual Identity Other	13 of 28	46% yes
Total	402 of 674	

Table 4 Abuse via disciplinary area

Disciplinary area	Participants indicating abuse	Rate of abuse
Humanities	88 of 146	60%
Social Sciences	102 of 170	60%
Business	14 of 23	61%
Science	7 of 11	64%
Law	42 of 71	60%
Health	34 of 136	25%
Fine arts	11 of 15	73%
Engineering	63 of 102	62%
Total	361 of 674	

Table 5 Abuse at career stage

Career stage	Participants indicating abuse	Rate of abuse
Early (pre-PhD to 5 years Post PhD	175 of 336	51%
Mid (5–10 years post-PhD)	186 of 273	68%
Late (10 years beyond PhD)	44 of 65	67%
Total	405 of 674	

participants 'Have you ever received comments that contained abusive, offensive, or derogatory words or phrasing?' While 59% of participants answered 'Yes', I chose to not set a time period regarding whether or not the abuse occurred, for example, in the last teaching period or year, so as to gain an idea of the frequency of abuse. As the next sections explore, the participants' survey responses discuss the regularity with which they receive abusive comments.

It would also be dismissive to suggest that the impact of abusive comments is directly related to volume or regularity of abuse. For example, 67% of participants who reported receiving abusive comments indicated that the stress or anxiety of receiving abusive comments, or how the comments and SET results were connected to promotion

and career, impacted on their wellbeing. Additionally, as SETs and SET comments occur at the end of each teaching period, participants regularly discussed the subsequent wellbeing issues as being a cycle that saw anxiety grow in the weeks/months before SET data was released, and then could take an equally long (or longer) amount of time to subside following the results. Though 67% of participants identified SET comments as negatively affecting their wellbeing, wellbeing is not a standard or definitive term, and is one that can be somewhat defined by the participant's own experiences (Tucker, 2014). However, 16% of participants stated that they had sought professional medical help due to work related stress, and highlighted that SET comments and data played a role in their stress.

That SET comments, and SETs more widely, contribute to wellbeing and mental health issues of academics is also somewhat expected, but SETs and SET comments can also impact on academic career progress and objectives. 52% of participants indicated that they had delayed promotion applications on at least one occasion due to unfavourable SET results and comments. Additionally, 30% of participants indicated SET results and comments had negatively impacted on their employment opportunities, or had been used by their faculty leaders as a reason to delay their promotion requests. In both examples, the academics to delay or be denied promotion due to SET results were in over 70% of circumstances women or marginalised academics.

In looking at the impact of SETs and SET comments on academics, it is often difficult to separate the two. Negative results are often paired with negative comments and one only exacerbates the negative impact of the other. Thus, though this is a paper primarily focused on comments in SETs, the wider impact of SETs on academic careers cannot be ignored or entirely set-aside in this discussion (Heffernan, 2020, 2021b, 2022. However, if we focus on what type of comments different demographics of academics receive, we see trends in the prejudice nature of SET results (Boring et al., 2016; Fan et al., 2019) be reflected in the tone of comments they receive.

Demographics and their relationship to abusive comments

Almost every piece of data generated in this study could itself be analysed by experts in each field. For example, scholars of gender theory, race, or disability theory could write journal articles or book chapters about what this study's generated data can tell us about people from different marginalised backgrounds working in the ostensibly white, middle-class, privileged domain of the university. This paper is nonetheless about highlighting the type, and extent, of problems that student comments cause, why they (and SET results) do not provide a clear representation of course quality or teacher effectiveness, the extent to which they are abusive, and why they should be removed to spare staff from discrimination and wellbeing issues.

An examination of the most populous participant groups (white, straight, able-bodied academics identifying as men, and white, straight, able-bodied academics identifying as women) gives an overall indication of the volume and type of abuse different groups can receive. The amount of abuse, as is evident from Tables 2, 3, 4 and 5, is that white, straight, able-bodied men reported abuse at approximately 55%, and women at 63%. However, the type of abuse and how each group reacted to the comments was very different.

Of the 63 male participants who fit this demographic group, approximately one-third of the abusive comments related to the class being taught, for example:

Class is dumb.
The course is stupid, it is useless for my future.

Around one-third of the comments related to personal attacks, for example:

This lecturer is out of touch. They don't understand the modern classroom.
If Dr (name) was a product that I had purchased at a retail store, I would have returned it by now and asked for my money back.
Dickhead.

At the worst level (for this demographic), participants spoke of regular fat-shaming or comments on appearance. Several participants noted being called 'gay' or 'faggot', while another participant wrote of regularly being referred to via names such as 'Ching Chong' as students perceived him to be of Asian descent. These participants wrote of the comments as being largely inconsequential because they were attacking elements that did not exist (such as their heritage or sexual identity because the participants identified as straight, white men), however, they regularly reflected on the comments people from these groups must receive. However, one participant, who indicated their Jewish heritage was evident via their name, also wrote of 'on occasion' having antisemitic comments or swastikas drawn in their student feedback, and noted that this increased in the 'wake of an attack on a nearby synagogue'. The participant spoke of these attacks being extremely confronting as they continue to receive regular attacks on them, their heritage, and their community.

Though 55% of white, straight, male participants indicated receiving abusive comments, these comments are on a spectrum. Remembering that the survey asked 'Have you ever received comments that contained abusive, offensive, or derogatory words or phrasing?', it is the participant's choice to decide how they felt about the comments. Thus, some readers of this paper may view some of the above comments more as harsh critiques, but that is not how the participant interpreted the comment, and it is their interpretation that impacts on their wellbeing, mental health, and career and promotion aspirations.

The tone of abuse alters significantly when examining the comments straight, white, able-bodied women academics received. Across the 191 participants who fit the category, we again see some comments focus on the class, for example:

The class is bullshit, I shouldn't have to study it.
[Name] is not a fair marker and gives better grades to people she likes [in a unit with a teaching team and blind marking].

However, while this type of comment made up around 30% of the comments men received, they made up less than 10% of the comments women received. By a significant margin of around 60%, a majority of the comments women receive are about their body, appearance, or age. Participants spoke at length of regularly receiving comments the reflect the tone of those below:

[Name] has a nice ass, I like to watch her walk around.
[Name] is too old to teach this class, give it to someone younger.
[Name] is going to die of a heart attack if she doesn't do something about her weight problem; I'm just saying this for her own good because I'm genuinely concerned about people

who are this overweight and don't do anything about it, and people are afraid to say something to overweight people these days because they will get criticized and attacked for it, because fat people claim that it's 'body shaming,' so I'm doing it here because it's anonymous.
[Name] is too fat to be a health lecturer.

In the worst category of abuse, participants spoke of receiving generally high-level abusive comments.

Comments including '[Name] is young and pretty' are also mixed with. '[Name] is a slut'.
Lecturer is hot and should deliver classes topless to improve the student experience.
[Name] is a bitch.
[Name] is an ugly fat cow.
[Name] should be scared in a dark alley [if the student came across them].
[Name] dresses too revealing and is a cunt.

The above is only a short sample of the type of comments women academics receive. Remembering that 59% of participants said they received abusive student comments, around 90% of those comments for women academics were about appearance, included threats, and the tone of abuse was significantly different to that received by men. The tone of comment also alters what each participant group perceives as abuse. For example, several male participants spoke of having their qualifications questioned as being a form of abuse. In contrast, several women specifically identified having their qualifications questioned as not being abusive and in fact being a 'best case scenario' when compared to their usual comments of being 'fat' or 'ugly'.

It is also necessary to consider the pushback that occurs from staff due to these comments. Zero male participants indicated that they had made attempts to have student comments (or the associated scores/results) removed from the system so that the comments were deleted and SET result scores not included in their class average. Conversely, around 10% of women academics wrote of unsuccessfully campaigning to have comments removed from the system and student's associated satisfaction survey results removed. One could suggest that male academics never pursued this course of action because their SET results and comments are not as bad or abusive at the same rate as female participants, and so, the pressure to seek change is not as strong. Nonetheless, it is important to consider the wider toll that the abuse and lack of support can have on women academics.

One participant wrote in detail of the trauma and lack of support they received. In response to the SET question 'What would improve the subject?' one academic received the student comment 'better looking gash'. The participant spoke of the difficulty of receiving this comment as an early career academic, but made clear that the lasting trauma came from then being reprimanded for the comment and having to explain the meaning of the comment to their supervisors. Following this, the participant attempted to have the student's survey removed due to clear bias, but the request was rejected, the participant being told that a student could write the abusive comment but be fair in scoring the other questions.

Discussion

When looking at the research questions that informed this paper's design and purpose, it is important to consider that the answers to these questions can be viewed as, (a) reasons why comments need to be removed or censored in SETs, and (b) further evidence of why SETs should be removed entirely due to their clear prejudices towards women academics and academics from marginalised backgrounds.

RQ1 sought to answer what groups are receiving abusive comments, the data summary in Tables 2, 3, 4 and 5 and the accompanying analysis makes clear women academics, academics from different gender and sexual identity groups, and disabled academics receive more comments than other groups. However, particularly when it is clear abusive comments impact on mental health and wellbeing, RQ2 and this paper's examination of the types of comments the most populous categories of white, straight men and women, makes it clear that some groups receive much more severe comments (and threats) than others. As was also made clear in the literature (Boring et al., 2016; Fan et al., 2019; Uttl & Smibert, 2017), it is important to remember that negative SET results are lower and abusive comments increase for more marginalised groups. RQ3 was concerned with the impact of the abuse and the study makes clear that SET comments need to be removed. 59% of participants report receiving abusive comments, 67% of that number stated that abusive comments negatively impact on their wellbeing, and 16% said SET comments contribute to them seeking professional mental health support. It is also crucial to note that the distribution of abuse, and type of abuse, is not evenly spread across different demographic groups – women academics and academics from marginalised backgrounds receive a bulk of the comments, and a majority of the highly abusive content. Without any other information, this finding is problematic for a sector that claims to be inclusive, diverse, and conscious of creating safe working environments.

We know academics are regularly receiving abusive comments, we know this is impacting on their mental health. This is also not new information; Felton et al. in 2008 reported that websites like RateMyProfessors.com (for all their faults) were nonetheless aware that the impact of abuse on academics was undefendable, and thus took the step of deleting abusive comments. However, this paper's findings also need to be considered in terms of the existing and growing knowledge we now have regarding SETs, career progression, and marginalised groups. Researchers have known for decades that SET results are prejudice towards an academic's ethnicity, age, sexual identity, and appearance (Anderson & Miller, 1997; Cramer & Alexitch, 2000; Worthington, 2002), and these are additionally the groups that receive the highest volume, and worst content, of abusive student comments. This is why it is important for readers of this article who are academics to consider that this problem may not be about the abuse and results you receive, but it is vital to understand the context in which negative results and abusive comments may be given to those around you or others in the sector.

The issues surrounding SET comments and results also go beyond the mental wellbeing and health issues, and fact that no one should be open to anonymous abuse in their workplace; they also impact on career aspirations. As was discussed in the literature review, SETs provide data that appears sound (Osoian et al., 2010) and is arguably why institutions use this data as a guide in terms of hiring, firing, and promoting staff (Jones et al., 2014; Uttl & Smibert, 2017). In this study, 52% of participants indicated that they had delayed promotion applications due to their SET results and comments during their 'promotion year', and 30% of participants said they had been denied promotion because of their SET

results and comments. Yet as this paper makes clear, marginalised groups within academia receive the highest number of abusive comments, a veracity of comment rarely experienced by more privileged groups, and receive lower SET results for conducting the same level of work; all while an increasing body of research (such as that discussed in the literature review) highlights the prejudice nature and flaws of the data being collected by SETs.

Conclusions

This paper is one of the largest explorations of the volume and tone of abusive comments in SETs to take place, and it adds to the existing literature that highlights the need for SETs to be removed because they are not an accurate measure of course content or teacher quality.

This study has shown that while 59% of the surveyed academics have indicated they receive abusive comments in their student feedback, it is clear that women and those from marginalised groups receive a higher number of abusive comments, and the attacks these groups receive are also of a more personal and often sexual nature.

This paper provides an evidence base that demonstrates to the sector that academics are being abused, the abuse is taking a wellbeing and mental health toll, and the abuse and prejudice nature of SETs is impacting on careers and career progression; particularly of women and marginalised academics. The levels of abuse evident from this study, and fact that some institutions are now taking steps to remove SETs or censor SET comments, is evidence both that the problem exists and there is some shift in the sector to make changes. However, the need for swift and major policy changes is time critical. The university sector's women academics and those academics from the most underrepresented groups are those subjected to the most abuse, and are also those most negatively impacted by the prejudice nature of student evaluations when it comes to employment and promotion.

Every teaching period where SETs and student comments are collected, is another teaching period where the sector is effectively condoning the abuse of its marginalised academics. The impact SETs have on careers and promotions also means the academy is facilitating the further marginalisation of its already marginalised academics – the exact groups so many institutions claim to value and protect.

Declarations

Disclosure statement The author wishes to disclose no financial benefit or interest from this work.

References

Adams, S., Bekker, S., Fan, Y., Gordon, T., Shepherd, L., Slavich, E., & Waters, D. (2021). Gender bias in student evaluations of teaching: 'Punish[ing] those who fail to do their gender right'. *Higher Education*. https://doi.org/10.1007/s10734-021-00704-9.

Anderson, K., & Miller, E. (1997). Gender and student evaluations of teaching. *Political Science and Politics, 30*(2), 216–219. https://doi.org/10.1017/S1049096500043407.

Arthur, L. (2009). From performativity to professionalism: lecturers' responses to student feedback. *Teaching in Higher Education, 14*(4), 441–454. https://doi.org/10.1080/13562510903050228.

Benton, S., Cashin, W., & Manhattan, K. (2012). Student ratings of teaching: a summary of research and literature. IDEA Center. http://www.ideaedu.org.

Boring, A., Ottoboni, K., & Stark, P. (2016). *Student Evaluations of Teaching (Mostly) Do Not Measure Teaching Effectiveness*. ScienceOpen Research. https://doi.org/10.14293/s2199-1006.1.sor-edu.aetbzc.v1.

Brandl, K., Mandel, J., & Winegarden, B. (2017). Student evaluation team focus groups increase students' satisfaction with the overall course evaluation process. *Medical Education, 51*(2), 215–227. https://doi.org/10.1111/medu.13104.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa.

Clarke, V., & Braun, V. (2017). Thematic analysis. *The Journal of Positive Psychology, 12*(3), 297–298. https://doi.org/10.1080/17439760.2016.1262613.

Cramer, K., & Alexitch, L. (2000). Student evaluations of college professors: identifying sources of bias. *Canadian Journal of Higher Education, 30*(2), 143–164.

Creswell, J. (2013). *Research design: qualitative, quantitative and mixed method approaches*. SAGE.

Cunningham-Nelson, S., Baktashmotlagh, M., & Boles, W. (2019). Visualizing student opinion through text analysis. *IEEE Transactions on Education, 62*(4), 305–311. https://doi.org/10.1109/TE.2019.2924385.

Denzin, N., & Lincoln, Y. (2011). *The SAGE handbook of qualitative research*. SAGE.

DiPietro, M., & Faye, A. (2005). Online student-ratings-of-instruction (SRI) mechanisms for maximal feedback to instructors. *30th annual meeting of the professional and organizational development network*. Milwaukee, WI.

Enache, I. (2011). Customer behaviour and student satisfaction. *Economic Sciences, 4*(53), 41–46.

Fan, Y., Shepherd, L., Slavich, D., Waters, D., Stone, M., Abel, R., & Johnston, E. (2019). Gender and cultural bias in student evaluations: Why representation matters. *PLoS ONE, 14*(2). https://doi.org/10.1371/journal.pone.0209749.

Felton, J., Mitchell, J., & Stinson, M. (2004). Web-based student evaluations of professors: the relations between perceived quality, easiness and sexiness. *Assessment & Evaluation in Higher Education, 29*(1), 91–108. https://doi.org/10.1080/0260293032000158180.

Felton, J., Koper, P., Mitchell, J., & Stinson, M. (2008). Attractiveness, easiness, and other issues: student evaluations of professors on RateMyProfessors.com. *Assessment & Evaluation in Higher Education, 33*(1), 45–61. https://doi.org/10.2139/ssrn.918283.

Hamermesh, D., & Parker, A. (2005). Beauty in the classroom: Instructors' pulchritude and putative pedagogical productivity. *Economics of Education Review, 24*(4), 369–376. https://doi.org/10.1016/j.econedurev.2004.07.013.

Heffernan, T. (2020). Examining university leadership and the increase in workplace hostility through a Bourdieusian lens. *Higher Education Quarterly*. https://doi.org/10.1111/hequ.12272.

Heffernan, T. (2021a). Sexism, racism, prejudice, and bias: a literature review and synthesis of research surrounding student evaluations of courses and teaching. *Assessment & Evaluation in Higher Education*. https://doi.org/10.1080/02602938.2021.1888075.

Heffernan, T. (2021b). 'There's no career in academia without networks': Academic networks and career trajectory. *Higher Education Research and Development, 40*(5), 981–994. https://doi.org/10.1080/07294360.2020.1799948.

Heffernan, T. (2022). *Bourdieu and Higher Education: Life in the Modern University*. Springer.

Hendrix, K. (1998). Student perceptions of the influence of race on professor credibility. *Journal of Black Studies, 28*, 738–764. https://doi.org/10.1177/002193479802800604.

Jones, J., Gaffney-Rhys, R., & Jones, E. (2014). Handle with care! An exploration of the potential risks associated with the publication and summative usage of student evaluation of teaching (SET)

results. *Journal of Further and Higher Education, 38*(1), 37–56. https://doi.org/10.1080/0309877x.2012.699514.

Kosinski, M. (2015). Will Facebook replace traditional research methods? Social media offers researchers a window into the human experience. Insights by Stanford Business. Available at https://www.gsb.stanford.edu/insights/will-facebook-replace-traditional-research-methods. Accessed 11 Mar 2021.

Kosinski, M., Matz, S., Gosling, S., Popov, V., & Stillwell, D. (2015). Facebook as a research tool for the social sciences: Opportunities, challenges, ethical considerations, and practical guidelines. *The American Psychologist, 70*(6), 543–556. https://doi.org/10.1037/a0039210.

Marsh, H. (2007). Students' evaluations of university teaching. dimensionality reliability, validity, potential biases and usefulness. In: Perry, R., & Smart J. (Eds). *The scholarship of teaching and learning in higher education: an evidence-based perspective* (pp. 319–383). Springer.

Miles, M., Huberman, A., & Saldana, J. (2013). *Qualitative data analysis*. SAGE.

Mitchell, K., & Martin, J. (2018). Gender bias in student evaluations. *Political Science & Politics, 51*(3), 648–652.

Osoian, C., Nistor, R., Zaharie, M., & Flueras, H. (2010). Improving higher education through student satisfaction surveys. *Proceedings of the 2nd International Conference on Education Technology and Computer*. https://doi.org/10.1109/icetc.2010.5529347.

Punch, K. (2013). *Introduction to social research: quantitative and qualitative approaches*. SAGE.

Rubin, D. (1998). *Readings in cultural contexts, Chapter Help! My Professor (or doctor or boss) doesn't. talk English*. Mayfield.

Schmidt, B. (2015). Is it fair to rate professors online? Retrieved on 16 February 2022. https://www.nytimes.com/roomfordebate/2015/12/16/is-it-fair-to-rate-professors-online.

Schmidt, B. (2022). Gendered language in teacher reviews. Retrieved on 16 February 2022. https://tinyurl.com/2p8fcpa2.

Shah, M., & Nair, C. (2012). The changing nature of teaching and unit evaluations in Australian universities. *Quality Assurance in Education, 20*(3), 274–288. https://doi.org/10.1108/09684881211240321.

Shulman, L. (1988). Disciplines of inquiry in education: An overview. In R. M. Jaeger (Ed.), *Complementary Methods for Research in Education* (pp. 3–17). American Educational Research Association.

Short, H., Boyle, R., Braithwaite, R., Brookes, M., Mustard, J., & Saundage, D. (2008). A comparison of student evaluation of teaching with student performance. In: MacGillivray, H. (ed). *Proceedings of the 6th Australian conference on teaching statistics*; *Melbourne* (p. 1–10).

Stark, P., & Freishtat, R. (2014). An evaluation of course evaluations. *ScienceOpen Research*. https://doi.org/10.14293/s2199-1006.1.sor-edu.aofrqa.v1.

Sunindijo, R. (2016). Teaching first-year construction management students: lessons learned from student satisfaction surveys. *International Journal of Construction Education and Research, 12*(4), 243–254. https://doi.org/10.1080/15578771.2015.11219371.

Tucker, B. (2014). Student evaluation surveys: anonymous comments that offend or are unprofessional. *Higher Education, 68*(3), 347–358. https://doi.org/10.1007/s10734-014-9716-2.

Uttl, B., & Smibert, D. (2017). Student evaluations of teaching: teaching quantitative courses can be hazardous to one's career. *PeerJ*. https://doi.org/10.7717/peerj.3299.

Valencia, E. (2021). Gender-biased evaluation or actual differences? Fairness in the evaluation of faculty teaching. *Higher Education*. https://doi.org/10.1007/s10734-021-00744-1.

Vivanti, A., Haron, N., & Barnes, R. (2014). Validation of a student satisfaction survey for clinical education placements in dietetics. *Journal of Allied Health, 43*(2), 65–71.

Worthington, A. (2002). The impact of student perceptions and characteristics on teaching evaluations: A case study in finance education. *Assessment & Evaluation in Higher Education, 27*(1), 49–64. https://doi.org/10.1080/02602930120105054.