

Academic Career Pathways and Student Evaluations of Teaching

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:

Teaching Criteria

High Quality Teaching Practice (Core)		
Level	Criteria	Evidence of achievement
Grade 8	Design, deliver and co-ordinate high quality teaching, assessment, feedback and student support to enable effective learning.	Having successfully completed Completion of the Postgraduate Certificate in Teaching for Learning in Higher Education (TLHE) (most commonly referred to as CiLT) you should consistently achieve excellent student feedback with a demonstrable contribution to excellent programme outcomes, and good levels of student progression and completion. This should be supplemented by other forms of evidence which should include: <ul style="list-style-type: none"> ● Excellent feedback from external examiners/assessors. ● Excellent feedback from peer review of teaching.
Grade 9	Lead and supervise delivery and assessment to ensure consistently high quality teaching, learning and student support.	Consistently excellent student feedback over a sustained period with a demonstrable contribution to excellent programme outcomes, and good levels of student progression and completion. This should be supplemented by other forms of evidence which should include: <ul style="list-style-type: none"> ● Sustained excellent feedback from external examiners/assessors. ● Sustained excellent feedback from peer review of teaching. ● Internal or external teaching awards.
Band 1 Professor	Lead and supervise delivery and assessment to ensure consistently high quality teaching, learning and student support.	Consistently excellent student feedback over a sustained period with a demonstrable contribution to excellent programme outcomes, and good levels of student progression and completion. This should be supplemented by other forms of evidence which should include: <ul style="list-style-type: none"> ● Sustained excellent feedback from external examiners/assessors. ● Sustained excellent feedback from peer review of teaching. ● Significant external teaching awards.
Band 2 Professor	Lead and supervise delivery and assessment to ensure consistently high quality teaching, learning and student support.	Consistently excellent student feedback over a sustained period with a demonstrable contribution to excellent programme outcomes, and good levels of student progression and completion. This should be supplemented by other forms of evidence which should include: <ul style="list-style-type: none"> ● Sustained excellent feedback from external examiners/assessors. ● Sustained excellent feedback from peer review of teaching. ● Prestigious external teaching awards.

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.’

See also Philip B. Stark and Richard Freishtat, ‘An Evaluation of Course Evaluations,’ *ScienceOpen Research* (2014), esp. 3–4, 1; Anne Boring, Kellie Ottoboni, and Philip B. Stark, ‘Student Evaluations of Teaching (Mostly) Do Not Measure Teaching Effectiveness,’ *ScienceOpen Research* 2016.1 (2017), esp. 1–2, 10; Bob Uttl, Carmela A. White, Daniela Wong Gonzalez, ‘Meta-analysis of Faculty’s Teaching Effectiveness: Student Evaluation of Teaching Ratings and Student Learning are Not Related,’ *Studies in Educational Evaluation* 54 (2017), esp. 38–40.

² Michela Braga, Marco Paccagnella, and Michele Pellizzari, ‘Evaluating Students’ Evaluations of Professors,’ *Economics of Education Review* 41 (2014), esp. 81–2; Scott E. Carrell and James E. West, ‘Does Professor Quality Matter? Evidence from Random Assignment of Students to Professors,’ *Journal of Political Economy*, 118.3 (2010), 412, 428–9.

³ Kristina M.W. Mitchell and Jonathan Martin, ‘Gender Bias in Student Evaluations,’ *PS: Political Science & Politics*, 51.3 (2018), 648–52; Anne Boring, ‘Gender Biases in Student Evaluations of Teaching,’ *Journal of Public Economics* 145 (2017), 27–41; Boring, Ottoboni, and Stark, ‘Student Evaluations of Teaching (Mostly) Do Not Measure Teaching Effectiveness,’ esp. 6–7, 8–9; Julianne Arbuckle and Benne Williams, ‘Students’ Perceptions of Expressiveness: Age and Gender Effects on Teacher Evaluations,’ *Sex Roles* 49.9 (2003), 507–16; Lillian MacNell, Adam Driscoll, and Andrea N. Hunt, ‘What’s in a Name: Exposing Gender Bias in Student Ratings of Teaching,’ *Innovative Higher Education*, 40.4 (2015), 291–303.

⁴ Therese A. Huston, ‘Race and Gender Bias in Higher Education: Could Faculty Course Evaluations Impede Further Progress toward Parity?,’ *Seattle Journal for Social Justice* 4.2 (2005), 598–9; Bettye P. Smith and Billy Hawkins, ‘Examining Student Evaluations of Black College Faculty: Does Race Matter?,’ *Journal of Negro Education* 80.2 (2011), 149–62; Landon D. Reid, ‘The Role of Perceived Race and Gender in the Evaluation of College Teaching on RateMyProfessors.Com,’ *Journal of Diversity in Higher Education* 3.3 (2010), 137–52; Daniel Storage, Zachary Horne, Andrei Cimpian, Leslie Sarah-Jane, ‘The Frequency of “Brilliant” and “Genius” in Teaching Evaluations Predicts the Representation of Women and African Americans across Fields,’ *PloS One* 11.3 (2016).

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,⁶ the instructor’s accent,⁷ the instructor’s perceived physical attractiveness,⁸ students’ grade expectations,⁹ the subject matter and discipline (e.g. humanities or STEM, quantitative or non-quantitative¹⁰), class size,¹¹ the physical condition of the classroom,¹² 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.

⁶ Stefano Bianchini, Francesco Lissoni, and Michele Pezzoni, ‘Instructor Characteristics and Students’ Evaluation of Teaching Effectiveness: Evidence from an Italian Engineering School,’ *European Journal of Engineering Education* 38:1 (2013), 49, 40–1; Arbuckle and Williams, ‘Students’ Perceptions of Expressiveness: Age and Gender Effects on Teacher Evaluations,’ 507–16.

⁷ John Ogier, ‘Evaluating the Effect of a Lecturer’s Language Background on a Student Rating of Teaching Form,’ *Assessment & Evaluation in Higher Education* 30.5 (2005), 477–88; Nicholas Close Subtirelu, “‘She does have an accent but...’: Race and Language Ideology in Students’ Evaluations of Mathematics Instructors on RateMyProfessors.com,” *Language in Society* 44.1 (2015), 35–62.

⁸ Daniel S. Hamermesch and Amy Parker, ‘Beauty in the Classroom: Instructor’s Pulchritude and Putative Pedagogical Productivity,’ *Economics of Education Review* 24.4 (2005), 369–76; Todd C. Riniolo, Katherine C. Johnson, Tracy R. Sherman, Julie A. Misso, ‘Hot or Not: Do Professors Perceived as Physically Attractive Receive Higher Student Evaluations?,’ *Journal of General Psychology*, 133.1 (2006), 19–35; Tobias Wolbring, and Patrick Riordan, ‘How Beauty Works: Theoretical Mechanisms and Two Empirical Applications on Students’ Evaluation of Teaching,’ *Social Science Research* 57 (2016), 253–72.

⁹ Boring, Ottoboni, and Stark, ‘Student Evaluations of Teaching (Mostly) Do Not Measure Teaching Effectiveness,’ 1, 7; Ross Vasta and Robert F. Sarmiento, ‘Liberal Grading Improves Evaluations But Not Performance,’ *Journal of Educational Psychology* 71.2 (1979), 207–11; A.C. Worthington, ‘The Impact of Student Perceptions and Characteristics on Teaching Evaluations: A Case Study in Finance Education,’ *Assessment & Evaluation in Higher Education* 27.1 (2002), 60–2.

¹⁰ Bob Uttl, Carmela A. White, Alain Morin, ‘The Numbers Tell it All: Students Don’t Like Numbers!,’ *PloS One* 8.12 (2013), 1–9.

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) [\[link\]](#).

¹² Mary Hill and Kathryn Epps, ‘The Impact of Physical Classroom Environment on Student Satisfaction and Student Evaluation of Teaching in the University Environment,’ *Academy of Educational Leadership Journal* 14.4 (2010), 65–79.

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.’¹⁴

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.’¹⁵

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.’¹⁶

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

¹⁴ Boring, Ottoboni, and Stark, ‘Student Evaluations of Teaching (Mostly) Do Not Measure Teaching Effectiveness,’ 1, 10.

¹⁵ Freishtat, *Expert Report on Student Evaluations of Teaching*, 8, 5. See also Wendy M. Williams and Stephen J. Ceci, “‘How’m I doing?’ Problems with Student Ratings of Instructors and Courses,’ *Change: The Magazine of Higher Learning* 29.5 (1997), 12–23.

¹⁶ Richard L. Freishtat, *Expert Supplemental Report on Student Evaluations of Teaching* (2016), 7 [[link](#)]; see also Stark, *Expert Report on Student Evaluations of Teaching (Faculty Course Surveys)*, 8.

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.’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.’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].’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.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.’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-

¹⁷ Freishtat, *Expert Report on Student Evaluations of Teaching*, 13.

¹⁸ Freishtat, *Expert Report on Student Evaluations of Teaching*, 9.

¹⁹ Quoted in Christopher K. Knapper, ‘The Origins of Teaching Portfolios,’ *Journal on Excellence in College Teaching* 5.1 (1995), 45–56.

²⁰ See, for instance: Carol O’Neil and Alan Wright, *Recording Teaching Accomplishment: A Dalbousie Guide to the Teaching Dossier* (5th edn., Halifax, Nova Scotia, 1995); John P. Murray, *Successful Faculty Development and Education: The Complete Teaching Portfolio* (Washington DC, 1995); Peter Seldin, J. Elizabeth Miller, and Clement A. Seldin, *The Teaching Portfolio: A Practical Guide to Improved Performance and Promotion/Tenure Decisions* (4th edn., San Francisco, 2010).

²¹ Freishtat, *Expert Report on Student Evaluations of Teaching*, 10.

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

²² Quoted in Christopher Knapper and W. Alan Wright, ‘Using Portfolios to Document Good Teaching: Premises, Purposes, Practices,’ *New Directions for Teaching and Learning* 88 (2001), 20–1.

²³ Knapper and Wright, ‘Using Portfolios to Document Good Teaching,’ 20.

²⁴ See, for instance: Freishtat, *Expert Report on Student Evaluations of Teaching*, 10–11; Stark and Freishtat, ‘An Evaluation of Course Evaluations,’ 4–6; Knapper and Wright, ‘Using Portfolios to Document Good Teaching,’ esp. 22–4; Seldin, Miller, and Seldin, *The Teaching Portfolio*, 10–20; Murray, *Successful Faculty Development and Education*, 19–36; O’Neil and Wright, *Recording Teaching Accomplishment*, 27–68.

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

²⁶ Knapper and Wright, ‘Using Portfolios to Document Good Teaching,’ 25; Seldin, Miller, and Seldin, *The Teaching Portfolio*, 61.