CAULLT Grant Resource

Student Evaluation of Learning and Teaching:

Good Practice Guide

2023

Grant title and year of award:

Investigating how leaders in L&T units across Australian regional universities employ SET to ensure fairness and optimise their L&T benefits (2022)

Lead institution:

University of the Sunshine Coast

Project leader and contact details:

Associate Professor Gail Crimmins (gcrimmin@usc.edu.au)

Partner institution(s):

Charles Sturt University   University of Southern Queensland

Central Queensland University   University of New England

Southern Cross University

Cite this document as

Crimmins, G., Casey, S., & Weber, I. (2023). *Student Evaluation of Learning and Teaching Good Practice Guide*. CAULLT.

CAULLT is a non-profit organisation that is not registered for GST   ABN 48 669 069 291.

PO Box 4121, Maroubra South, NSW, 2035   www.caullt.edu.au   secretariat@caullt.edu.au

Acknowledgements

CAULLT is a non-profit organisation that is not registered for GST   ABN 48 669 069 291.

PO Box 4121, Maroubra South, NSW, 2035   www.caullt.edu.au   secretariat@caullt.edu.au

Student Evaluation of Learning and Teaching:

Good Practice Guide

Developed by Associate Professor Gail Crimmins, Dr Sarah Casey and Dr Ian Weber, based on research funded by the Council of Australasian University Leaders in Learning and Teaching (CAULLT).

Student Evaluation of Learning and Teaching:

Good Practice Guide

This Student Evaluation of Learning and Teaching: Good Practice Guide is specifically designed for leaders in learning and teaching in Australia. However, the main principles and guidance transcend the Australian context. We adopt a broad definition of who is considered a leader in learning and teaching within universities to include all staff working within universities who adopt the following roles:

Deputy Vice-Chancellor Academic

Pro Vice-Chancellor, Students

Faculty and School Deans

Associate or Deputy Dean (L&T)

Director of a Learning and Teaching Unit or Centre

Discipline Lead

Program Coordinator

Academic and Educational Developer

Principal or Senior Fellow of the Higher Education Academy

Student Evaluation of Teaching (SET) instruments are the most widely used assessment tool for collecting Higher Education student feedback to inform learning and teaching effectiveness, quality assurance and improvement, promotion and tenure, and recruitment processes. Responding to student feedback at the management and academic levels is designed to improve the student experience and raise student satisfaction based on instructional performance, capability, and competence.

Research on perceptions and experiences of academics and learning and teaching leaders involved in facilitating SET instruments informs the development of this Good Practice Guide and recommendations for using learning and teaching evaluation instruments in regional university contexts. However, the principles and recommendations are applicable outside of regional locations. Research outcomes allow the Council of Australasian University Leaders in Learning and Teaching (CAULLT) members to reflect on current assumptions of SET use within their institutional contexts and build knowledge around the changing responses in regional universities' student cohorts.

This research established a need to adjust and adapt SET practices for robust, more informed, improved student engagement and learning experiences using more meaningful and valuable data. Leveraging SET for continual improvement processes in learning and teaching is an opportunity for student-centred course improvements and, therefore, likely to be of interest to CAULLT members and general academic and faculty staff. We also suggest that CAULLT members may be able to apply the research learnings by considering the recommendations as part of the Good Practice Guide for their universities.

The guide thus considers:

- Why evaluate student learning and teaching (SET) in universities
- How SET is typically undertaken
- Limitations of typical SET practice
- Guidance on good practice to evaluate quality learning and teaching
- Good practice checklist.

Table of Contents

Why evaluate student learning and teaching in universities?

TEQSA Overview

TEQSA is Australia's national regulator of higher education, an independent agency established under the *Tertiary Education Quality and Standards Agency Act 2011* (TEQSA Act).

TEQSA regulations are designed to protect students' interests and the reputation and standing of Australian higher education. Its approach to quality assurance and regulation is or is based on:

- Standards-based, risk-reflective and transparent
- Positioned to promote and facilitate a culture of effective self-assurance as an integral part of a provider's operations
- Intervening only to the extent necessary to achieve our regulatory purpose
- A model of regulatory partnerships with individual providers and the sector overall.

Compliance frameworks TEQSA regulates

In addition to the TEQSA Act, it is primarily responsible for regulation by ensuring providers comply with the following:

- Higher Education Standards Framework, including the *Higher Education Standards Framework (Threshold Standards) 2021* (Threshold Standards)
- *Education Services for Overseas Students Act 2000* (ESOS Act) and associated instruments (ESOS Framework).

Compliance with the Threshold Standards

All providers must comply with the Threshold Standards to manage higher education activities and risks. This compliance includes matters such as the adequacy of facilities, staffing levels, support services, and academic and corporate governance.

Compliance with the ESOS Framework

All providers who deliver higher education to overseas students studying in Australia, including ELICOS and Foundation Programs, are expected to comply with the ESOS Framework.

Providers offering higher education courses to overseas students must be registered on the Commonwealth Register of Institutions and Courses for Overseas Students (CRICOS). CRICOS identifies both the providers and the courses that have been registered.

The ESOS Framework sets out the obligations relevant to these providers and focuses on the unique needs of overseas students studying in Australia (see the ESOS Act).

How TEQSA regulates the Higher Education sector

It regulates the sector through:

- Assessing risks to the sector via data collection and analyses
- Registering and re-registering providers
- Accrediting courses (where a provider does not have authority to self-accredit)
- Sharing information with the sector about how to improve and maintain compliance
- Monitoring compliance

- Working with providers to return to compliance and taking action to enforce compliance if necessary (see TEQSA - How we regulate)

TEQSA publishes a series of guidance notes to provide greater clarity for Higher Education providers in interpreting and applying selected standards. Guidance notes draw on TEQSA's regulatory experience and knowledge of experts in the Higher Education sector.

Guidance notes usually focus on a single topic (e.g., academic governance, integrity, leadership, quality assurance). The guidance note typically outlines the following:

- the nature of the topic
- the relevant standards in the Higher Education Standards Framework (HESF) that relate to the topic
- the underlying intent of those standards
- the risks to the quality of education if the topic is not addressed sufficiently by a provider
- the evidence that TEQSA is likely to look for to be satisfied that the requirements of the HESF are being met (see TEQSA – Guidance notes).

Assessment functions of SET

Accountability for outcomes of learning and effective delivery of information is an essential administrative responsibility and core mission of any academic institution. A strategy for assessment of curriculum content, course structure, and individual teaching effectiveness should be in place to meet an institution's academic obligations.

SET as a method of seeking feedback from learners and participants has the potential advantage of speed (in administration), anonymity and/or confidentiality (of response) and standardisation (for purposes of comparison between cohorts). The shortcomings can include poor response rate and validity of the outcomes if the survey is not designed with care (for purpose and focus) and if surveys are over-used (i.e., "questionnaire-fatigue"). Instructors might also implement a self-administered survey focusing on an innovative or new learning activity to help evaluate the success or areas for further development. These are often administered mid-term or mid offering of study unit; the findings from which can feed into the learning design of the remainder course of study.

Importance of evaluating the quality of learning as part of a quality assurance process

In the 1970s, the SET instrument became the primary method for formative assessment to improve and shape the quality of courses and measure the capability/competence of teachers and teaching practice (Hornstein, 2017; Kayas et al., 2022). Over the years, SET has evolved into a summative evaluation tool and the primary performance management tool and indicator for the promotion and tenure of academic staff in Australia, the United States, and Canada (see Heffernan, 2018a, b, c). In many cases, SET has become the sole performance indicator of teaching competence and effectiveness (see Berk, 2018; Galbraith et al., 2012; Spooren et al., 2013).

Recognition of students as critical stakeholders in universities

Students' learning experience and evaluation of teaching effectiveness constitute the root and source of student evaluation data, and meaningful and active participation of students is essential. The usefulness of student evaluation data is severely undermined unless students willingly provide quality input (Chen & Hoshower, 2003).

Resetting the role of student voice in shaping learning and teaching design

Adopting the SET tool became popular because of the potential perceived benefits for stakeholders – university administrators/managers, students, and academics/instructors. SET provides managers with insights to validate new programs and improve courses; measure academic performance (i.e., teaching capability/competence); provide academics/instructors with feedback to enhance courses and improve individual teaching practice leading to promotion and tenure (Johnson 2000); and empower students with a source of information to aid their choice of where and what to study (Collini, 2012) and a voice for their student experience, including dis/satisfaction with course design and content and academic teaching performance. However, Hornstein (2017) suggests that under the pressure of neo-liberalism, attempts to corporatise universities encourage students to view themselves as customers/consumers. However, the relationship between students and academics is not analogous to the customer/consumer covenant (Sproul, 2000). Furthermore, academics can and do view SET as a form or means of top-down surveillance with severe implications for their tenure or promotion. For example, Kayas et al. (2021) applied surveillance theory to analyse academic perceptions of SET implementation in four university business schools. The research found that top-down vertical surveillance, imbued with disciplinary procedures, involved university managers scrutinising academics through SET surveys. Emerging from such 'surveillance' is a power dynamic in which academics recognise the importance institutions and administrators place on responses and the crucial and sometimes abusive role student comments play in the process (see Cunningham et al., 2023; Heffernan, 2023; Kayas, 2021; Tucker, 2014). The research also identified how academics respond to the heavily weighted student comments and academics' engagement to disrupt or resist the effects of unfavourable, harmful, or abusive responses (Kayas et al., 2021).

In this context, forms of dissent and resistance can be thought of as a struggle against the management-imposed labour processes to subdue or minimise managerial encroachments through challenging or disrupting power assumptions, discourses, and relations in increasingly pervasive organisational contexts (Mumby, Thomas & Martí 2017). Resistance manifests through employees' asserting their individual and collective autonomy (i.e., behaviours) (Kayas et al., 2021; Sewell et al., 2011; Heffernan, 2020b, 2021, 2023). Such manifestations, Heffernan (2018c, 2020b, 2022, 2023) and Heffernan and Bosetti (2020) contend, have resulted in institutional hierarchical structures, fault lines, and rifts between stakeholders (i.e., university administrators, academics, and students) in the negotiation of habitus, fields, and accumulation of capital (i.e., social, economic, and political). As Heffernan (2020b) suggests, 'if habitus and capital are about the elements that create how someone became who they are, and what they might do in the future, … 'field', in its most basic form, whereby someone can enter a field, what their habitus and capital are worth and how it can be leveraged, is determined by the field they have entered' (see Bourdieu, 1988). Accumulation of capital and how it is leveraged within the field of teaching and learning is revealed in Steinhardt's et al. (2017) mapping of the quality assurance of teaching and learning in higher education. The review identifies 'antagonistic tensions' (i.e., resistance) between the education strand (i.e., academics) and management strand (i.e., university leaders and administrators). Negotiating and navigating such spaces of power and resistance has resulted in dissatisfied responses by academics to the intensification of their jobs and the detrimental effect it has on their mental health and wellbeing (see Heffernan, 2021, 2022, 2023; Heffernan & Bosetti, 2020).

How SET is typically undertaken

SET surveys are administered worldwide in many different formats. However, a recurring theme is using Likert scales to elicit responses, generally using a five-point scale.

Critical limitations of Student Evaluation of Learning and Teaching practice

SET used as a proxy for quality Learning & Teaching

SET assessment tools were intended primarily for *formative* purposes (Berk, 2005), that is, to improve and shape the quality of teaching. However, since the 1970s, SET has become the primary indicator for summative evaluation as the key performance indicator that decides staff promotion and tenure. In other words, SET has evolved into the dominant and, in many cases, the sole indicator of teaching competence (Berk, 2005; Spooren et al., 2013).

This evolution appears to have resulted from the ease with which data are collected, presented and interpreted. However, the interpretations are questionable on conceptual and statistical grounds. Few articles in the literature challenge the interpretation that student satisfaction ratings (SET) equate to teaching competence. Fewer still note that SET data, being categorical, cannot be evaluated validly using parametric statistics. Nonetheless, Hamermesh and Parker (2005) argue that SET is used for faculty tenure/promotion evaluations and pay determination regardless of whether the evaluations correspond to legitimate measures of underlying teaching quality/competence. Sproule (2002) argues that adjustment is necessary when student evaluations are used since the factors being assessed are not controlled by faculty. Sproule (2000) points out that while SET encourages students to view themselves as customers/consumers of education, the relationship between student and instructor is not analogous to the customer–client covenant, notwithstanding administrations' attempts to corporatize the university.

Lack of survey instrument validity, reliability, and application of data

SET assessment tools are often used in assessing faculty members' job performance and promotion and tenure decisions. However, the debate over this use of student evaluations has centered on the data's validity, reliability, and application in assessing teaching performance. Accordingly, SET implementation remains a delicate topic in higher education and education research literature. Many stakeholders are not convinced of SET's usefulness and validity/reliability for formative and summative purposes (Ching, 2018; Clayson, 2009; Spooren et al., 2013; Zhao et al., 2022).

Clayson (2009) argues that no consensus has been reached about the validity of the SET process, specifically the relationship between evaluations and learning, which should be validly correlated (see also Kreitzer and Sweet-Cushman's, 2021; Stark & Freishtat, 2014; Uttl et al., 2017). Clayson's et al. (2009) review found that attempts to find such a nomological relationship have been complicated by practice, methodology design, and interpretation of data. Further, Clayson's (2009) meta-analysis of the literature 'shows that a small average relationship exists between learning and the evaluations but that the association is situational and not applicable to all teachers, academic disciplines, or levels of instruction' (p. 16). A more significant concern is that the more objective the learning is measured, the less likely it is to be related to SET evaluations (Clayson, 2009). Ching's (2018) research also supports such conclusions, highlighting the complexity of the factors affecting SET design, implementation, interpretation, and the antecedents involved. As a result, 'no single perfect tool to adequately reflect what is happening in the classroom, specifically the effectiveness of teaching and learning with the

likelihood that different SETs should be designed for different disciplines, courses, and subjects' (Ching, 2018, p. 63).

Another analysis by Spooren et al. (2013) of the SET literature from 2000 indicates the research on SET has thus far failed to provide clear answers to several critical questions concerning the validity and reliability (i.e., the dimensionality debate, the 'bias' question, and questionnaire) of data derived from the evaluation tool. Zhao's et al. (2022) literature review of SET research also found multiple shortcomings in the student evaluation of teaching from purpose, indicator system, application of results, and process management. It concluded that within the 'academic circle', there is 'no unified opinion or standard on selecting the specific content of evaluation indicators, and fewer scholars have paid attention to the differentiation of the indicators for different courses' (Zhao, 2022, p. 9).

Reviews of the SET literature indicate that SET tools remain abstract and lack pertinence to the teaching and learning assessment improvement and resulting professional advancement. Jones' et al. (2014) research on SET validity raises essential issues for university administrators and academics to consider before interpreting and releasing survey results and using them summatively. The research explores relevant legal issues (namely, defamation, breach of the duty to take reasonable care for an employee's welfare, breach of the duty of trust and confidence, breach of the right to privacy and, if the lecturer is forced to resign because of such infringements, constructive dismissal). Possible litigation has considerable implications for exacerbating Steinhardt's et al. (2017) notion of educative-management 'antagonistic tensions' and consequential and ongoing deterioration of the employer-employee relationship of trust whilst continuing to employ SET assessment for quality assurance and learning and teaching purposes.

Gendered biases in responses

One of the critical issues emerging within the research literature is the presence and impact of gender bias on SET results. According to social psychological theory, gender biases in SET assessment may occur because of a lack of fit between gender stereotypes and individuals' professional roles, leading to negative evaluations (see Kwok & Potter, 2022; Renström et al., 2021). Eagly and Karau's (2002) congruity theory of prejudice suggests that gender bias and prejudice relate to socially constructed values and behaviours and the perceived incongruity between gender, leadership, and occupational roles. Such communally held assumptions and expectations embody socially constructed beliefs about men's and women's traits and behaviours. People tend to connect women with concern for the welfare of others while associate men with agentic characteristics, such as assertive, controlling, and confident tendencies (Eagly & Karau, 2002). Such expectations also tend to emerge in strongly patriarchal beliefs in leadership roles. As a result, female leaders face prejudice in two ways: 1) women's leadership potential is evaluated less favourably than men's, and 2) female leaders receive less favourable evaluations because expected leader behaviours conflict with expected female behaviours (Eagly & Karau, 2002).

Drawing on Biernat and associates' (2003; Biernat & Eidelman, 2007; Biernat, Fuegen & Kobrynowicz, 2010; Biernat & Vescio, 2002) shifting standards theory, which proposes that evaluative standards can change due to stereotyping effects, Kwok and Potter's (2022) research found that gender stereotyping contributed to students being more likely to nominate teachers of the same gender, but also that male students were disproportionately less likely to nominate a female teacher. The study concluded that student conceptions generally conformed to gender biases, particularly for male students, with students' perceptions of high-quality teaching inextricably linked to sociocultural influence (Kwok & Potter, 2022). The results support Khazan's et al. (2019) investigation into the relationship between bias, prejudice, and gender in student evaluations of an extensive, asynchronous online course. Khazan et al. (2019) argue that such bias and prejudice make it more difficult for women to successfully secure leadership roles and recognition as effective leaders. Similarly, such expectations for gender-typical occupational roles translate to education disciplines and the classroom. For example, female faculty in male-dominated disciplines 'face prejudice like that experienced by female leaders, resulting

in similar consequences. The incongruity of these perceived roles can result in fewer women entering male-dominated disciplines, and women who do enter these disciplines receive poorer performance evaluations' (Khazan et al., 2019, p. 423).

Other contemporary literature details the ongoing influence of gender bias against female academics (see Boring et al., 2016; Kreitzer and Sweet-Cushman, 2021; Boring, 2016; Mitchell & Martin, 2018; Renström, 2021; Shreffler et al., 2019; Stark & Freishtat, 2014; Uttl et al., 2017). As far back as the 1980s, research indicated the presence and impact of gender bias in student evaluations of teaching responses (Bennett, 1982). A more contemporary examination of the literature on bias in student evaluations by Kreitzer and Sweet-Cushman (2021) from 2000 produced a novel dataset of over 100 articles, providing a nuanced review of this broad but established literature. The review indicated that researchers and scholars, using different data and methodologies, routinely identify women faculty and other marginalised groups who face significant biases and are thus disadvantaged in SET responses.

Other researchers concur, highlighting how instructor gender and ethnicity, students' gender, and grade expectations influenced SET responses (see Boring, 2017; Boring et al., 2016; Shreffler et al., 2019; Stark & Freishtat, 2014). These findings raise concerns about the accuracy of SET in measuring teaching effectiveness and student learning (Stark & Freishtat, 2014; Uttl et al., 2017). Boring's (2017) research also revealed how students scored women lower than men for the same level of teaching effectiveness, noting that bias may be context-dependent. Shreffler et al. (2019) point to gender as one of three factors (in addition to work-family conflict and perceived organisational support) disproportionally impacting career advancement to student impressions and expectations of them.

Research by LaPaglia et al. (2002) focused on student perceptions of the instructor's teaching competency in facilitating lecture delivery on student learning and instructor ratings. The study found that participants rated the female speaker significantly lower than the male speaker, but only when the speaker was disfluent. The results concur with Renström's et al. (2021) study of how student comments on feminine or masculine behaviours led to gendered evaluations of the lecturer. Lecturers displaying feminine behaviours were expected to be more approachable, while lecturers displaying male behaviours were instead perceived as being more competent, better pedagogues, and leaders. Researchers conclude that the results 'should not be used as sole indicators of the pedagogic ability of a lecturer for promotion and hiring decisions because they may be gender-biased (Renström et al., 2021, n.p.).

Longitudinal studies comparing quantitative and qualitative student responses in SET assessment revealed gendered characteristics and gender bias previously invisible in statistical results. Sigurdardottir et al. (2023) analysed five years of SET responses for gendered communication characteristics between students and teachers in an Icelandic higher education setting. The quantitative results showed that male students rated female teachers lower than male ones. Differences were found in qualitative student comments on male teachers' subject knowledge capabilities, while comments on female teachers related to student service and relatability (Sigurdardottir et al., 2023, p. 954). Gelber's et al. (2022) analysis of four years of SET responses, initially reported in quantitative form, asked whether the same evaluations would produce different results when analysed qualitatively rather than quantitatively and do students evaluate male-identified and female-identified teachers differently, and if so, what are the differences? The qualitative results revealed gender bias that 'is invisible in quantitative analysis' (Gelber et al., 2022, p. 199). Female-identified staff were evaluated by students more positively than their male counterparts for undertaking time-intensive, stereotypically feminine, emotional labour. In contrast, male-identified staff were assessed more positively for their technical expertise and teaching style. Gelber et al. (2022) concluded that SET assessment evaluates gender-stereotypical behaviour rather than teaching quality, which has significant implications for their use in universities.

Impact of SET on academic appointments/promotions and mental health and wellbeing

Research indicates concerns and the impact of negative student comments in SET assessment by highlighting the use of offensive, unprofessional, abusive and malicious language, exhibiting homophobic prejudice, gender bias, sexism and racism in Australian universities (see Cunningham et al., 2023; Heffernan, 2022, 2023; Tucker, 2004). Tucker's (2014) research manually assessed more than 30,000 student comments to identify offensive and unprofessional responses within SET. The study found that 'most students do not abuse the privilege of giving anonymous feedback' (p. 347). However, the research also recognised the need to address negative comments through a process of educating or training students and teachers in appropriate and professional ways of working together by providing professional feedback that improved the student experience, teaching and learning effectiveness and support through mentoring teachers in their academic careers (Tucker, 2014). However, later research by Heffernan (2022) found that increasing abusive comments were mainly directed towards women and those from marginalised groups, making student surveys a growing cause of stress and anxiety for academics. The research indicates that SET assessment is influenced heavily by student demographics, teaching academic culture and identity, and other aspects not associated with course quality or teaching effectiveness. Heffernan (2022) argues that student evaluations that 'openly prejudiced against the sector's most underrepresented academics … contribute to further marginalising the same groups universities declare to protect, value and are aiming to increase in their workforces' (p. 199).

Heffernan's (2023) ongoing research on the volume, type, and impact of anonymous student comments on 674 academics highlights the volume and type of abusive comments academics receive more widely. The research outlines how women and marginalised groups received the highest volume, most derogatory, and most threatening abuse. Equally troubling is that the research underestimates the magnitude and severity of abusive comments that academics receive. Other research by Cunningham's et al. (2023) focuses less on the volume of malicious and abusive comments and more on the impact of the comments on the mental health and wellness of academic staff who receive the responses. Under the 'first, do no harm' theme, the research documented a machine-learning (supported by a final human checkpoint) approach to screening 100,000 student comments in 2021. This project has prevented 100 abusive comments from reaching academic staff members. Some might dismiss the small percentage of unacceptable comments made by students as minimal and thus not significant (statistically). However, Cunningham et al. (2023) suggest monitoring and reporting to ensure students know when their comments are unacceptable to the university' (p. 387).

Under typical workplace protocols, harassing, threatening or abusive language, including profanities and language that intimidates or discriminates (i.e., sexual orientation, gender identity, intersex status, disability, ethnicity, marital status, nationality, age, religion and/or political persuasion) is deemed inappropriate and unprofessional (see Cunningham et al., 2023; Tucker, 2014). In a university context, such protocols relate to personal attacks on staff related to appearance or other matters unconnected to courses, teaching, or learning experiences (Cunningham et al., 2023). It must be remembered unmonitored malicious and abusive language causes harm to the wellbeing and career prospects of academics, particularly women and other minority groups (see Adams et al., 2022; Cunningham et al., 2023; Heffernan, 2022, 2023). Cunningham et al. (2023) and Heffernan (2022, 2023) also remind universities and administrators that they have a duty of care to protect educators' wellbeing from emotional trauma and psychological harm.

Moment-in-time assessment

"Single-moment-in-time" refers to how learning is evaluated on a specific date or within a designated time, often at the end of a defined teaching phase for assessing competence up to that point. The high-stakes nature describes how assessment outcomes substantially affect the individual learners involved, with limited retake options in the proximal period. Written and oral formats are primarily used to measure knowledge. In contrast, objective structured examinations have been used to assess behavioural competencies, such as professionalism and communication, in addition to knowledge and skills (Sidhu & Fleming, 2023).

Good practice guide to evaluate quality learning and teaching in universities

Research indicates a range of strategies can be used by institutions and faculty to reduce measurement bias (i.e., validity and reliability) and equity bias (i.e., gender) in student evaluations of teaching, personnel reflections (i.e., responses to student comments) and professional decision-making (i.e., hiring, promotion, tenure). Despite the developing nature of testing interventions and strategies to mitigate biases, emergent research indicates some promise. Several studies advocate multiple sources of evidence to evaluate teaching (Berk, 2018; Kreitzer & Sweet-Cushman, 2021). For example, Berk (2018) suggests a broad range of alternatives beyond student ratings or SET in the delicate and potentially career-determining decision-making processes exist. Such alternatives include observations (Miller & Seldin, 2014; Hornstein, 2017), instructor portfolios (Seldin et al., 2010), internal and external review of course material (Chism, 2007), and focus and nominal group technique (Varga-Atkins et al., 2017). However, such current and future interventions require substantially more rigorous testing to make such assessments more feasible, reliable, fair, and equitable. As Kreitzer and Sweet-Cushman (2021) argue, 'more caution should be taken in the use of SETs in hiring, tenure, and promotion decisions and alternatives assessments of teaching should be further utilized' (p. 80).

In selecting data-gathering methods for teaching effectiveness, it is essential to account for the possible inherent bias of each source. As Esarey and Valdes (2020) argue, selecting multiple data-gathering approaches should avoid those that systematically bias the results in the same way. This approach ensures whichever combination of sources is triangulated to minimise the weaknesses inherent in each process or, in the case of SET, a single-method approach. The research also notes that unbiased, reliable, and valid student evaluations can be unfair. Kreitzer and Sweet-Cushman (2021) support using multiple data collection methods around student evaluations of courses and teaching, although alternatives are more laborious, time-consuming, and cost-ineffective than survey methods.

Recommendations for adjustments to typical SET practice

Research on intervention strategies employed to address the limitations of SET as solely a summative assessment tool for measuring the effectiveness of teaching include:

1. Eliminating omnibus items related to 'overall teaching effectiveness' and 'value of the course' from teaching evaluations, as the results are misleading (Hornstein, 2017).
2. Avoiding averaging or comparing averages of student rating scores, as these averages do not result in statistical reliability and validity of results. Instead, report the distribution of scores, the number of responders and the response rate to validate the significance of the results (Hornstein, 2017).
3. Avoiding extrapolating student evaluations to the entire class because extrapolation is unreliable when response rates are low in smaller cohorts of 15 or fewer students (Hornstein, 2017).
4. Avoiding comparing teaching in courses of different types, levels, sizes, functions, or disciplines (Hornstein, 2017).

Recommendations for processes that might replace or augment SET

1. Supplement student evaluations of teaching with pre–and post-learning measures that examine the relationship among student grades, learning (as measured by the pre– and post-assessment), and student evaluations of teaching (Stark-Wroblewski et al., 2007).
2. Employ mid-semester course evaluations to generate rich, high-quality, elaborate student feedback for enhancing learning and teaching, with anonymity and confidentiality of the process encouraging more reflective practice for students and minimising stress for instructors (Sozer et al., 2019).

Feedback literacy for students

1. Educate students on what constitutes teaching effectiveness and teachers in appropriate and professional ways of working together by providing professional feedback that improves the student experience in teaching and learning and supports and mentors teachers in their academic careers (Tucker, 2014).
2. Improve the effects of SET results by combining evaluation with individual counselling (non-education) in an institutional development approach (Rindermann et al., 2007).
3. Combine focus group and Nominal Group Technique methods for enhancing quality assurance through issue exploration, data analysis, and increased student ownership (Varga-Atkins et al., 2017).

Socio-emotional support for academics

1. Pay careful attention to student comments to understand their scope and limitations. Students are the authorities on their experiences in class but are not well situated to evaluate teaching competence and effectiveness pedagogically (Cunningham et al., 2022; Heffernan, 2022, 2023).
2. Employ machine learning to monitor, identify and screen malicious and abusive students to prevent academic emotional trauma and psychological harm (Cunningham et al., 2023).

Opportunity for academics to respond to SET reports (if used as part of staff appraisals)

1. Utilise the virtual learning environment (i.e., virtual worlds, gaming, simulation) to encourage faculty reflection and improve the student learning experience, engagement, and retention (Winchester & Winchester, 2012).
2. Adopt an automated visualisation methodology of students' free text comments from course satisfaction surveys, focusing on sentiment to reveal learning and teaching aspects of the course that may require improvement or are performing well and providing educators with a simple, systematic way to monitor their courses and make pedagogically sound decisions on teaching strategies (Cunningham-Nelson et al., 2019).
3. Use teaching or instructor portfolios as a course and teaching review process component (Seldin et al., 2010).
4. Employ external and internal assessors to observe classroom interactions as part of milestone reviews (Chism, 2017; Miller & Seldin, 2014).

Data to be used alongside a suite of other indicators

Quality Indicators for Learning and Teaching (QILT) includes a suite of higher education surveys covering the student lifecycle from commencement to employment. QILT provides the government and the sector with robust, nationally consistent performance data to uphold and drive quality improvement. The Social Research Centre administers the QILT surveys on behalf of the Australian Government Department of Education. Data collected through the QILT surveys drives the ComparED website.

Prospective students can use ComparED to explore and compare Australian higher education institutions and study areas based on the real-life experiences of current students and recent graduates (Australian Government, 2023).

1. **Student Experience Survey (SES):** SES is Australia's only comprehensive survey of current higher education students. It focuses on aspects of the student experience that are measurable, linked with learning and development outcomes, and potentially able to be influenced by higher education institutions. Information collected in the SES helps higher education institutions and the government improve teaching and learning outcomes for students.
2. **Graduate Outcomes Survey (GOS):** GOS is completed by graduates of Australian higher education institutions approximately four to six months after finishing their studies. The survey measures short-term employment outcomes, including skills utilisation, further study activities, and graduate satisfaction.
3. **Graduate Outcomes Survey – Longitudinal:** GOS-L is completed by graduates of Australian higher education institutions approximately three years after completing their studies. The survey supplements the GOS by measuring graduates' medium-term employment outcomes and further study activities.
4. **Employer Satisfaction Survey (EES):** ESS is the only national survey that measures how well graduates from Australian higher education institutions meet employer needs. It uses a unique methodology to link graduates' experiences to their direct supervisors' views. ESS data are used to better understand the specific skills and attributes needed in business today, how well higher education prepares graduates for the workforce, and the varied employment pathways graduates take after completing their studies (Australian Government, 2023; see QILT resource site).

Good practice for evaluating learning and teaching feedback: A checklist

Evaluation of teaching and learning involves collecting evidence from various stakeholders to improve the effectiveness of the teaching-learning process. A successful evaluation generates valid, reliable, and action-oriented outcomes for improving and developing teaching and learning. Research prioritises **seven critical questions** when considering the practical issues of evaluating teaching.

What is the purpose of the evaluation?

Good practice: clarify the purpose, or purposes, of the evaluation to all stakeholders. The evaluation should centre on the following:

- **Quality of the educational provision** (the product) - which could be the whole program, a course (module), a class (lecture, seminar, laboratory, etc).
- **Performance of the provider**(s) - the academic staff, tutors, and support staff involved in delivering this program/course/class.
- **Experience of the learners** as partners in the process - their experience with the provider, what is provided, and their motivation and approach to learning.
- **Combination and clarity of factors** - provided the various purposes are clearly and comprehensively stated to evaluation participants (Vanderbilt University, 2023).

What is the key focus of the evaluation?

Good practice: make clear to participants the critical focus of an evaluation. Are multiple areas focused on in the evaluation? Organise these questions into clusters, identifying the focus of each cluster. Avoid too many focus areas, as this will confuse the evaluation. Help the evaluator to concentrate their work on a maximum of three key questions and provide you with a constructive and reflective response for each one.

For example, you might want to know about these areas by asking these questions:

Area of practice to be considered	Questions you could ask
Learning experience	Are the aims of the session and the learning outcomes clearly stated at the outset of learning and aligned with the learning and teaching activities?
	Is the curriculum content meaningful, relevant and pitched at the appropriate level for the course and learners' existing knowledge level?
	Are learners motivated and actively engaged in learning? Are they attentive and participating when required?
	Is the content presented effectively and engagingly, employing various learning and teaching methods?
Evaluation of learning	Is the assessment method clear, transparent, and valid?
	Are the assessment criteria accessible and at the correct QILT?
	Are the assessment criteria 'constructively aligned' with the learning and teaching activities and the intended learning outcomes or objectives?
	Is the quality of feedback appropriate and linked to improving learner performance?
Curriculum	Is the curriculum challenging enough to maintain learners' interest?
	How reasonable is the workload involved?
	Does the curriculum develop skills, knowledge and experience relevant to the program and individual professional development?
	Do learners take advantage of support and resources?

(University of Exeter, 2023)

Who will be asked to make the evaluation?

An evaluation of teaching and learning is typically designed for learners as the primary participants in the learning experience. However, there is a significant advantage in seeking evaluation by others. Feedback from colleagues and other staff, as well as from learners, allows for the triangulation of different perspectives, adding to the reliability and validity of the outcomes of the evaluation process. Such correlation provides instructor insight into the level of harmony or disharmony of perceptions between the stakeholders in the teaching-learning process. Internal and external peers can be invited to participate in the evaluation, thus introducing a broader perspective on the academic standards of educational provision and teaching effectiveness.

Good practice: correlate evaluation outcomes from different viewpoints (data) wherever possible.

How do we motivate students to complete evaluations and provide useful feedback?

Good practice: talk with students about the importance of course evaluations and how those evaluations are used to improve learning and teaching effectiveness.

- Advise students that you value their honest and constructive feedback to improve course learning and teaching. Share examples of how you have changed your courses because of student feedback, if possible.
- Let your students know you are interested in positive and negative feedback on the course. What aspects of the course and/or instruction helped them learn? What aspects might be changed to help future students learn more effectively?
- Describe the feedback you find most useful. In most cases, specific feedback with examples is more valuable than general statements. If your university does not have a "Providing helpful feedback to your instruction" handout, consider developing the resource to support constructive feedback.
- Remind students that evaluations are usually designed to be anonymous or confidential (so that their identities are not linked with any evaluative marks or comments they make), and results cannot be viewed until after final grades have been submitted. Many students don't understand the process of SET implementation and review.
- Advise students that you are the primary audience for their feedback but that others will potentially read their evaluations, including department and school administrators and managers. Course evaluations are critical in personnel evaluations, curriculum design, planning, and revision. Whilst this may, potentially impact feedback response rates, it might support clearly considered responses.
- Consider language usage in your syllabus that addresses student evaluations. This usage alerts the students to pay attention to their learning experiences throughout the semester and makes them more mindful of their responses in the course evaluations.

How do we make sense of SET feedback?

One of the most challenging tasks for faculty who look at their end-of-semester student evaluations is interpreting what students say in written comments. Many instructors may be tempted to dismiss the critical information these comments provide about their teaching and students' learning because they feel students do not know enough to judge their teaching and "these written comments show just how unreliable they are!". On the other hand, instructors also say that they get more information from student written comments than from the scaled items typically found on SET forms (Lewis, 2001).

Instructors can assess their teaching effectiveness when they receive student evaluations during mid-semester and the end of the semester. Making sense of student feedback can be challenging, so we offer tips for examining evaluations and using the information to improve teaching effectiveness.

When considering student evaluations:

- Choose a good time to review the feedback in which you have enough time to digest the information, have privacy, and ensure you establish mental 'space' to analyse the information.
- Track and compare quantitative results against your stated teaching goals and objectives. Consider how each item's summary rating scores relate or fit with your teaching goals and your department or school's stated expectations for teaching effectiveness.
- Identify patterns in students' comments—identify trends, noting what has been done well and what needs improvement.
- Reflect on your experience and account for this in your response to feedback. Suppose you are new to teaching, the school, or even the course. In that case, you may still be learning about various aspects of being an instructor, such as course design, teaching skills, student interaction and motivation, and departmental or school expectations.
- Account for the context and characteristics of your course. Research shows that student evaluations often are more positive in courses that are smaller rather than larger and elective rather than required. Also, instructors receive more positivity in courses in which students do well.

When dealing with negative student feedback:

- Remember that almost *all* faculty members, including senior and highly successful academics, receive negative feedback at some point in their teaching careers.
- Allow yourself to acknowledge that student comments can be hurtful and make you angry but may point to critical areas for pre-SET implementation and continued teaching development.

When deciding how to further your development as a teacher:

- Consider the most frequently mentioned areas for teaching improvement in analysing student evaluations within and across universities related to 1) more transparent, more specific in-class communication and 2) more transparent, more explicit organisation of course content.
- Consider scheduling an appointment with the course convenor/coordinator, senior academic mentor, or university learning and teaching support unit to help interpret the evaluation results. Research suggests that teachers who consult with someone about evaluation results are more likely to score higher on subsequent SET responses than those who do not discuss the results.

When planning steps to improve the feedback you receive in evaluations, consider the following options:

- Use "a moment in time" evaluations after selected class sessions, asking students to note the main idea learned in class, two ideas about a significant concept, theory, or construct, a question about content, and so forth.
- Give a self-administered "midterm evaluation" of the course (in addition to the official university SET) to check how the class is progressing. You can use the information to make changes.
- Discuss the interim mid-term feedback with the class and explicitly implement one of the student suggestions.
- Before the final course evaluation, explain to the class the importance you place on their constructive input to teaching effectiveness.

Good practice: Course evaluations should be thought of as a part of a larger classroom narrative, one that focuses on improving students' learning experiences from beginning to end along intertwined paths: improving the student experience, providing feedback and improving teaching effectiveness (Vanderbilt University, 2023).

Who will see the evaluation outcomes, and who will act upon them?

Good practice: make clear, from the outset to all involved in the evaluation process who will see the evaluation report, who will be responsible for actioning the results, and who will monitor the effectiveness of this action. Avoid undertaking an evaluation of matters where there is no realistic prospect of any action following the outcomes of the SET implementation.

Before designing an evaluation, it is necessary to define:

- How and by whom the raw data will be processed to generate outcomes?
- To whom the outcomes will be reported, and in what detail and form?
- Who bears responsibility for acting on the outcomes?

Both the evaluator and the evaluated, as well as the person(s) responsible for managing the quality of the educational provision, must have access to the evaluation outcomes. This approach should include the Head of Department and Academic Lead or PDR Reviewer. Use the Annual Review of Teaching proformas to ensure that you provide an appropriate level of detail.

What methods of evaluation are available to support staff promotion and tenure review?

Good Practice: Provide clear guidelines on how to best document instructors' demonstrated teaching effectiveness in preparation for promotion and tenure review.

Applying the pre-post method approach to measuring teaching effectiveness presents a viable method of supplementing SET scores with learning outcome measures in an efficient manner that provides several instructional benefits of providing students with a sample of exam or task questions and a quick overview of the course, while simultaneously providing instructors with a pre-course assessment of students' familiarity with and knowledge of the subject matter (i.e., diagnostic evaluation).

Stark-Wroblewski et al. (2007) offer an example of a pre-post assessment approach. Students are asked to complete a pre-test on the first day of class. These items are also used to (a) introduce students to the course by illustrating topic areas that will be covered during the semester and (b) illustrate the types of questions that will be included on exams administered in the course (adaptable to non-exam courses). Post-test scores are calculated by selecting scores for the relevant items across each of the five exams and then summing these scores at the end of the semester. In this way, the post-test is essentially 'embedded into' exams that students complete in class, allowing the instructor to examine student learning without usurping valuable class time. At the end of the semester, pre-and post-test scores are compared to assess student learning in the course. Additionally, individual instructors can present a standard metric, in the form of a *d*-score, as an indicator of student learning.

In the absence of clear, agreed-upon guidelines for documenting teaching effectiveness for promotion and tenure processes, individual faculty members are encouraged to consider supplementing SETs with measures of student learning. Preparation of documentation for promotion and tenure review by faculty members should include pre-post assessment methods to measure student learning, along with SETs to document teaching effectiveness more convincingly and comprehensively.

Some points to consider

Who should design the survey? The purpose of the evaluation determines the answer and is, most commonly, the person(s) responsible for delivering the education under evaluation. It is *good practice* to seek the views of the intended evaluators and teaching teams on their suitability for the evaluative purpose

Should the questions be designed for the response on a rating scale (e.g., a positive statement with the response on a scale of "agree" to "disagree")?

Rating scales lend themselves to rapid processing and ease of comparability across cohorts/years, especially when using online survey tools, but they limit the range of responses.

Should the questions be designed for freeform response (e.g., "What did you find most challenging about …")?

Freeform responses allow a more subtle range of responses and raise issues beyond those in the questionnaire. However, they take longer to complete, longer to process and much longer to report. A *good compromise* is a questionnaire mainly of rating-scale format (for speed and consistency) with some opportunities for freeform response.

It is ***good practice*** for the processing and reporting by a third party who is not closely associated with or involved in course evaluation.

How long should the survey instrument be?

So that the purpose and focus remain clear, it is **good practice** to keep a survey short - 10 questions would allow for a rating-scale format, but less if the questions allow freeform student responses.

When should it be administered?

The timing of the SET depends on the purpose (i.e., mid-semester and end-of-semester). For example, evaluation on completion of the module provides a more comprehensive picture. However, it is too late for that cohort to benefit from the information - evaluation part-way through the module or after individual classes (i.e., instructor designed and initiated) gives an incomplete picture but would enable the instructor to adjust instructional strategies or content to benefit that student cohort. The purpose and focus of the evaluation also determine the best frequency of administration. However, it is unwise to overload to the extent that "survey fatigue" impacts student motivation to participate in SET responses (i.e., students complete the SET for each course they participate in).

It is ***good practice*** for a department or school to plan the evaluation schedule with a higher frequency of evaluation where there is cause for concern and a lower frequency where evaluation provides stable, positive outcomes.

Structured focus group (Nominal Group Interview technique)

The nominal group technique is a meeting with learners or participants where they are asked to give their views about a program, course, or class. The meeting must be planned and structured carefully to generate constructive debate so that learners/participants feel free to express their views without personal risk. Typically, learners are asked to work in small groups to reflect upon the positive and negative features of the educational provision (e.g., program, course, or module), its delivery, and their performance and learning experience. A spokesperson from each group relays the considered views of the focus group to the meeting. The role of the member of staff leading the meeting is to compile a summary of such views, to validate them at the meeting, and, later, to produce a short report of the primary outcomes. It is advantageous for this person to be someone from outside the department and school or teaching team to support anonymity and provide a safe environment for learners to express their views honestly. Learners who take part should receive a copy of the written report.

A critical benefit of this structured group interview is the provision of learners with greater freedom of expression than a SET survey and more opportunity to make constructive suggestions for improvement. It typically requires a meeting of about an hour, but the processing is done during that hour, and the time needed for producing a report is short.

It is ***good practice*** for a group interview to be led by an experienced leader who is not involved in the delivery of the educational provision being evaluated, and preferably not in the same department or

school. However, this provision depends on staff availability and the time required to complete the focus group and reporting processes.

Student-Staff Liaison Committee (SSLC)

Academic departments or schools should incorporate a Student/ Staff Liaison Committee (SSLC) into its learning and teaching approach. This committee provides an opportunity for learners and staff to partner in the teaching/learning process and:

- Identify priorities for scheduling evaluations,
- Propose purposes and the focus of an evaluation,
- Comment on the outcomes of evaluations, their validity and reliability, and
- Explore the following steps and options for action.

It is *good practice* for SSLC meetings to include a default agenda item on student evaluation of learning and teaching to legitimise topics raised at the meeting. Points raised at SSLC can be compared with outcomes of other evaluation methods (University of Exeter, 2023; see example of a Code of Practice for SSLCs).

Self-evaluation and peer evaluation

The course/module team can undertake the same evaluation the learners undertake (i.e., complete the same SET survey or conduct a self-evaluation using the same format of a structured Nominal Group Interview). Harmonising staff responses and the outcomes of the learners' evaluation indicates staff awareness of learner perceptions. Items of disharmony suggest that staff understanding of learners' needs and concerns needs to be addressed before attending to issues emerging from the evaluation.

Regular peer evaluation of the teaching/learning process to support self-evaluation is a valuable activity. A colleague focusing on the process that is taking place adds a dimension of evaluation that may escape the members of staff and the learners, who are generally too busy with the business of teaching and learning to observe the process itself. However, casual 'dropping-in' on a class is not the best approach to peer evaluation. Both self-evaluation and peer evaluation can align with and add value to other forms of evaluation of teaching.

It is **good practice** for peer evaluation to be a planned and structured process involving the separate stages of:

- **A briefing meeting** should set the context and identify the aims and learning outcomes of the process to be observed,
- **An observation** should include a checklist of critical features to monitor, agreed upon at the briefing meeting
- **A debriefing meeting** should provide feedback on the observation and compare the perceptions of the observer and observed perceptions to enhance clarity around teaching effectiveness.

Reference material

Australian Government. (2023). Quality Indicators for Learning and Teaching (QILT). Available at https://www.qilt.edu.au (accessed 20 October 2023).

Adams, S., Bekker, S., Fan, Y., Gordon, T., Shepherd, L., Slavich, E., & Waters, D. (2022). Gender bias in student evaluations of teaching: 'Punish[ing] those who fail to do their gender right'. *Higher Education*, 83(4), 787-807. 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/S104909650004340

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

Ball, K. (2010). Workplace surveillance: An overview. *Labor History*, 51(1), 87-106.

Bennett, S.K. (1982). Student perceptions of and expectations for male and female instructors: Evidence relating to the question of gender bias in teaching evaluation. *Journal of Educational Psychology*, 74(2), 170-179.

Biernat, M. (2003). Toward a broader view of social stereotyping. *American Psychologist*, 58(12), 1019–1027. https://doi.org/10.1037/0003-066X.58.12.1019

Biernat, M. & Eidelman, S. (2007). Translating subjective language in letters of recommendation: The case of the sexist professor. *European Journal of Social Psychology*, 37, 1149-1175. https://doi.org/10.1002/ejsp.432

Biernat, M., Fuegen, K. & Kobrynowicz, D. (2010). Shifting standards and the inference of incompetence: Effects of formal and informal evaluation tools. *Personality and Social Psychology Bulletin*, 36(7), 855–868. https://doi.org/10.1177/0146167210369483

Biernat, M. & Vescio, T.K. (2002). She swings, she hits, she's great, she's benched: Implications of gender-based shifting standards for judgment and behavior. *Personality and Social Psychology Bulletin*, 28(1), 66-77. https://doi.org/10.1177/0146167202281006

Benton, S., Cashin, W. & Manhattan, K. (2012). *Student Ratings of Teaching: A Summary of Research and Literature*. IDEA Center. http://www.ideaedu.org

Berk, R.A. (2018). Start spreading the news: Use multiple sources of evidence to evaluate teaching. *Beck's Law*, 32(1), 73-81.

Boring, A. (2017). Gender biases in student evaluations of teaching. *Journal of Public Economics*, 145, 27-41. http://dx.doi.org/10.1016/j.jpubeco.2016.11.006

Boring, A., Ottoboni, K. & Stark, P. (2016). Student evaluations of teaching (mostly) do not measure teaching effectiveness. *Science Open Research*. https://doi.org/10.14293/s2199-1006.1.sor-edu.aet-bzc.v1

Bourdieu, P. (1988). *Homo Acadmicus* (P. Collier, Trans.). Cambridge, UK: Polity Press.

Bourdieu, P. (1989). Social space and symbolic power. *Sociological Theory*, 7(1), 14–25.

Bourdieu, P. (1996). *The State Nobility. Elite Schools in the Field of Power* (L.C. Clough, Trans.). Cambridge, UK: Polity Press.

Bourdieu, P. (2006). The forms of capital. In Lauder, H., Brown, P., Dillabough, J.A. & Halsey, A.H. (Eds), *Education, Globalisation and Social Change* (pp. 105–119). Oxford, UK: Oxford University Press.

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.1310

Braun, V. & Clarke, V. (2006) Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77-101.

Braun, V. & Clarke, V. (2023). Thematic analysis. In Denzin, N.K., Lincoln, Y.S., Giardina, M.D. & Canella, G.S. (Eds), *The Sage Handbook of Qualitative Research* 6th ed. (pp. 385-402). Los Angeles, CA: Sage Publishing,

Chen, Y. & Hoshower, L.B. (2003). *Student Evaluation of Teaching effectiveness: An assessment of student perception and motivation*. Assessment & Evaluation in Higher Education, 28(1), 71-88.

Ching, G. (2018). A literature review on the student evaluation of teaching: An examination of the search, experience, and credence qualities of SET. *Higher Education Evaluation and Development,* 12(2), 63-84. https://doi.org/10.1108/HEED-04-2018-0009

Chism, N.V.N. (2007). *Peer Review of Teaching. A Sourcebook*. Bolton, MA: Anker.

Clayson, D.E. (2009). Student evaluations of teaching: Are they related to what students learn? A meta-analysis and review of the literature. *Journal of Marketing Education*, 31(1), 16-30. https://doi.org/10.1177/0273475308324086

Cone, C., Fox, L.M., Frankart, L.M., Kreys, E., Malcom, D.R., Mielczarek, D. & Lebovitz, L. (2022). A multicenter study of gender bias in student evaluations of teaching in pharmacy programs. *Currents in Pharmacy Teaching and Learning*, 14(9), 1085-1090. https://www.sciencedirect.com/science/article/abs/pii/S1877129722002003

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

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

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

Esarey, J. & Valdes, N. (2020). Unbiased, reliable, and valid student evaluations can still be unfair. *Assessment & Evaluation in Higher Education,* 45(8), 1106-1120. https://doi-org.ezproxy.usc.edu.au/10.1080/02602938.2020.172487

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): e0209749. https://doi.org/10.1371/journal.pone.0209749.

Foucault, M. (1977, 2020). *Discipline and Punish: The Birth of the Prison*. United Kingdom: Penguin Classics.

Galbraith, C., Merrill, G., & Kline, D. (2012). Are student evaluations of teaching effectiveness valid for measuring student outcomes in business related classes? A neural network and Bayesian analyses. *Research in Higher Education*, 53, 353–374.

Gelber, K., Brennan, K., Duriesmith, D. & Fenton. E. (2022). Gendered mundanities: Gender bias in student evaluations of teaching in political science. *Australian Journal of Political Science*, 57(2), 199-220. https://doi.org/10.1080/10361146.2022.2043241

Heffernan, T.A. (2017). A fair slice of the pie? Problematising the dispersal of government funds to Australian universities. *Journal of Higher Education Policy and Management*, 39(6), 658–673. https://doi.org/10.1080/1360080x.2017.1377965

Heffernan, T.A. (2018a). Approaches to career development and support for sessional academics in higher education. *International Journal for Academic Development*, 23(4), 312–323. https://doi.org/10.1080/1360144X.2018.1510406

Heffernan, T.A. (2018b). Using university rankings as a potential indicator of student experiences in American higher education. *Perspectives: Policy and Practice in Higher Education*, 23(1), 12–17. https://doi.org/10.1080/13603108.2018.1517108

Heffernan, T.A. & Heffernan, A. (2018c). Language games: University responses to ranking metrics. *Higher Education Quarterly*, 72(1), 29–39. https://doi.org/10.1111/hequ.12139

Heffernan, T. (2020b). Examining university leadership and the increase in workplace hostility through a Bourdieusian lens. *Higher Education Quarterly,* 75(2), 199-211. https://doi.org/10.1111/hequ.1227

Heffernan, T.A., & Bosetti, L. (2020). The emotional labour and toll of managerial academia on higher education leaders. *Journal of Educational Administration and History*, 1–16. https://doi.org/10.1080/00220620.2020.1725741

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

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

Hornstein, H.A. (2017). Student evaluations of teaching are an inadequate assessment tool for evaluating faculty performance. *Cogent Education*, 4(1), 1304016, https://doi.org/10.1080/2331186X.2017.1304016

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.

Khazan, E., Borden, J., Johnson, S. & Greenhaw, L. (2019). Examining gender bias in student evaluations of teaching for graduate teaching assistants, *NACTA Journal*, 64, 422-427.

Kreitzer, R.J. & Sweet-Cushman, J. (2021). Evaluating student evaluations of teaching: A review of measurement and equity bias in SETs and recommendations for ethical reform. *Journal of Academic Ethics,* 20, 73–84. https://doi.org/10.1007/s10805-021-09400-w

Kwok, K. & Potter, J. (2022). Gender stereotyping in student perceptions of teaching excellence: applying the shifting standards theory. *Higher Education Research & Development,* 41(7), 2201-2214. https://doi.org/10.1080/07294360.2021.2014411

LaPaglia, J., Miller, K. & Protexter, S. (2022). Gender bias interacts with instructor disfluency to negatively affect student evaluations of teaching. *Frontiers in Education*, 7, 817291. https://doi.org/10.3389/feduc.2022.817291

Lewis, K.G. (2001). Making sense of student written comments. *New Directions for Teaching and Learning*, 87 (Fall), 25-32.

Mao, Y. (2017). Intercoder reliability techniques: Holsti Method. In Allen, M (ed), *The Sage Encyclopedia of Communication Research Methods* (pp. 741-743). Thousand Oaks, CA: SAGE Publications, Inc.

Mihas, P. (2023). Qualitative multi-method and mixed methods research. In Tierney R.J., Rizvic, F. & Erikan, K. (eds), *International Encyclopedia of Education* 4th ed (pp. 302-313). Amsterdam, Netherlands: Elsevier Inc.

Miller, J. & Seldin, P. (2014). Changing Practices in faculty evaluations: Can better evaluation make a difference? *Academe*, 100(3), 35–38.

Mitchell, K.M.W. & Martin, J. (2018). Gender bias in student evaluations. *Political Science & Politics*, 51(3), 648-652. https://doi.org/10.1017/S104909651800001X

Mumby, D.K., Thomas, R. & Martí, I. (2017). Resistance redux. *Organization Studies* 38(9), 1157–83.

Padgett, S.M. (2020). 'He just teaches whatever he thinks is important': Analysis of comments in student evaluations of teaching. *Nursing Inquiry,* 28, e12411. https://doi.org/10.1111/nin.12411

Renström E.A., Gustafsson Sendén, M. & Lindqvist, A. (2021). Gender stereotypes in student evaluations of teaching. *Frontiers in Education*, 5, 571287. https://doi.org/10.3389/feduc.2020.571287

Rindermann, H., Kohler, J. & Meisenberg, G. (2007). Quality of instruction improved by evaluation and consultation of instructors. *International Journal for Academic Development*, 12(2), 73–85.

Seldin, P., Miller, J. E., & Seldin, C. A. (2010). *The Teaching Portfolio: A Practical Guide to Improved Performance and Promotion/Tenure Decisions*. John Wiley & Sons

Sewell, G. (1998). The discipline of teams: The control of team-based industrial work through electronic and peer surveillance. *Administrative Science Quarterly,* 43(2), 397–428.

Sewell, G. (2012). Organization, employees and surveillance. In Ball, K., Haggerty, K. & Lyon, D. *Routledge Handbook of Surveillance Studies* (pp. 303–312). Oxon, UK: Routledge.

Sewell, G., Barker, J.R. & Nyberg, D. (2011). Working under intensive surveillance: When does 'measuring everything that moves' become intolerable? *Human Relations,* 65(2), 189–215.

Shreffler, M.B., Shreffler, J.R. & Murfree, J.R. (2019). Barriers to advancement of female faculty in higher education: An examination of student evaluations of teaching, work-family conflict and perceived organizational support. *Journal of Higher Education Theory and Practice*, 19(8), 72-85.

Sidhu, N.S. & Fleming, S. (2023). Re-examining single-moment-in-time high-stakes examinations in specialist training: A critical narrative review. *Medical Teacher*. https://doi.org/10.1080/0142159X.2023.2260081

Sigurdardottir, M.S., Gudbjorg, L.R., Jónsdóttir, A.H. & Kristofersson, D.M. Student evaluation of teaching: Gender bias in a country at the forefront of gender equality. *Higher Education Research & Development*, 42(4), 954–967. https://doi.org/10.1080/07294360.2022.2087604

Sozer, M.E., Zeybekoglu, Z. & Kaya, M. (2019). Using mid-semester course evaluation as a feedback tool for improving learning and teaching in higher education. *Assessment & Evaluation in Higher Education*, 44(7), 1003-1016. https://doi.org/10.1080/02602938.2018.1564810

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, 598–642. http://dx.doi.org/10.3102/0034654313496870

Stark, P.B. & Freishtat, R. (2014). An evaluation of course evaluations. Science Open Research. https://doi.org/10.14293/S2199-1006.1.SOR-EDU.AOFRQA.v1

Stark-Wroblewski, K., Ahlering, R.F. & Brill, F.M. (2007). Toward a more comprehensive approach to evaluating teaching effectiveness: Supplementing student evaluations of teaching with pre–post learning measures, *Assessment & Evaluation in Higher Education*, 32(4), 403-415. https://doi.org/10.1080/02602930600898536

Steinhardt, I., Schneijderberg, C., Goˈtze1, N., Baumann, J. & Kruˈcken, G. (2017). Mapping the quality assurance of teaching and learning in higher education: The emergence of a specialty? H*igher Education*, 74, 221–237. https://doi.org/10.1007/s10734-016-0045-5e

Taylor, F.W. (1947). *Scientific Management*. New York and London: Harper and Brothers Publishers.

Taylor, F.W. (2004). *Scientific Management: The Early Sociology of Management and Organizations Volume 1*. New York: Routledge.

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-

Uttl, B., White, C.A. & Gonzalez, D.W. (2017). Meta-analysis of faculty's teaching effectiveness: Student evaluation of teaching ratings and student learning are not related. *Studies in Educational Evaluation*, 54, 22-42. https://doi.org/10.1016/j.stueduc.2016.08.007

Varga-Atkins, T., McIsaac, J. & Willis, I. (2017). Focus group meets Nominal Group Technique: An effective combination for student evaluation? *Innovations in Education and Teaching International*, 54(4), 289-300. https://doi.org/10.1080/14703297.2015.1058721

Valencia, E. (2021). Gender-biased evaluation or actual differences? Fairness in the evaluation of faculty teaching. *Higher Education 83, 1315-1333.* https://doi.org/10.1007/s10734-021-00744-1.

University of Exeter. (2023). *Teaching Quality Assurance Manual*. Available at https://as.exeter.ac.uk/academic-policy-standards/tqa-manual/lts/studentstaffliason/ (accessed 20 October 2023).

Vanderbilt University. (2023). *Student Evaluations of Teaching*. Available at https://cft.vanderbilt.edu/guides-sub-pages/student-evaluations/ (accessed 20 October 2023).

Winchester, T.M. & Winchester, M.K. (2012). Utilising the virtual learning environment to encourage faculty reflection and improve the student learning experience. In Wankel, C. & Blessinger, P. *Increasing Student Engagement and Retention using Immersive Interfaces: Virtual Worlds, Gaming and Simulation* (pp. 341-368). Bingley, England: Emerald Group Publishing.

Zhao L., Xu P., Chen Y. & Yan S (2022). A literature review of the research on students' evaluation of teaching in higher education. *Frontier Psychology*, 13, 1004487. https://org/10.3389/fpsyg.2022.1004487