In the Teaching Criteria of the University of Sheffield (TUOS) Academic Career Pathways Framework that was published in July 2018, ‘consistently excellent student feedback’ was prioritised as evidence of high quality teaching practice at all four levels:

At all four levels, other forms of evidence – feedback from external examiners/assessors, feedback from peer review of teaching, and teaching awards – were thus indicated to be supplementary to the core, priority requirement of ‘consistently excellent student feedback.’

At the same time that the Academic Career Pathways Framework is being introduced, student evaluations of teaching (SETs) are being standardised across TUOS, with all departments required to use a set of common core questions set at University level.

There is extensive academic literature on student evaluations of teaching (SETs), and on best practice for evaluating teaching in employment decisions. This paper provides a brief overview of relevant aspects of that literature, and its implications for the formulation and implementation of student evaluations and for the use of student evaluations in employment decisions such as promotion. As a university, TUOS should adopt policies that are evidence- and research-based and reflect international best practice.
Concerns Regarding Student Evaluations of Teaching

Contrary to the implications of the Academic Career Pathways Framework published in July 2018, student evaluations do not measure ‘high quality teaching practice.’ Student evaluations, concludes Richard Freishtat, the Director of the Center for Teaching and Learning at the University of California, Berkeley, are ‘primarily measures of student satisfaction with their experience in a [module].’ This is not the same thing as high quality teaching practice: ‘there is no compelling correlation between student learning and more highly rated instructors.’1 Indeed, some studies have found negative correlations between high student evaluation ratings and other measures of teaching effectiveness.2

Whether students express satisfaction with their experience of a module depends on many factors that are irrelevant to high quality teaching practice. There is extensive evidence that student evaluations of teaching are strongly (though non-uniformly) biased on the basis of personal characteristics of the instructor. These include gender and race, both of which are legally protected characteristics. Multiple studies have found that female instructors3 and instructors of colour4 tend

---

1 Richard L. Freishtat, *Expert Report on Student Evaluations of Teaching (SET)* (2016) [link], 2; see also Philip B. Stark, *Expert Report on Student Evaluations of Teaching (Faculty Course Surveys)* (2016) [link], 4–6. This briefing paper draws heavily upon Freishtat’s expert report, which was prepared for an arbitration case in Canada between Ryerson University and Ryerson Faculty Association. On the basis of the expert evidence presented by Freishtat and Stark, the *Ontario arbitrator found* in June 2018 that ‘most meaningful aspects of teaching performance and effectiveness cannot be assessed by [student evaluations],’ and directed Ryerson University to ensure that student evaluations results ‘are not used to measure teaching effectiveness for promotion or tenure.’


to be rated lower in student evaluations of teaching. These biases affect answers even to seemingly ‘objective’ questions, such as the promptness with which marks are returned.\(^5\)

Studies have found many other factors that can also affect student evaluations of teaching and that are beyond the control of the instructor and/or are irrelevant to high quality teaching practice. These include the instructor’s age, the instructor’s accent, the instructor’s perceived physical attractiveness, students’ grade expectations, the subject matter and discipline (e.g. humanities or STEM, quantitative or non-quantitative), class size, the physical condition of the classroom, etc. Instructors who address ‘sensitive, challenging, and controversial topics’ that challenge students’ beliefs, or who engage in innovation in teaching methods, also tend to be rated lower in student evaluations.\(^13\)

\(^{5}\) MacNell, Driscoll, and Hunt, ‘Exposing Gender Bias,’ 300; Boring, Ottoboni, and Stark, ‘Student Evaluations of Teaching (Mostly) Do Not Measure Teaching Effectiveness,’ 1, 8; Mitchell and Martin, ‘Gender Bias in Student Evaluations,’ 651.


\(^{11}\) James Monks and Robert Schmidt, ‘The Impact of Class Size and Number of Students on Outcomes in Higher Education’ (working paper, Cornell University, School of Industrial and Labor Relations, 2010) [link].


\(^{13}\) Freishtat, Expert Report on Student Evaluations of Teaching, 7–8; Su L. Boatright-Horowitz and Sojattra Soeung, ‘Teaching White Privilege to White Students can Mean Saying Good-Bye to Positive Student Evaluations,’ American Psychologist 64.6 (2009), 574–5.
Biases in student evaluations – conclude Anne Boring, Kellie Ottoboni, and Philip Stark – ‘can be large enough to cause more effective instructors to get lower SET than less effective instructors.’ And ‘given the many sources of bias in SET and the variability in magnitude of the bias by topic, item, student gender, and so on, as a practical matter it is impossible to adjust for biases to make SET a valid, useful measure of teaching effectiveness.’

Far from encouraging high quality teaching practice, prioritising student evaluations in employment decisions has a negative impact on teaching quality. Freishtat concludes that ‘an over-reliance on SETs as a measure of [staff] performance in teaching serves to deter pedagogical improvement and innovation.’ Instead, ‘teaching to SETs occurs. Instructors are disincentivized to improve and innovate teaching, and are instead incentivized to focus on approaches not driven by increasing student learning (e.g. lower course rigor) that are highly correlated to increased student ratings.’

Implications for the formulation and implementation of student evaluations

These concerns regarding student evaluations of teaching have important implications for how student evaluations should be formulated and implemented.

On the kinds of questions that should be used in student evaluations, Freishtat concludes: ‘Appropriate items [in student evaluation questionnaires] ask the student to comment on themselves, their background, and their experience only. Inappropriate items ask the student to comment on the course or instructor, and the impact on them/their learning. These types of questions invite the most bias, and include areas of teaching and learning that students do not have sufficient expertise to comment upon.’

Questions used in student evaluations at Sheffield should be formulated in accordance with these guidelines. Two of the four core questions that all TUOS departments are required to ask at the end of each module in 2018–19 – ‘The tutor(s) were approachable and helpful’ and ‘The teaching (was interesting and challenging and) helped me learn’ – are particularly inappropriate and invite bias.

Freishtat notes that ‘the driving purpose of standardizing SET questions across any institution is to make comparisons of teaching effectiveness in merit and promotion decisions.’ Such standardisation across an institution is ‘very problematic’:

Administering SET in a standardized way across an institution depersonalizes and ignores the complexity of teaching. It also ignores the contexts… that affect ratings (e.g., qualitative versus quantitative course, required versus elective course). Instead, it asserts that everyone must teach in the same way to be rated well on standardized items, and that all instructors have equal opportunity to garner high ratings regardless of context. It is misleading to standardize SET across an institution because of what it will necessarily further obscure (e.g., biases that affect ratings based on course...
subject, type, format, level, content, etc.), and that the obscuring penalizes and ultimately
discourages pedagogical experimentation and innovation – both things we want to encourage in
order to promote increased student learning.

For these reasons, some experts recommend that questions used in student evaluations should be
formulated ‘at the department-level, and not beyond it.’\textsuperscript{17}

**Implications for the use of student evaluations in employment decisions**

Concerns regarding student evaluations of teaching also have important implications for how
student evaluations are used in employment decisions such as promotion. As Freishtat observes,
the expert ‘consensus is that a teaching dossier is the ideal tool for assessing teaching
effectiveness.’\textsuperscript{18}

In response to concerns about overreliance on student evaluations of teaching in employment
decisions, teaching dossiers (portfolios) were developed as a method of assessing teaching
effectiveness in employment decisions by colleagues in our sister union, the Canadian Association
of University Teachers (CAUT), in the 1970s and 1980s. Bruce Shore, the professor of educational
psychology who co-ordinated the CAUT’s efforts in this field in the 1970s, explained that ‘The idea
is not to list the things an instructor must do. Rather, the aim is to create a larger selection of such
types of evidence that any one instructor could use himself [sic].’\textsuperscript{19}

Following the CAUT’s initiative, portfolio-style approaches are now used to document teaching in
higher education institutions around the world, including here at the University of Sheffield in the
Learning & Teaching Professional Recognition Scheme (LTPRS), TUOS’s internal accreditation
process for professional recognition by the Higher Education Academy (HEA). There is an
extensive academic literature on the preparation, interpretation, and evaluation of portfolios.\textsuperscript{20}

Portfolio-style approaches to documenting teaching recognise that there is no easy shortcut or
single metric that can be used to assess high quality teaching practice. Instead they bring together
multiple forms of evidence in order to provide ‘complementary perspectives on various aspects of
teaching.’\textsuperscript{21} Such approaches thus draw upon a much wider range of sources than the four
categories listed in the TUOS Academic Career Pathways Framework (student evaluations, peer
review, feedback from external examiners, teaching awards). Famously, the CAUT’s original *Guide
to the Teaching Dossier: Its Preparation and Use* (1980) – which first popularised the use of portfolio-

\textsuperscript{17} Freishtat, *Expert Report on Student Evaluations of Teaching*, 13.
style approaches – listed forty-nine categories of evidence that could be used to document good teaching. These were grouped under three main headings:

‘The products of good teaching’ (for example, student work and achievements),
‘Material from oneself’ (description of teaching duties, course syllabi, instructional innovations, and so on), and
‘Information from others’ (including students, colleagues, alumni, even employers).22

The CAUT were not suggesting that every teaching portfolio should be comprised of all forty-nine categories of evidence; indeed, the CAUT’s original guide argued that a portfolio should be no longer than three pages. (In many institutions today, teaching portfolios are longer than this: Knapper and Wright suggest a limit of around ten pages, excluding appendices.23)

There is extensive literature on the forms of evidence that can be used to document teaching.24 No list of possible components should be regarded as a one-size-fits-all template, however: Knapper and Wright note that ‘a key principle of the teaching portfolio is that the content, organization, and presentation are controlled by the individual teacher.’25

Whereas prioritising student evaluations in employment decisions has a negative impact on teaching quality, portfolio-style approaches to documenting teaching combine formative and summative functions: ‘Even when prepared largely for summative purposes [such as promotion], the very act of collecting information and interpreting it inevitably leads to self-appraisal and thoughts about possible changes.’26 The use of a broader range of types of evidence of high quality teaching practice, instead of TUOS’s plan to prioritise and require the use of student evaluation ratings, would not only lead to better promotion decisions, it would also lead to better teaching.

Simon Stevens
SUCU Committee


25 Knapper and Wright, ‘Using Portfolios to Document Good Teaching,’ 22.