https://studentsuccessjournal.org/
Volume **14** (3) 2023 https://doi.org/10.5204/ssj.2756

Allegations, Abuse and Discrimination: Using Student Evaluation of Teaching Surveys to Support Student and Educator Wellbeing

Samuel Cunningham, Abby Cathcart and Tina Graham
Queensland University of Technology, Australia

Abstract

Student Evaluation of Teaching surveys (SETs) are used at universities to inform teaching practice and subject design. However, there is increasing concern about the impact of allegations, abuse, and discrimination in survey open text components. Here we discuss the implementation of an automated screening mechanism using a combination of dictionary and machine learning approaches. We present both a process diagram detailing how the screening is performed, as well as a form of categorisation for comments that are unacceptable or indicate a potential risk of harm. Examples of real comments in each of these categories are presented to demonstrate the depth of the challenge and potential cause for concern. Ultimately, we argue that student and educator wellbeing are inextricably connected and exposing staff to abusive and discriminatory comments causes harm. Furthermore, SETs are an important channel for students to raise concerns about their own wellbeing and potentially unsafe experiences in the learning environment.

***Keywords**:* Student Evaluation of Teaching; wellbeing; teacher evaluation; risk of harm.

Introduction

Student Evaluation of Teaching surveys (SETs) are one of the most common mechanisms used by universities to gain insights into the student learning experience (Spooren et al., 2013). However, in recent years there have been growing concerns around the validity and reliability of SETs in measuring student experiences, or teaching quality. There have been calls to acknowledge the negative impact that SETs may have on educator wellbeing due to the growing recognition that SETs results can include abusive, discriminatory or cruel feedback to staff members (Lakeman et al., 2021; 2022b). More recently researchers have pointed to the importance of educator wellbeing in modelling and supporting student wellbeing (James et al., 2019). This article highlights the emotional toll that SETs place on educators and argue that this is doing significant damage to staff and student relationships and undermining strategies to promote the sense of belonging, trust and community that are central to ideas of student success. We outline an approach taken by an Australian university to reframe practice and protocols relating to SETs to protect staff from unintended consequences and to develop screening processes which promote constructive feedback practices and alert the institution to students that may need additional support.

Background

There is a growing body of literature focused on student wellbeing which is viewed as a core domain of student success (van der Zanden et al., 2018). The mental health of students has become a focus for researchers globally with repeated studies highlighting anxiety, depression and psychological distress at levels far greater than those found in the broader population

(Larcombe et al., 2016; Stallman, 2010). Researchers have pointed to the need for institutions to help students build positive perceptions of self, and coping skills, and to ensure there are structures to provide student services and social support (Baik et al., 2019; Taylor, 2017).

More recently, researchers have focused on the relationship between student wellbeing and the wellbeing of the educators that teach them. For example, James et al. (2019) in a study of law academics in the UK and Australia posited that if we expect law student wellbeing to improve "then it is reasonable to hope and expect law teachers' wellbeing to be sound" (p.76). Similarly, research exploring the experiences of university-affiliated mentors for pre-service teachers' placements highlighted key inhibitors to staff wellbeing and the need for more attention on support for staff as well as students (Gillett-Swan & Grant-Smith, 2020). Others have suggested that constructive connections between educators and students promote occupational wellbeing and a commitment to ongoing educator skills development (Cui, 2022).

There has long been concern in the university sector about the pressures on staff shaped by the casualisation of academic employment, rising performance expectations, increased student demands and broader neoliberal agendas that promote concepts of students as customers rather than as partners in the learning experience (Finney & Finney, 2010; Matthews et al., 2018; Troiani & Dutson, 2021). The 2022 Times Higher Education work-life balance survey reported that 74% of academics felt that work had a negative effect on their mental health (Williams, 2022). The transition period as students start university is a well-known period of vulnerability and decline in psychological wellbeing and this in turn can create additional pressure on staff (Brooker & Woodyatt, 2019). Additionally, the COVID-19 pandemic has led to a period of significant disruption in almost every aspect of the lives of educators and students including the rapid transition to wholly or largely online teaching. As the sector moves into its 'new normal' this transition period is shaped by structural changes adding to the pressure on academics and professional staff. For example, in Australia in the first half of 2021, 41,000 university jobs were lost, more than 60% of them to women (Littleton, 2022). In Australia, pre-pandemic, international students contributed 27% of the revenue and 30% of the student population in public universities (Littleton, 2022, p.28). Despite significant efforts to retain international student numbers, they are not returning to pre-covid numbers, let alone growing in line with the previous trajectory, and it is estimated that over the next two years international student revenue will be 50% lower than if the pandemic had not happened (Littleton, 2022). This ongoing revenue challenge is likely to continue to put pressure on staffing levels leading to increased workload challenges for staff.

Significantly, the post-pandemic university experience is also shaped by a radical disruption to forms of teaching and an ongoing focus on blended and online learning (García-Morales et al., 2021). This too should be understood as a period of vulnerability where staff and students may experience declines in psychological wellbeing. For example, a recent study exploring the emotions experienced by academics as they transitioned to online teaching applied the Kübler-Ross (1973) model of grief and found strong alignment with staff experiencing denial, anger, bargaining, depression, acceptance and ultimately (for some at least) hope (Taylor et al., 2022). It is important to highlight that the move to online teaching has been a period of significant transition and that time and dialogue about emerging forms of teaching and support for learning is needed to make sense of loss, grief and growth (Taylor et al., 2022). It is within this context that the sector has reported growing concerns about the use of SETs (James et al., 2019).

SETs are widely used in academic promotion, performance discussions, quality enhancement and assurance initiatives and in recruitment (Cunningham et al., 2023). SETs typically include a number of quantitative questions about teaching and curriculum as well as an opportunity for students to provide open text comments (Laundon et al., 2023). There is a growing concern, however, about the negative impact that malicious and unacceptable comments have on educators (Heffernan, 2021, 2023; Lakeman et al., 2022a). For example, a recent study on educator wellbeing in the UK and Australia clearly identified SETs as contributing to stress through "mischievous and untrue criticisms that damage the morale of teachers" (James et al. 2019, p. 79). Others have pointed to the "emotionally devastating" impact of SETs on mental health, particularly for precariously employed female educators (Smele et al., 2021, p. 975) and "stress, distress, disorder and coping" responses with clear impact on educator wellbeing (Lakeman et al., 2022b, p. 1490).

There is a body of research dedicated to identifying strategies to mitigate the impact of SETs on academic wellbeing. These vary from calls to abandon or completely reform SETs (Heffernan, 2023; Lakeman et al., 2022c; NTEU, 2018), through to analysis of cognitive and behavioural approaches designed to help educators to cope (Lakeman et al., 2022b). Others have focused on the style of SETs reports, for example by providing visualisations of qualitative student comments to direct attention on feedback themes rather than individual comments (Cunningham-Nelson et al., 2021).

Researchers have pointed to the limitations of much of the research on SETs as characterised by a narrow or single focus and recent scholarship has attempted to adopt a more macro stance, using situational mapping to identify key actors and elements

and the discursive constructions which drive the relationships between them (Lloyd & Wright-Brough, 2022). They highlight the discursive constructions that shape the perceptions of SETs as a form of surveillance within the university sector, noting the importance of considering the future and silent actors including, for example, promotion panels who may be given access to SETs data. This mapping further highlights the wider literature which has connected SETs to occupational stress (Lee et al., 2022; NTEU, 2018). Despite the growing evidence base linking SETs to occupational stress, there is little indication that the sector is moving away from the use of SETs to provide both quantitative and qualitative feedback on teaching and curricula.

Furthermore, there is limited research which focuses on institutional responses to detect and remove abusive or discriminatory comments before they are seen by educators. A small number of studies have focused on leveraging machine learning to facilitate comprehensive screening of SETs in efforts to identify and remove threatening and discriminatory comments (Cunningham et al., 2023; Gibson et al., 2022). This research therefore seeks to move beyond the critique of SETs and instead proposes that their continued use be dependent on a renewed focus on mechanisms that can be implemented to reduce stress and risk associated with institutional SETs processes. It extends previous work in which a screening mechanism is presented (Cunningham et al., 2023). In this work, direct implications for students and educators are further discussed with a clear focus on educator and student wellbeing. Thus, this research projects aims to outline the process of how unacceptable and risk of harm comments are managed at a large metropolitan university in order to advance approaches to categorising and responding to SETs in ways that promote educator and student wellbeing.

Process and Context

Review of Existing Practice

The Queensland University of Technology (QUT) undertook a comprehensive and consultative review of unit and teaching evaluations in 2019. The review was informed by scholarly literature on evaluation, teaching quality, student voice, stakeholder consultation, and a sector scan of evaluation practices. As part of the review, concerns were noted from the academic community about the impact on staff of harassing, threatening, abusive and discriminatory comments, including personal attacks relating to appearance or other matters unconnected to units and teaching. In consultation with senior university leaders, concerns were raised about students who make evaluation comments that imply a risk of harm to self or others.

Staff members were expected to notify the university's evaluations unit if they identified an unacceptable comment once the survey results were shared with them. This meant that harm was potentially caused when a staff member read the comment. Comments were then reviewed and removed manually, leading to delays in their removal and potential further distress for staff members. Students were not notified when unacceptable comments were identified and removed, thus missing an opportunity to educate students on professional conduct and constructive feedback. Comments were not screened for risk of harm to self or others.

Recommendations adopted from the review included pre-screening unit and teaching comments for unacceptable and risk of harm comments, and adopting an educative approach in response to students making unacceptable comments, so they had an opportunity to reflect on and learn about professional feedback.

The evaluations unit then took a proactive partnership-based approach in 2020 to scope the issue of unacceptable and risk of harm comments within the institution. Advice was sought from the University's Office of the Registrar, Equity Committee, Human Resources and Health Safety and Environment (HSE) teams in considering the current processes and potential risk of harm to staff and students. Balancing the rights of students to express both positive and negative views against potential harm to either themselves or others, student representatives were also included in the consultations. Academic representatives from each faculty were offered the opportunity to provide their views via committee and stakeholder meetings. The evaluations unit also drew on past experience of managing student evaluations of teaching surveys over time. From these consultations, a set of working definitions and a dictionary, which were essential tools in the screening process, were built. This strategy was then developed and trialled with procedures in place to respond to identified problems. Human research ethics approval for this research (Approval No. 2000000997) was obtained through QUT.

Definitions of Unacceptable and Risk of Harm Comments

Developing a definition of the kinds of comments that were deemed unacceptable was a complex and challenging process involving extensive consultations with staff and students. The definitions below were created based on existing policies on staff and student conduct, and an analysis of comments that had been requested for removal by staff over a 3-year period. A definition of 'risk of harm' was developed to capture the rare but vital occasions when a student made a comment that raised

key concerns that they needed urgent help or were a potential risk to others. These definitions have evolved and been further refined since the first intervention in Semester 1, 2021. The most notable change was the inclusion of allegations concerning staff behaviour as a number of these comments were detected in the first year of screening. Both the unacceptable and risk of harm categories may contain negative or discriminatory elements, but do not have to.

Unacceptable comments are defined as (but are not limited to):

a. *Harassing, threatening or abusive language, including profanities, and language that is intimidating or discriminatory. Discriminatory comments include comments relating to sexual orientation, gender identity, intersex status, disability, ethnicity, marital status, nationality, age, religion and/or political persuasion.*
b. *Personal attacks relating to appearance or other matters unconnected to units, teaching or the learning experience.*

Risk of harm refers to the physical, psychosocial and emotional wellbeing of students and staff. Risks may be immediate or potential and the risk may be to the student themselves, to other students or to staff members.

Risk of harm (staff or student) comments are defined as (but not limited to):

a. *References to self-harm or suicide.*
b. *Threatening acts of violence or intimidation towards staff, other students and/or QUT property.*
c. *Allegations concerning staff behaviour including discrimination, bullying and harassment*

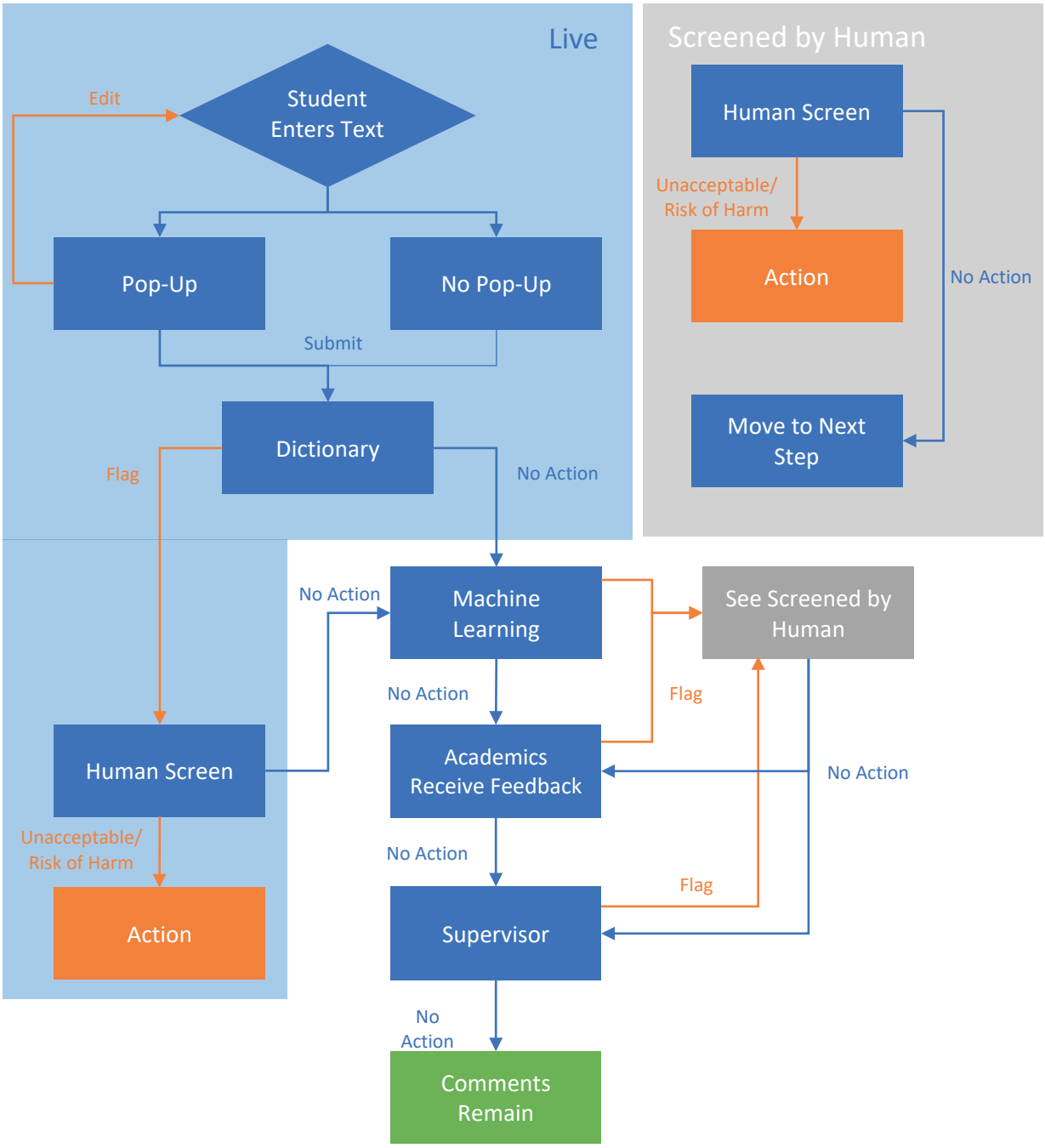
Screening Process for Unacceptable and Risk of Harm Comments

The screening process consists of a combined dictionary approach and a machine learning approach. This built on emerging capabilities to leverage technology for screening comments in other contexts including for example of abuse on X (Twitter) (Bashar et al., 2020). Utilising a list of potentially unacceptable or risk of harm keywords, and a machine that had been trained on past examples of unacceptable or risk of harm comments, more comments that contain potentially unacceptable or risk of harm sentiments can be detected.

Figure 1 shows an overview of the screening process used. The blue shaded section represents the elements which occur whilst the survey is open or live. Additionally, several times throughout the process, certain comments are screened by a human. These are represented in the grey block of the diagram. Further details about the technical elements of the entire screening process can be seen in previous work (Cunningham et al., 2023).

Figure 1

Screening Process Overview



In 2021, Qualtrics was introduced as the university's enterprise survey tool. The dictionary component of the screening process was implemented in Semester 1, 2021 using the TextIQ feature within Qualtrics. Qualtrics allows for students to first be notified via an onscreen pop up whilst completing the survey if a potentially unacceptable word or term is used. Students have the option to edit or continue with submission if the popup is triggered. Once feedback is submitted, the comment is then matched further against an extensive list of potentially unacceptable words and phrases. The dictionary of potentially unacceptable terms was progressively refined since first deployment in Semester 1, 2021 using existing lists available online, and words known from past comments to the central evaluations team. It contains elements such as: profanities, words related to mental health and derogatory terms. The dictionary is an important part of the screening process as it ensures that selected

keywords or phrases will be screened by a person. Where potentially unacceptable or risk of harm comments are identified through the automated screening of survey data, a senior central evaluations team member checks the flagged comments. Comments that are identified as unacceptable or risk of harm can have action taken. Potential actions that can be taken will be discussed in the results and discussion section of this article. This dictionary screening process occurs live whilst the survey is ongoing to maximise opportunities for students to be contacted and invited to revise their comments before the survey closes.

A machine learning model was developed to complement the dictionary approach. Broadly, machine learning applies pattern recognition and classification processes to group or identify data (Langley, 1996). The type of model selected for this application is known as a RoBERTa model. RoBERTa models have advantages as they can be trained in one domain, and then adjusted to suit a more nuanced domain (Liu et al., 2019). This machine learning model is trained on three datasets to gain a purposeful understanding. It is initially trained on a large dataset of Wikipedia to understand English corpus, then movie reviews to develop a polar decision-making sense, and then finally examples of student comments. In this case, prior examples of comments were given to a machine learning model, with labels assigned for both acceptable, and unacceptable or risk of harm comments. The model then learns from these examples, as well as general samples of the English language, to make a decision if the comments are acceptable or not. Detailed information about the machine learning model used can be found in previous work (Cunningham et al., 2023). The machine learning model scans all student comments after the survey has closed.

It is important to note that both the dictionary and machine learning approaches are expected to improve over time, as the dictionary is refined, and as further examples of relevant comments are used to train the machine learning model. For the dictionary component of the process, additional words or phrases can be added at any point to be flagged. For the machine learning element, new examples of full comments can be added, and then the model can be trained again, with a wider set of training data. This machine learning model was updated once at the end of the first semester, and again at the end of the second semester. Furthermore, as language develops over time, it is important to ensure that the dictionaries and examples of comments are kept up to date, to ensure that they remain current.

Findings and Discussion

In 2021 more than 100,000 comments were screened, with approximately 4,000 identified as potentially unacceptable or risk of harm by the dictionary approach, and a further 4,000 identified by the machine learning approach. After these comments were reviewed manually by an evaluations staff member, 100 total comments were identified as being unacceptable or risk of harm, and some action was taken.

It is important to remember that there are many types of comments presented in SETs, including categories such as irrelevant or unrelated comments and superficial comments (Lakeman et al., 2022a). This current study focuses on the aforementioned categories of unacceptable and risk of harm comments. As with any written text some comments are ambiguous and may not fit exactly into one of the above definitions. This highlights the importance of context and ensuring that robust definitions are used, as well as the human element of the screening process for interpretation.

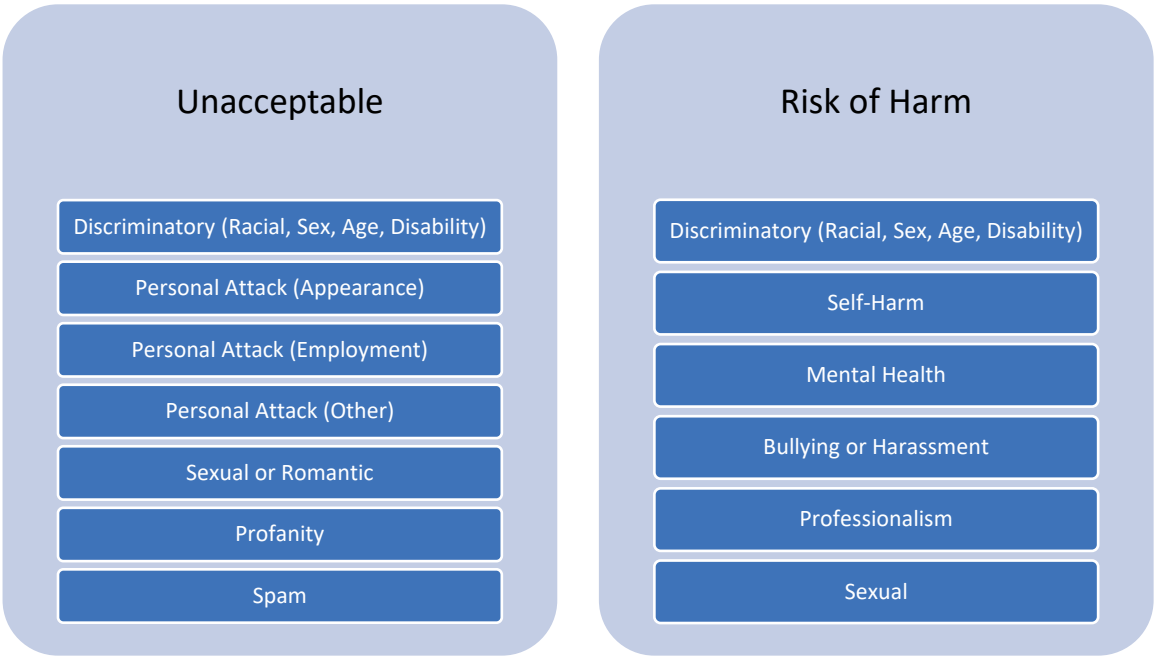
Categorisation of Comments

To further understand the types of both unacceptable and risk of harm comments being received a range of categories for these comments were determined. The categories are important because they allow for further investigation of the types of comments that are made in SETs. Figure 2 below shows a breakdown of these categories. This categorisation was developed through an iterative process as further comments were read. Comments were categorised into two groups: *Unacceptable* and *Risk of Harm*.

Discriminatory comments could be categorised either as unacceptable or risk of harm and have grouped into four categories, based on Australia's anti-discrimination law (*Age Discrimination Act 2004* Cth); *Disability Discrimination Act 1992* (Cth); *Racial Discrimination Act 1975* (Cth); *Sex Discrimination Act 1984* (Cth)). Unacceptable comments categories are shown on in Figure 2 with categories related to discrimination, personal attack, sexual, profanity and spam. Risk of harm comments are also shown in Figure 2, including discriminatory, self-harm, mental health and others. We present this categorisation as a starting point for educators to categorise risk of harm or unacceptable SET comments.

Figure 2

Categories of, Unacceptable and Risk of Harm Comments



Unacceptable Comments

Unacceptable comments may contain elements such as harassment, threats, discrimination, abusive language, and personal attacks. Unacceptable comments can have a large negative impact on educators, especially marginalised groups such as sessional educators and women (Heffernan, 2023; Kreitzer & Sweet-Cushman, 2022; Lakeman et al., 2022c). Previous studies have noted that attacks on educators' perceived job competency are common (Gibson et al., 2022). Our intervention reached similar conclusions and some examples of unacceptable comments detected through the screening process include: "You aren't the VC, start dealing with students or find a new job" and "Go back to North America where you belong". These comments were detected (and removed) before they were shared with educators.

Where a comment meets the definition of an unacceptable comment, the student who made the comment receives a message reminding them of their obligations under the QUT Student Code of Conduct and if the survey is still live, students are provided with the opportunity to edit or retract their comment during the survey period. If the student declines to edit or retract the unacceptable comment, a senior evaluations staff member decides whether the whole or part of the comment will be redacted. In 2021, 50% of students who were contacted through this process edited their comment. At the end of the survey any remaining comments that may be a serious violation of the code of conduct are referred to the Executive Dean for consideration and action in consultation with the Registrar's office.

Despite the robust methodology and definitions used to shape screening decisions there are still many comments that are a borderline and require expert input before determining whether they are unacceptable. This could be due to several reasons, such as context that may not be clear or potential different understandings of the meanings of words or phrases as language evolves. Other researchers have attempted to create an assessment protocol to assist reviewers in categorising comments ranging from low to high severity and covering self-harm, cyber bullying and cyber threats (Gibson et al., 2022). Although this is a helpful step, it is clear that within these categories there remains a high level of ambiguity in determining the seriousness of comments. For example, the authors state that a comment such as "Needs to learn how to speak English" would be categorised as low severity cyber bullying. However, if you are the educator on the receiving end of that comment you may have a very different interpretation including that it is discriminatory, potentially career damaging, and therefore unacceptable. Similarly, the student that made that comment may be surprised to hear that the university views it as either bullying or discriminatory as they saw it as feedback on poor communication skills.

Table 1 below shows these unacceptable categories and example comments.

Table 1

Unacceptable Categories and Examples

Category	Example Comment
Discriminatory	**1 -** "She shouldn't be running an online class. Unfortunately, she doesn't understand how to keep an online class engaged (perhaps due to age)." **2 -** "Nothing against [tutor], he's definitely a little out of his time period …" **3 -** "[Educator] strikes me as someone who has undiagnosed ADHD as he often loses track of time and his train of thought."
Personal Attack (Appearance)	**1 -** "This guy needs to lose some weight and go do a degree in teaching." **2 -** "Needs to comb his hair."
Personal Attack (Employment)	**1 -** "Complete idiot. Should be fired." **2 -** "[name] is the most incompetent unit coordinator in the Bachelor degree. She is lazy, unskilled and disorganised and should not be in a unit coordinator position."
Personal Attack (Other)	**1 -** "Whoever marked my assignment I hate." **2 -** "[Educator's] voice on the recorded tutes is the worst thing I think I've ever heard. She starts by putting on this toffee well-spoken accent which soon fades into her regular voice, where she speaks AT you."
Sexual or Romantic	**1 -** "Daddy? Yes." **2 -** "….is too attractive. It made it hard to concentrate on whatever he was saying while I was mentally undressing him …"
Profanity	**1 -** "The tutor was shit!!!" **2 -** "Go f**k yourself [educator]."
Spam	**1 -** [long section of text from a movie transcript]

Comments Indicating a Potential Risk of Harm to Students or Staff

Comments given in SETs can have the potential to either cause harm or may identify existing or potential harm. Although these surveys are not designed as a mechanism for students to report potential harm, it is important that universities acknowledge that SETs may be used for this purpose, and that these comments should be taken seriously. Notably, literature around SET surveys mention comments that are malicious or unacceptable (Lakeman et al., 2022a), and some of these comments do mention a potential harm to educators, however there has been very little discussion around the potential risk of harm to students highlighted in SET comments. Some examples of comments detected through the screening process include, "I was not able to attend lectures due to chronic depression and suicidal tendencies. Please be more forgiving in future.". In this instance, an email was sent to the student detailing information about the range of support available at the university. The student was not identified to any teaching staff members, and the email was sent from a central evaluations staff member.

Significantly, as shown in Table 1, our research identified that a number of students used the SETs to make allegations about discriminatory or inappropriate behaviour by the people teaching them. We categorised this as 'Risk of Harm (Staff Related)' to emphasise the potential risk to the student or to other students if the allegation was accurate or the behaviour repeated. For example, "[Tutor] made a lot of sexual comments, was very weird and [they] make a lot of students uncomfortable". In this instance, the Registrar recommended that the faculty investigate further, and the student was contacted by trained personnel and offered further support.

Table 2 below shows these risk of harm categories and example comments.

Table 2

Risk of Harm Categories and Examples

Category	Example Comment
Discriminatory	**1 -** "She commented to a student that she believed another student's English was not very good. In actual fact, the student she was referring to, who was of Asian appearance, had an Australian accent and had been born here." **2 -** "… However, during a discussion about perceptions of LGBTQI+ in different cultures, he said he still finds it uncomfortable seeing two men holding hands or kissing …" **3 -** "On one occasion on zoom, we were having a conversation regarding an assessment which I was quite upset over, to which he responded "I'm sorry to make your makeup run…"
Self-Harm	**1 -** "… I almost wanted to kill myself and that's also due to the uni pretending to care by offering mental health …"
Mental Health	**1 -** "… I myself have been in near breakdowns in class because I've been struggling with the content …" **2 -** "… My depression has re-emerged somewhat triggered by a sense of overwhelming hopelessness and not knowing how I'll ever make a difference without burning out…"
Bullying or Harassment	**1 -** "in my tutorial we were extremely intimidated and it was not a safe environment ...the tutor was extremely passive aggressive towards the class and would call you out in front of everyone if the answer was wrong..." **2 -** "I have been yelled at by a particular workplace supervisor, they have a habit of banging tables or walls when angry which is frightening, as well as throwing objects…"
Professionalism	**1 -** "Stop saying 'fuck' every second word. You are in a professional environment where there are many easily influenced young adults …" **2 -** "Stop disparaging your fellow colleagues and stop suggesting to students that it's okay to cheat on exams because 'everyone does it'."
Sexual	**1 -** "Good guy overall. Cursed at me a few times, made creepy comments towards female students." **2 -** "… made flirtatious comments towards a male student in our first tutorial class."

It is important to recognise the risks associated with psychological injury and the impact on learning and the student experience. Our strategy was to ensure that psychosocial risks were identified and managed for both students and staff. In 2022, a formal risk assessment was undertaken on the comment screening process currently undertaken. A panel was convened of student volunteers, including representatives of the Student Guild, and senior staff members from across the university including HSE and Human Resources to undertake this risk assessment. Ultimately several risks were identified, assessed, and controls enacted to remove or reduce the risk.

As an example of risk, the Head of School or delegate may fail to respond to a risk of harm (staff conduct) flagged comment and there is a risk that the problematic behaviours or actions continue. A control to mitigate the 'failure to respond' risk was implemented by requiring the Head of School to provide a response to the evaluations office about actions taken on the flagged comment prior to release of results. The risk of screening failing, and staff being exposed to a comment that is upsetting, led to the creation of a new risk control. This control is a formal professional development module about engaging with survey results, seeking support, and engaging with a supervisor. Additionally, results are never released on a Friday, to ensure collegial and formal support is available to the staff member during the working week.

Measures of Success

Through this screening process, and categorisation of comments, the priority was to help create a safer environment for both students and educators by dealing with unacceptable comments before they are shared with educators. In 2021 there was an 84% decrease in requests to remove unit feedback comments, compared to 2019, demonstrating a drastic reduction to the number of potentially unacceptable comments seen by educators.[1] This is particularly relevant in the light of other recent research which demonstrated that the number of negative comments made to female educators increased during the COVID-

[1] Institutional SETs were not deployed in 2020 due to COVID-19 related service disruptions and an institutional review of the process.

19 pandemic when teaching was predominately online (Tangalakis et al., 2022). Fifty per cent of students sent a retake link in 2021 updated their comment to remove previously included profanities. This shows two key benefits, with both the number of unacceptable comments being reduced, and students being offered the opportunity to amend their comment and improve their feedback literacy. A further critical measure of success is that through this process students who are in need of support are identified and directed to relevant support as needed.

Across an 18-month period, 284 comments were identified through the screening process as either being unacceptable or risk of harm. These comments were all attended to before they reached educators and students received additional support where appropriate. One important aspect through this process is that the tool becomes more effective through further refinement of the dictionary, and further examples of unacceptable and risk of harm comments used in the machine learning process.

Conclusions

This research has provided insight into the categories of unacceptable and risk of harm comments found in SETs and the implications of this feedback for the wellbeing of both educators and students. Although unacceptable comments represent a tiny proportion of SETs feedback, the examples highlighted here illustrate how cruel, abusive and discriminatory they can be. Even one unacceptable comment can have a detrimental impact on an educator's wellbeing. Additionally, students may use SETs to communicate a risk of harm connected to their own mental or physical wellbeing or to make serious allegations of risks experienced as part of their courses. While there are other avenues for students to raise concerns, our screening demonstrates that some students may use SETs to alert the institution. It is important to listen to students regardless of the mechanism used to raise a concern. Universities have a duty of care to ensure that systems are in place to monitor and manage these situations. This is particularly important when considering rises in concern related to student mental health. If universities intend to continue to use SETs it is crucial that screening is implemented to minimise the risk of harm to both students and educators. By developing institutional definitions of what is acceptable and unacceptable, universities can communicate more effectively and build confidence amongst staff and students that their wellbeing is prioritised. Creating a dictionary ensures university management, educators and students can develop a shared understanding of these definitions.

The process outlined in this article uses a combination of dictionary and machine learning to detect comments that may fit one of these definitions. However, as language changes over time the dictionary needs to be refined with further examples of comments to improve the machine learning model. The machine learning model, is also currently trained based on sensitive examples of unacceptable and acceptable comments, making sharing the model, or example comments themselves difficult to share widely. The work presented here also aligns more broadly with wider work in responding and dealing with online hate (Vidgen et al., 2021). Future work includes improvements to the machine learning model, in terms of training with more publicly available comments, so that it can be shared more widely.

The findings from this research detail an approach which is reactive in nature as the process is only triggered once a comment has been made by a student. We recommend this approach is paired with other proactive approaches, such as feedback literacy for students with particular emphasis on framing negative comments in a polite and constructive way (Carless et al., 2018). For example, at QUT, students are invited to provide feedback that is 'Actionable, Specific and Kind'. Additionally, support should be provided both for educators, and their supervisors to develop strategies for reflecting on negative but potentially actionable feedback. It is important to acknowledge that each person (student or educator) brings their own experiences into how they write and read comments.

In conclusion our findings here align with conclusions reached by Gillett-Swan and Grant-Smith (2020): "Wellbeing, while an individual outcome, is impacted by broader institutional and societal expectations and practices" (p. 406). If universities want to create environments conducive to student success, they need to first focus on institutional practices that see educator wellbeing as both a prerequisite and a central feature of structures and strategies.

Acknowledgements
The authors would like to thank Pip Jenkins, Ashleigh Quail, Ronan Kelly and Rick Somers for their support in the categorisation of unacceptable comments. They would also like to thank Md Abul Bashar and Richi Nayak for their support in developing the screening mechanism.

References

Age Discrimination Act 2004 (Cth). https://www.legislation.gov.au/Details/C2020C00283

Baik, C., Larcombe, W., & Brooker, A. (2019). How universities can enhance student mental wellbeing: The student perspective. *Higher Education Research & Development*, *38*(4), 674-687. https://doi.org/10.1080/07294360.2019.1576596

Bashar, M. A., Nayak, R., & Suzor, N. (2020). Regularising LSTM classifier by transfer learning for detecting misogynistic tweets with small training set. *Knowledge & Information Systems*, *62*(10), 4029-4054. https://doi.org/10.1007/s10115-020-01481-0

Carless, D., & Boud, D. (2018). The development of student feedback literacy: Enabling uptake of feedback. *Assessment & Evaluation in Higher Education*, *43*(8), 1315-1325. https://doi.org/10.5204/ssj.v10i3.1419

Brooker, A., & Woodyatt, L. (2019). 2019 Special issue: Psychological wellbeing and distress in higher education. *Student Success*, *10*(3), i-vi. https://doi.org/10.5204/ssj.v10i3.1419

Cui, L. (2022). The role of teacher-student relationships in predicting teachers' occupational wellbeing, emotional exhaustion, and enthusiasm. *Frontiers in Psychology*, *13*, 896813. https://doi.org/10.3389/fpsyg.2022.896813

Cunningham-Nelson, S., Laundon, M., & Cathcart, A. (2021). Beyond satisfaction scores: Visualising student comments for whole-of-course evaluation. *Assessment & Evaluation in Higher Education*, *46*(5), 685-700. https://doi.org/https://doi.org/10.1080/02602938.2020.1805409

Cunningham, S., Laundon, M., Cathcart, A., Bashar, M. A., & Nayak, R. (2023). First, do no harm: Automated detection of abusive comments in student evaluation of teaching surveys. *Assessment & Evaluation in Higher Education*, *48*(3), 377-389. https://doi.org/10.1080/02602938.2022.2081668

Disability Discrimination Act 1992 (Cth). https://www.legislation.gov.au/Details/C2018C00125

Finney, T. G., & Finney, R. Z. (2010). Are students their universities' customers? An exploratory study. *Education + Training*, *52*(4), 276-291. https://doi.org/10.1108/00400911011050954

García-Morales, V. J., Garrido-Moreno, A., & Martín-Rojas, R. (2021). The transformation of higher education after the Covid Disruption: Emerging challenges in an online learning scenario. *Frontiers in Psychology*, *38*. https://doi.org/10.3389/fpsyg.2021.616059

Gibson, M. J., Luong, J., Cho, H., Moh, B., Zanin, S., Djatmiko, M., & Aandahl, R. Z. (2022). Where's the harm? Screening student evaluations of teaching for offensive, threatening or distressing comments. *Australasian Journal of Educational Technology*, *38*(2), 35–48. https://doi.org/10.14742/ajet.6133

Gillett-Swan, J., & Grant-Smith, D. (2020). Addressing mentor wellbeing in practicum placement mentoring relationships in initial teacher education. *International Journal of Mentoring & Coaching in Education*, *9*(4), 393-409. https://doi.org/10.1108/ijmce-02-2020-0007

Heffernan, T. (2021). Sexism, racism, prejudice, and bias: A literature review and synthesis of research surrounding student evaluations of courses and teaching. *Assessment & Evaluation in Higher Education*, *47*(1), 144-154. https://doi.org/10.1080/02602938.2021.1888075

Heffernan, T. (2023). Abusive comments in student evaluations of courses and teaching: The attacks women and marginalised academics endure. *Higher Education*, 85, 225-239. https://doi.org/10.1007/s10734-022-00831-x

James, C., Strevens, C., Field, R., & Wilson, C. (2019). Student wellbeing through teacher wellbeing: A study with law teachers in the UK and Australia. *Student Success*, *10*(3), 76-83. https://doi.org/10.5204/ssj.v10i3.1338

Kreitzer, R. J., & Sweet-Cushman, J. (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. https://doi.org/10.1007/s10805-021-09400-w

Kübler-Ross, E. (1973). *On death and dying*. Routledge.

Lakeman, R., Coutts, R., Hutchinson, M., Lee, M., Massey, D., Nasrawi, D., & Fielden, J. (2021). 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. https://doi.org/10.1080/02602938.2021.2012643

Lakeman, R., Coutts, R., Hutchinson, M., Massey, D., Nasrawi, D., Fielden, J., & Lee, M. (2022a). 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. https://doi.org/10.1080/02602938.2022.2060936

Lakeman, R., Coutts, R., Hutchinson, M., Massey, D., Nasrawi, D., Fielden, J., & Lee, M. (2022b). Playing the SET game: How teachers view the impact of student evaluation on the experience of teaching and learning. *Assessment & Evaluation in Higher Education*, 1-11. https://doi.org/10.1080/02602938.2022.2126430

Langley, P. (1996). *Elements of machine learning*. Morgan Kaufmann.

Larcombe, W., Finch, S., Sore, R., Murray, C. M., Kentish, S., Mulder, R. A., Lee-Stecum, P., Baik, C., Tokatlidis, O., & Williams, D. A. (2016). Prevalence and socio-demographic correlates of psychological distress among students at an Australian university. *Studies in Higher Education*, *41*(6), 1074-1091. https://doi.org/10.1080/03075079.2014.966072

Laundon, M., Cunningham, S., & Cathcart, A. (2023). Institutional approaches to evaluation of learning and teaching: A sector scan of Australasian universities. *Journal of Higher Education Policy and Management*, 1-18. https://doi.org/10.1080/1360080x.2023.2196646

Lee, M., Coutts, R., Fielden, J., Hutchinson, M., Lakeman, R., Mathisen, B., Nasrawi, D., & Phillips, N. (2022). Occupational stress in university academics in Australia and New Zealand. *Journal of Higher Education Policy & Management*, *44*(1), 57-71. https://doi.org/10.1080/1360080x.2021.1934246

Littleton, E. (2022). *At the crossroads: What is the post-COVID future of Australia's public universities?* Centre for Future Work, The Australia Institute. https://futurework.org.au/wp-content/uploads/sites/2/2022/04/At_the_Crossroads_-_FINAL_2.pdf

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., & Stoyanov, V. (2019). RoBERTa: A robustly optimized BERT pretraining approach. *arXiv preprint*. https://doi.org/10.48550/arXiv.1907.11692

Lloyd, M., & Wright-Brough, F. (2022). Setting out SET: A situational mapping of student evaluation of teaching in Australian higher education. *Assessment & Evaluation in Higher Education*, 1-16. https://doi.org/10.1080/02602938.2022.2130169

Matthews, K. E., Dwyer, A., Hine, L., & Turner, J. (2018). Conceptions of students as partners. *Higher Education*, *76*(6), 957-971. https://doi.org/10.1007/s10734-018-0257-y

NTEU. (2018). *Staff experience of student evaluation of teaching and subjects/units*. National Tertiary Education Union. https://issuu.com/nteu/docs/staff_experience_of_student_evaluat

Racial Discrimination Act 1975 (Cth). https://www.legislation.gov.au/Details/C2016C00089

Sex Discrimination Act 1984 (Cth). https://www.legislation.gov.au/Details/C2014C00002

Smele, S., Quinlan, A., & Lacroix, E. (2021). Engendering inequities: Precariously employed academic women's experiences of student evaluations of teaching. *Gender & Education*, *33*(8), 966-982. https://doi.org/10.1080/09540253.2021.1884194

Spooren, P., Brockx, B., & Mortelmans, D. (2013). On the validity of student evaluation of teaching: The state of the art. *Review of Educational Research*, *83*(4), 598-642. https://doi.org/10.3102/0034654313496870

Stallman, H. M. (2010). Psychological distress in university students: A comparison with general population data. *Australian Psychologist*, *45*(4), 249-257. https://doi.org/10.1080/00050067.2010.482109

Tangalakis, K., Kelly, K., KonYu, N., & Hall, D. (2022). The impact of teaching from home during the COVID-19 pandemic on the student evaluations of female academics. *Journal of University Teaching & Learning Practice*, *19*(1), 160-175. https://doi.org/10.53761/1.19.1.10

Taylor, D. (2017). A chronicle of just-in-time information: The secret to building first year university student wellbeing and resilience based on a three year initiative. *Student Success*, *8*(1), 35-47. https://doi.org/10.5204/ssj.v8i1.373

Taylor, D., Bearman, M., Scarparo, S., & Thomas, M. K. E. (2022). The emotional transition to online teaching: Grief, loss, and implications for academic development. *International Journal for Academic Development*, *27*(2), 176-190. https://doi.org/10.1080/1360144x.2022.2083141

Troiani, I., & Dutson, C. (2021). The neoliberal university as a space to learn/think/work in higher education. *Architecture & Culture*, *9*(1), 5-23. https://doi.org/10.1080/20507828.2021.1898836

van der Zanden, P. J. A. C., Denessen, E., Cillessen, A. H. N., & Meijer, P. C. (2018). Domains and predictors of first-year student success: A systematic review. *Educational Research Review*, *23*, 57-77. https://doi.org/10.1016/j.edurev.2018.01.001

Vidgen, B., Burden, E., & Margetts, H. (2021). *Understanding online hate: VSP regulation and the broader context*. The Alan Turing Institute. https://www.ofcom.org.uk/__data/assets/pdf_file/0022/216490/alan-turing-institute-report-understanding-online-hate.pdf

Williams, T. (2022). THE work-life balance survey 2022. *The Times Higher Education.* https://www.timeshighereducation.com/depth/work-life-balance-survey-2022

Please cite this article as:

Cunningham, S., Cathcart, A., & Graham, T. (2023). Allegations, abuse and discrimination: Using student evaluation of teaching surveys to support student and educator wellbeing. *Student Success*, *14*(3), 92-103. https://doi.org/10.5204/ssj.2756

This article has been peer reviewed and accepted for publication in *Student Success.* Please see the Editorial Policies under the 'About' section of the Journal website for further information.

Student Success: ***A journal exploring the experiences of students in tertiary education.***