Student Evaluations of Teaching: Understanding Limitations and Advocating for a Gold Standard for Measuring Teaching Effectiveness

Judith Marychurch
The University of Melbourne

Kelley Burton
University of the Sunshine Coast

Michael Nancarrow
Macquarie University

Julian Laurens
The University of Melbourne

STUDENT EVALUATIONS OF TEACHING: UNDERSTANDING LIMITATIONS AND ADVOCATING FOR A GOLD STANDARD FOR MEASURING TEACHING EFFECTIVENESS

JUDITH MARYCHURCH,[*] KELLEY BURTON,[+] MICHAEL NANCARROW,[^] JULIAN LAURENS[*]

I   INTRODUCTION

Student Evaluation of Teaching instruments (SETs) are a commonly used measurement tool in the neo-liberal university that are applied with a general uniformity across all disciplines in an institution and promise an efficient, relatively straightforward measurement of teaching effectiveness as judged by students.[1] Most questions on SETs are measured on a Likert scale for ease of data collection, with a small number of open-text questions.[2] Their emphasis on providing largely quantitative data for comparative analysis of teacher-to-teacher performance has allowed SETs to be infused with an aura of objectivity, providing 'scientific' and pedagogical justification for their continued favour by the neo-liberal university. This objectivity has been challenged – research consistently finds that SETs are not a valid form of evaluation as they are currently applied and they are, in many cases, statistically unreliable.[3] Institutions have increasingly 'doubled down'

[*]   Melbourne Law School, The University of Melbourne
[+]   School of Law and Society, University of the Sunshine Coast
[^]   Macquarie Law School, Macquarie University

[1]   A presentation was given at the October 2020 Bond University Professional Legal Education Conference and the April 2021 Asian Law Schools Association Symposium by the LEAD Executive, upon which this paper is based.

[2]   A Likert scale typically allows for a five-to-seven-point scale, commonly including a range of 'strongly agree, agree, neither agree nor disagree, disagree, strongly disagree' and is used commonly in the social sciences.

[3]   Bob Uttl, Kelsey Cnudde and Carmela A. White, 'Conflict of interest explains the size of student evaluation of teaching learning correlations in multisection studies: a meta-analysis' (2019) 7 *Peer Journal*:e7225 <http://doi.org/10.7717/peerj.7225>; Bob Uttl, Armela A. White and Daniela Wong Gonzalez, 'Meta-analysis of faculty's teaching effectiveness: student evaluations of teaching ratings and student learning are not related' (2017) 54 *Studies in Educational Evaluation* 22-42, 40; Anne Boring, Kellie Ottoboni and Phillip B. Stark, 'Student evaluations of teaching (mostly) do not measure teaching effectiveness' (2016) *ScienceOpen Research,* 1.

to the extent that SETs now appear to have 'transcended' their quantitative limitations and been ascribed a qualitative dimension as a measure of teaching effectiveness, both implied and made explicit, a purpose for which they were not originally designed.[4]

In March 2020, the potentially transformative impact of the COVID-19 pandemic on teaching at many, if not most, tertiary institutions, was beginning to become apparent. Within a matter of days, universities closed campuses and academics were required to move to remote teaching using available technologies. Academics reported significantly increased working hours associated with the transition to online learning and many had to learn new technologies with minimal support, away from both academic and professional colleagues. Teaching and learning resources that normally would have taken a year or more to develop in conjunction with educational designers and with high-level technical support were developed in a matter of weeks. The resources, of course, were likely at a lower level in terms of visual and audio quality than if developed with the aid of a teaching and learning grant. However, teachers have worked extremely hard to meet the needs of their students during a time that presented an unprecedented challenge on a global scale, and which was adversely affecting staff and student wellbeing.

It was in the face of these challenges that multiple universities paused the use of SETs in response to staff concerns about their use in unprecedented circumstances. The invalidity of comparison to previous (and subsequent) SETs, was felt to have a consequential impact on confirmation and promotion prospects of teaching staff who often rely on SET data in making their argued case.

SETs are not a new topic of concern for law academics. At a meeting just prior to the COVID-19 pandemic, members of the Legal Education Associate Deans network (LEAD)[5] identified concerns with the inappropriate way that SETs were being used. These included the impact on staff wellbeing by injudicious student comments, and the

---

<https://doi.org/10.14293/S2199-1006.1.SOR-EDU.AETBZC.v1>; *Ryerson University v Ryerson Faculty Association* (Ontario Labour Arbitration Awards) William Kaplan Arbitrator, 10 April 2018 CanLII 58446 (ON LA), 5 ('*Ryerson*').

4 Regarding the 'original purpose [of SETs] as tools to inform pedagogical practice, and their current mis(use) as performance markers' see Warwick Fisher et al, 'Student Evaluations: Pedagogical Tools or Weapons of Choice?' (2020) 29 *Legal Education Review* 7, 7. These authors note that SETs measure student *perceptions* of teaching effectiveness, as opposed to teaching quality or student learning, also at 7 (emphasis added).

5 The Legal Education Associate Deans network (LEAD) was established in 2010. It is comprised of the Associate Deans (Education/Teaching and Learning) (or equivalent) of Australian law schools. The LEAD Executive consists of Co-convenors who are Associate Deans Teaching and Learning (or equivalent) and a Project/Executive officer. The purpose of LEAD is to promote collaborative approaches to teaching and learning in Australian legal education. The LEAD network intends to lead, encourage, and document good practice in learning and teaching in the discipline of law and to create processes and strategies to sustain the network as a Sub-Committee of the Council of Australian Law Deans (CALD). It previously received funding from the Office for Learning and Teaching (OLT), and now receives assistance from CALD pursuant to a Memorandum of Understanding. See <http://lawteachnetwork.org/>.

seemingly disproportionate weight given to SET scores by managers when reviewing performance. LEAD members agreed broadly at that time that this was as an area requiring future coordinated change. LEAD was therefore a natural point of contact for Associate Deans Learning and Teaching (or equivalent) when colleagues began raising concerns about the risks of the continued use of SETs in the unprecedented circumstances of COVID-19 where teachers were re-designing delivery of subjects at short notice.

LEAD sought to assist members by preparing a statement [6] supporting a pause in the use of SETs that they could then use in advocacy at their own institutions. Many institutions did pause the use of SETs during the pandemic, which allowed some teachers, academic divisions and institutions to develop alternative measures of teaching effectiveness, and, in some cases, to revise the SET instrument.

The purpose of this paper is to identify the gold standard for, or best practice in, measuring teaching effectiveness and promote LEAD's initiatives in this space. In this paper, Part II identifies the gold standard for measuring teaching effectiveness as articulated in the Canadian arbitration case *Ryerson*[7] in which an academic challenged the use of SET for career advancement. An equivalent Australian case has not emerged. Part III analyses the strengths and weaknesses of SETs, noting that they were particularly problematic during the COVID-19 pandemic and should not be the primary data source of teaching effectiveness. Part IV shares the LEAD policy recommendations on best practice in SETs for Australian law schools during the COVID-19 pandemic. Finally, Part V reflects on other qualitative and quantitative measures of teaching effectiveness, especially a teaching portfolio and Peer Review of Teaching (ProT), which were endorsed in *Ryerson*.[8]

II   A 'GOLD STANDARD' FOR MEASURING TEACHING EFFECTIVENESS

The LEAD discussion on the use of SETs as measures of teaching effectiveness at its September 2019 meeting centred around the recent Canadian arbitration decision *Ryerson*[9] which examined the strengths and weaknesses of SETs and advanced a 'gold standard' for measuring teaching effectiveness. The 'gold standard' is a reference to what should be considered as the best practice approach, in this case, to measuring teaching quality and effectiveness.[10] A gold standard for measuring teaching effectiveness in tertiary education has not been explicitly articulated in Australia. As a result, the Canadian decision in the context of collective bargaining between a University and that University's Association for staff members, provides a useful benchmark. In *Ryerson,* Arbitrator Kaplan rejected the use of SETs as a measure of

---

[6]   This is provided in Appendix 1.

[7]   *Ryerson* (n 3).

[8]   Ibid 6.

[9]   *Ryerson* (n 3)

[10]   Ibid 8.

teaching effectiveness for the purposes of confirmation and promotion of academic staff.[11] Kaplan determined that '[t]he expert evidence convincingly establishes that the best way to assess teaching effectiveness is through the careful assessment of the teaching dossier and in-class peer evaluations.'[12] This was advanced as the 'gold standard for measuring teaching effectiveness.'[13]

Kaplan considered it 'axiomatic' that the evidence of high quality in teaching practice is essential for tenure and promotion and that decisions on confirmation and promotion should be based on the best evidence.[14] The value of SETs in providing information directly from students on their perceptions of their educational experience was noted, as was the importance of student satisfaction to the mission of the University and the need for University management and teachers to know this.[15] However, the interpretation of SET responses 'is challenging – they tell a story and add depth but the information needs to be carefully contextualised.'[16] Kaplan then turns to examine the strengths and weaknesses of SETs having regard to the evidence presented. Despite their air of objectivity and the appearance of being scientifically sound in that they suggest an objective basis for correlating numerical data, Kaplan noted 'serious and inherent limitations' arising from 'numerous factors, especially personal characteristics … such as race, gender, accent, age and "attractiveness" skew SET results'.[17] Questions were raised about the reliability of SETs completed online compared with those completed in class, a factor particularly relevant to teachers due to seek confirmation, or considering applying for promotion in 2020 or subsequent years, as well as the low response rates that commonly render the results less reliable at best, if not unreliable, given that the views of responding students cannot necessarily be applied to non-responding students.[18]

Ryerson is significant in that it lends critical momentum to moves challenging the use of SETs in the neo-liberal university, particularly in the context of using them for academic promotion. *Ryerson* gives weight to a considerable body of research that calls into question core assumptions about the ability of SETs to measure teaching effectiveness in a manner that is methodologically valid.[19] In making his award, Kaplan identified that:

> The expert evidence led at the hearing persuasively demonstrates that the most meaningful aspects of teaching performance and effectiveness cannot be assessed by SETs. Insofar as assessing teaching effectiveness is

---

11 Ibid.
12 Ibid 8.
13 Ibid.
14 Ibid 4.
15 Ibid.
16 Ibid 5.
17 Ibid 6.
18 Ibid.
19 See, eg, Uttl, Cnudde and White (n 3).

concerned – especially in the context of tenure and promotion – SETs are imperfect at best and downright biased and unreliable at worst.[20]

This is where the concept of a gold standard or best practice in assessing teaching effectiveness that draws on other evidence aside from SETs is essential. Kaplan considered it 'far from clear whether students have the expertise to comment on course content and teaching methods and assignments,'[21] raising peer review of teaching practice as the logical alternative. Peer evaluation also overcomes the risks associated with teachers modifying their behaviour to achieve desired outcomes on SETs.[22] This risk was particularly apparent in the rapid transition to online learning in many institutions nationally and internationally as the COVID-19 pandemic took hold in 2020 and extended periods of lockdown began. Such transition, often with little to no technological or educational design support, meant that academics had to innovate and take risks in their teaching practice in a short period of rapidly changing circumstances which ordinarily would have required many months of resource intensive development, often via teaching development grants. The COVID-19 pandemic created a unique opportunity for development of a teaching portfolio comprising a variety of evidence, including peer review, to become best practice in measuring teaching effectiveness as many institutions paused the use of SETs.

This paper uses the term 'teaching effectiveness' instead of 'teaching quality' for consistency but it is acknowledged that both terms have been utilised interchangeably in Australian higher education and are intended to capture the quality of individual teaching.[23] Because SETs have been found to be both an unreliable quantitative indicator of teaching effectiveness, [24] and not a valid qualitative evaluation of teaching effectiveness and student learning,[25] the authors argue that SETs should not be the sole or main method of measuring teaching effectiveness in a neoliberal university. While SETs do retain some usefulness, albeit limited, as a measure of student *perceptions* of teaching quality and of their experience in a subject, the assumption that teaching *effectiveness* and *student learning* can be measured in a valid way by SETs is problematic and can be mistaken for an evaluation approach 'that does no more than reinforce a neoliberal notion of students as customers paying for a service, turning faculty teaching into a form of entertainment that plays to … "the applause meter."'[26]

---

[20] *Ryerson* (n 3) 5.
[21] Ibid 6.
[22] Ibid 7.
[23] Denise Chalmers, *Indicators of University Teaching and Learning Quality*, (Australian Learning & Teaching Council, April 2008). 'Effectiveness' is often used in conjunction with 'efficiency' although there are clear differences between the terms. Effectiveness connotes a measure of the degree to which something succeeds in producing the desired result, whereas efficiency measures input versus output and is an inherent in the neoliberal focus on economic efficiency.
[24] Warwick Fisher et al, 'Student Evaluations: Pedagogical Tools or Weapons of Choice?' (2020) 30 *Legal Education Review* 1, 9.
[25] Ibid 7.
[26] Ibid.

In the context of problematic SETs, the agenda of LEAD was to raise Australian law school awareness of the *Ryerson* determination regarding the gold standard for measuring teaching effectiveness. More specifically, the *Ryerson* determination was discussed at the September 2019 meeting of LEAD, where members recognised the importance of the decision given the high-stakes nature of academic career advancement. LEAD made a commitment to explore the implications of the Canadian decision in 2020. The authors' initial intent was to raise awareness of *Ryerson* amongst Australian and other law academics from the wider Asia-Pacific and Oceania region and advocate for a teaching portfolio and PRoT for best practice in measuring teaching effectiveness.[27]

Ideally, a teaching portfolio would embrace a variety of evaluative mechanisms that were 'fit for purpose' for the teaching and learning tools being used by an individual academic and/or teaching team. Peer review would be an important part of a teaching portfolio and may consist of a review of classroom teaching (as the traditional notion of peer review recognised in *Ryerson*), but also peer review of, for example, curriculum design; online learning design; learning resources; or assessment practices (including feedback tools). Peer review would ideally be engaged in as a regular process of professional development and tailored to the academic teacher's needs at the time, whether for ongoing reflective practice or specifically for purposes of confirmation, promotion, or as evidence for a teaching award or teaching grant application. Finally, peer review may be internal or external to the teacher's own discipline or institution, each potentially being valid depending on what is being reviewed and the purpose of the review. The unprecedented events of 2020 and 2021 have expanded and galvanised this mission to seek change in measures of teaching effectiveness that meet the present and emerging challenges in the post-COVID-19 world.

The impact of the COVID-19 pandemic on teaching and learning at tertiary institutions around the globe has been profound. In many cases, traditional face-to-face classes were moved to online learning within a week, and asynchronous, synchronous, and blends of asynchronous and synchronous learning tools replaced traditional face-to-face delivery. Some institutions in Australia abandoned SETs in 2020 to allow teachers to innovate without fear of repercussions in SET scores and the consequent impact on professional development appraisals ('please explain'), confirmation and promotion processes. These institutions are now considering whether or not to re-introduce SETs in 2021, given the ongoing uncertainty in the teaching and learning landscape due to the restrictions on student travel to their institutions of choice, which often dictates dual delivery of programs on-campus and online. An immediate return to traditional SETs is not necessarily being

---

[27] As noted previously (n 1), a presentation was given at the October 2020 Bond University Professional Legal Education Conference and the April 2021 Asian Law Schools Association Symposium by the LEAD Executive, upon which this paper is based.

contemplated at all institutions. Even when the impact of the COVID-19 pandemic passes, it is likely that the effects on teaching and learning will be ongoing and some academics are embracing this period of change as an opportunity to create and experiment.[28] This means the pandemic has given rise to a unique opportunity to re-think not just teaching and learning, but also how we assess effectiveness in teaching and learning. Taken in conjunction with the decision in *Ryerson*, the present circumstances create an opportunity to revisit how to measure teaching effectiveness, particularly the reliance on SETs primarily in tertiary institutions. This is an argument which the authors believe law academics are well placed to lead at their own institutions. This paper seeks to arm our colleagues with the arguments and evidence to support this advocacy.

III   SETs - The Ultimate Irony: More Weaknesses than Strengths

SETs are often prioritised as evidence of high-quality teaching practice, whereas feedback from external assessors, examiners and peer review is often placed in the *supplementary* category for the purpose of academic promotion.[29] However, in providing expert evidence in relation to the *Ryerson* arbitration, Stark concluded that 'the use of the results [of SETs] in employment decisions should be discouraged, if not forbidden'.[30] Academics are vulnerable to the vagaries of SETs – and this vulnerability is heightened in periods of unprecedented change in the nature and volume of work, such as those precipitated by the COVID-19 pandemic. This part will identify some key strengths and weaknesses of SETs and concerns that were exacerbated during the COVID-19 pandemic.

In tertiary educator evaluation, what is valued by management in reviewing and using SET metrics is 'consistently excellent student feedback'. The main advantage of SETs for managers is the metrics that SETs generate can easily translate into management-driven key performance indicators (KPIs) for teaching staff and thus seamlessly integrate into academic performance reviews. For example, the metrics can be translated into various presentation formats (for eg, graphs, charts, tables and colour coding) to quickly allow comparison across whole of university course and teaching offerings. However, as an evaluation yardstick for teacher performance, '[g]enerally, [SETs] lack appropriate controls, do not use randomisation, use inappropriate

---

[28]   Blair Major, 'Making Something New: Legal Education in a Pandemic' (2020) 25(4) *Lex Electronica* 93-98, 95.

[29]   Simon Stevens, *Academic Career Pathways and Student Evaluations of Teaching* (SUCU Committee, University of Sheffield, 2018) 1. <http://ucu.group.shef.ac.uk/wp-content/uploads/Stevens-Academic-Career-Pathways-and-Student-Evaluations-of-Teaching.pdf>

[30]   Ibid.

statistical tests, and conflate statistical significance with effect size.'[31] And so, comparisons of SETs across a university may prove meaningless given that the underlying data and metrics are not a reliable measure of learning effectiveness, innovation and creativity. Further, studies have confirmed that SETs contain inherent bias (for example, bias based on gender, ethnicity, race and sexual orientation) using generic questions that fail to incorporate differentiation and individuality in educator-student interactions.[32]

It has been argued that SETs have the benefit of being designed using objective criteria and thus are not affected by potentially value-laden subjectivity.[33] However, this argument is flawed on two grounds. First, as we have argued, SETs carry in-built bias and are thus not genuinely objective. Second, the search for a purely objective evaluation strategy is misconceived. While the method of determining effectiveness in teaching and learning should be grounded in criteria, it should accommodate diverse pedagogic styles. SETs may also discourage teachers approaching confirmation and promotion from pursuing innovation, a critical feature of effective teaching, because innovation often creates an initial uncertainty for teachers and students, as learning occurs from trial and error, and teaching innovations are refined over time.

In effect, SETs have become a measure of 'customer satisfaction' of students. Students rate their 'customer satisfaction' of an individual teacher rather than the actual educational learning that transpires in teacher-student interactions. This is particularly evident in online tools like Rate My Professor,[34] which allows students to rate their professors online, anytime. Students self-select to participate in such anonymous online evaluations and tend to offer extreme, unrepresentative and unmoderated views without any fear of reprisals.[35] Tertiary education has become commodified.[36] In this environment, learning is equated with having a pleasurable experience, where engagement between an educator and their students is evaluated based on how 'appealing' a teacher is and how 'pleasant' the student's experience was.[37] It is

---

[31] Phillip B, Stark, 'Expert report on Student Evaluations of Teaching', Submission in *Ryerson Faculty Association v Ontario Confederation of University Faculty Association*, 10 October 2016.

[32] Kerry Chavez and Kristina MW Mitchell, 'Exploring Bias in Student Evaluations: Gender, Race, and Ethnicity' (2019) 53(2) *PS: Political Science & Politics* 270.

[33] Cf Lorraine Daston and Peter Galison, *Objectivity* (Zone Books, 2010), 363-82 on the evolution of scientific objectivity.

[34] See <https://www.ratemyprofessors.com> (accessed 9 March 2021). See also Pamela Leong, *Rating Professors Online: How Culture, Technology and Consumer Expectations Shape Modern Student Evaluations* (Palgrave MacMillan, 2020).

[35] Leong (n 34) 39-40.

[36] Cf Raewyn Connell, *The Good University* (Monash University Publishing, 2019) 119, who argues that public universities have become corporatized from within. See also Jurgen Habermas, Thomas McCarthy (translator), *The Theory of Communicative Action, Vol 2: Lifeworld & System: A Critique of Functional Reason* (Beacon Press, 1981) regarding societal commodification at the meta level.

[37] See, eg, Leong, (n 34) 143 -168, where the author identifies key characteristics of the 'ideal' teacher, emerging from analysis of reviews on the Rate My Professor online tool, such as (not exhaustive) accessibility/ availability; how accommodating the

acknowledged that evaluating the validity of SETs is a complex task,[38] and that well-designed SETs may provide useful information about student perceptions of teaching and of subjects where they are appropriately applied and the data is properly analysed, having regard to their inherent limitations.

In addition to the inherent bias that has historically plagued the configurations of SETs, the way in which these biases are played out in remote and online learning needs to be addressed. Educators that are challenging and innovative can find themselves more open to criticism through the standardised matrix of SET questions as currently designed. One reason for this is that these teachers tend to make their students' lives and learning journeys more intellectually challenging and difficult. SET questions are frequently slanted towards how easy it is for students to achieve certain outcomes, such as to navigate websites, or how well assessments were explained in advance and that assessments were returned in a timely fashion.

The configuration of SETs generally fails to account for nuances in learning relationships and spontaneous teaching and learning innovations. During the COVID-19 pandemic world, uploading existing learning materials on a learning management system missed the opportunity to include students as partners and co-create organic learning materials and strategies. The same could be said about the post-COVID-19 world if learning materials need to be developed and uploaded on a learning management system in advance of the current cohort commencing their studies. Authentic educators are responsive not only to developments in content but also cognisant of students as partners in the co-creation of learning strategies and materials over time.[39]

A significant failure of the current SET strategy is that it does not acknowledge the deep differences between conventional classroom-based learning and virtual learning.[40] It is now accepted that the COVID-19 pandemic (a spectre that continues to haunt our daily lives in manifest ways) moved higher education into 'remote' spaces mostly inhabited by attendees in Zoom sessions.[41] What were initial hurdles to conducting classes and student communications and engagements to online virtual encounters have emerged as a normalised environment of higher education. In the post-COVID-19 pandemic world of remote

---

teacher is; approachability; how caring the teacher is; how calm and pleasant the teacher's voice or intonation is.

[38] Peter Spooren, Bert Brockx and Dimitri Mortelmans, 'On the Validity of Student Evaluation of Teaching: The State of the Art' (2013) 83(4) *Review of Educational Research* 598, 29, 32.

[39] See generally, K Matthews et al, 'Toward theories of partnership praxis: an analysis of interpretive framing in literature on students as partners in teaching and learning' (2019) 38(2) *Higher Education Research & Development* 280.

[40] Encompassing fully online and remote learning.

[41] In the early shift to Zoom-based learning, what became known as 'zoom bombing' became a not uncommon development. At some institutions, despite its clearly unwelcome and at time unacceptable presence in a learning session, some university administrators and senior managers welcomed it as a marketing and 'engagement' opportunity. Law academics were advised to use a password on Zoom meetings to avoid 'zoom bombers.'

learning, there is little correlation between universities promotion of more individualised and personalised learning experiences for students and the standardised one-size-fits-all approach adopted by SETs for educator evaluation of teaching and learning achievements.[42]

Another effect of COVID-19 on SETs has been the financial and resourcing impact upon all universities. Law schools have lost significant numbers of qualified, experienced and committed educators. In some cases, these full-time staff have been replaced by sessional academic staff. In many others, remaining academic colleagues have been required to take on heavier teaching loads, often at short notice. These consequences of COVID-19 in higher education inevitably impacts upon the student learning experience in various way: loss of face-to-face instruction; loss of the opportunity to be taught by experienced full-time academic-focused educators; a sense of isolation through all learning and contact moving to the online/remote context; and implications for the wellbeing of all those involved in higher education, both students and educators. In the midst of these disruptions, different approaches to using SETs emerged.

At some institutions, SETs were suspended, driving academics to identify alternative measures of teaching effectiveness. Other institutions replaced SETs with an alternative survey instrument in a hurried manner, raising concerns about whether the alternative survey instrument elicited evidence needed by academics. Many institutions, however, continued with SETs (either modified or unmodified) during the COVID-19 pandemic, despite the fact that the teaching and learning experience over that time had changed.

IV   LEAD ACTION ON SETs DURING THE COVID-19 PANDEMIC

At the September 2019 meeting of LEAD, multiple members raised concerns about the use of SETs as an evaluative tool for staff performance and the impact of SET results and, in particular, students' comments in open-text sections of SET instruments. The pandemic accelerated work on this issue, as LEAD members reported concerns to the LEAD Co-convenors about the use of any SETs administered during 2020. The issues related to the use of the normal evaluations in a context which was not the conventional teaching and learning environment an academic would normally work within. With the huge acceleration and uplift in teachers' use of technology, it was felt that teachers needed to be encouraged to experiment with mechanisms to keep students engaged in their learning in a remote context, and to innovate without being concerned about the impact on SET scores. In particular, the effect on academics needing to apply for confirmation or considering a promotion application in the near future would be impacted by SET scores from 2020, though the effect could last into the future with SET scores mapped over time for any individual teacher.

---

[42]   Certainly, in our COVID-19 informed learning context, well-being should figure more prominently in SET formulations.

Comparison of past (and potentially subsequent) SET results to those of 2020 would likely be inappropriate due to the unusual and pervasive context of teaching and learning in 2020, rendering them even more problematic than already identified. The risk of injudicious use of, at the time, 2020 SET data had to be balanced against allowing the opportunity for institutions and teachers to efficiently gather evidence of the impact of new teaching and learning tools for future development given the uplift in technology usage, the acceleration of change and the anticipated permanent impact of COVID-19 on pedagogy.

LEAD members appreciated that institutions would need to ascertain how students were coping with the forced change to remote and online learning, and that academics may want to evaluate the impact of new teaching and learning resources provided for subjects. There was no desire to interfere with this sort of evaluation; indeed, LEAD wanted to encourage such evaluation in the interests of student learning, teacher professional development, and to ensure academics could provide evidence of effectiveness in teaching during the COVID-19 pandemic. As a result, LEAD prepared a best practice statement on the use of SETs during the COVID-19 pandemic, with the intention of providing Associate Deans – Teaching & Learning (or equivalent) and Deans of law schools a tool they could use for advocacy within their own institutions for appropriate use of SETs during the COVID-19 pandemic. The statement issued by LEAD is provided in Appendix 1. This was distributed via the LEAD network and also the Council of Australian Law Deans (CALD).

Following release of this statement outlined in Appendix 1, LEAD members anecdotally reported that the statement was useful to refer their Deans and other senior members of their institutions to, and had aided them in engaging others in conversations about appropriate use of SETs during the COVID-19 pandemic. [43] While we have no confirmation that the statement actually influenced change, the articulation of a best practice statement was appreciated by LEAD members. For this reason alone, the work in developing the statement was worthwhile. LEAD members in their respective law schools play an essential role in motivating and inspiring teachers every year and particularly in a year with significant upheaval in professional as well as personal lives. Resourcing those who support the teachers, who in turn support our students, is essential work for LEAD as a collegiate and supportive national network of Associate Deans - Learning & Teaching (or equivalent).

The next stage of this work, which is particularly important given the continuing effect of the COVID-19 pandemic on modes of teaching and pedagogy, is to empower our network of Associate Deans – Learning & Teaching in law to be advocates at their own institutions for best practice in measuring achievement of excellence in teaching

---

[43]   However, while many Law Schools were responsive to these suggestions, where law schools were located in wider organisational units and with less autonomy, these suggestions were less well acknowledged. It became evident that some institutions intended to make no change to the operation of SETs despite the widespread disruptive impact of the COVID-19 pandemic.

and learning, which focuses on a teaching portfolio, including PRoT. This should not be seen as advocating for the removal of the student voice in measuring teaching effectiveness, but for seeing SETs as only a part of a portfolio of evidence of teaching effectiveness rather than the primary quantitative and qualitative data source. The 'pause' or partial change in traditional SETs during the 2020 and 2021 COVID-19 pandemic period provides us with the opportunity to re-conceive both the form and use of SETs as one measure of teaching effectiveness, amongst others. Indeed, there will be a cohort of new academics whose typically three-year probationary periods will be impacted significantly by the COVID-19 context of SETs, making statistical comparison with previous or subsequent SET results invalid. This cohort of teachers need to be supported through PRoT and development of a teaching portfolio reflective of their lived experience of the COVID-19 pandemic as a tertiary teacher.

V   A TEACHING PORTFOLIO: OTHER MEASURES OF TEACHING EFFECTIVENESS

The weaknesses of SETs and the rapid transition to remote learning as a result of the COVID-19 pandemic, has provided an opportunity to explore other measures of teaching effectiveness that are evidence-based and grounded in the literature. For example, in a higher education context, five Western Australian universities led an Office for Learning & Teaching Project that developed a framework to provide examples of evidence for teaching criteria for each by promotional level on the academic ladder.[44] Some examples of teaching effectiveness evidence outlined at the Professor E level include 360-degree leadership of teaching feedback, feedback from mentees, contribution to a committee or leadership role, teaching fellowship application, professional development transcripts, teaching and learning grants, teaching and

---

[44]   'Indicative Standards by Promotional Level: Professor (E)', Australian University Teaching & Criteria & Standards' (Web Page, 18 November 2021) <http://uniteachingcriteria.edu.au/framework/career-promotions/indicative-standards-promotional-level/new-professor-e/>. The framework underwent several iterations resulting from trials and feedback. The teaching criteria included 1. Design and planning of learning activities; 2. Teaching and supporting student learning; 3. Assessment and giving feedback to students on their learning; 4. Developing effective learning environments, student support and guidance; 5. Integration of scholarship, research and professional activities with teaching and in support of student learning; 6. Evaluation of practice and continuing professional development; and 7. Professional and personal effectiveness. The seven criteria were grounded in an extensive literature review, including the United Kingdom Higher Education Authority (HEA) Professional Standards Framework. The contact for this Office for Learning & Teaching Project is Emeritus Professor Denise Chalmers. The five Western Australian universities are Curtin University, Edith Cowan University, Murdoch University, The University of Notre Dame Australia and The University of Western Australia. For an example of another framework, see U21 Educational Innovation Steering Group, *U21 Teaching Standards Framework* (Report, February 2018) <https://universitas21.com/sites/default/files/2018-03/U21-TSF-Full-Description-MASTER.pdf>.

learning publications, reflective teaching portfolio[45] and an invitation to present at a conference as a keynote.[46] These examples are not unique to the Professor E level and have also been charted as examples of evidence appropriate for other promotional levels on the academic ladder. The examples of evidence are relevant to all teaching and learning endeavours, including confirmation; promotion; learning and teaching grant applications; teaching awards at an institutional, discipline and national level; and fellowships of organisations such as the Higher Education Research and Development Society of Australasia (HERDSA) and Advance Higher Education.[47]

The collection of evidence for all teaching and learning endeavours should be cumulative and demonstrate sustained effectiveness of teaching over the relevant timeframe. For example, when preparing an Advance Higher Education Fellowship, the teaching effectiveness evidence should be connected with the reflective narrative to show how the prospective fellow meets the United Kingdom Professional Standards Framework.[48] A further example, is that a nomination for an Australian Award for University Teaching (AAUT) such as a Citation should link the nominee's teaching philosophy with evidence of scholarly practice because that is required by the AAUT Assessment Matrix.[49] Measures of teaching effectiveness should consist of both quantitative and qualitative evidence. Such evidence may draw on data and student comments from SET instruments, but this would only be part of a comprehensive teaching portfolio and would need to be contextualised with additional information, such as response rates and

---

[45] The ability to engage in reflective practice has received greater attention in law: see, eg, Anna Huggins, Sally Kift and Rachael Field, 'Implementing the Self-Management Threshold Learning Outcome for Law: Some International Design Strategies from the Current Curriculum Toolbox' (2011) 21 *Legal Education Review* 183, 207; Sally Kift, Mark Israel and Rachael Field, *Learning and Teaching Academic Standards Project: Bachelor of Laws Learning and Teaching Academic Standards Statement December 2010*, (Council of Australian Law Deans, January 2011) 10 <https://cald.asn.au/wp-content/uploads/2017/11/KiftetalLTASStandardsStatement2010.pdf>. A teaching portfolio could be a subset of a professional portfolio that may consist of other academic pursuits such as research and engagement.

[46] Ibid.

[47] *Higher Education Research and Development Society of Australasia* (Website) <https://www.herdsa.org.au/>; *Advance Higher Education* (Website) <https://www.advance-he.ac.uk/>. Some Australian higher education institutions, for example, the University of the Sunshine Coast, support law academics to apply for an Advance Higher Education/Higher Education Academy fellowship. Law academics lacking access to such professional development opportunities are encouraged to advocate within their institution for the time and space to participate in such programs, utilising the outcomes of the professional development as justification.

[48] Advance HE, 'UK Professional Standards Framework (PSF)' (Web Page, 18 November 2021) <https://www.advance-he.ac.uk/guidance/teaching-and-learning/ukpsf>.

[49] Australian Government, Department of Education, Skills and Employment, *2021 AAUT Program Overview* (Report, 2021) <https://www.universitiesaustralia.edu.au/wp-content/uploads/2021/05/2021-AAUT-Program-Overview-Final-30Apr.pdf>.

details of changes in a subject that may explain differences in SET results over time.

Anecdotally, in the authors' experience as applicants and assessors in teaching and learning processes, applications tend to include more qualitative evidence than quantitative evidence in teaching and learning endeavours. Qualitative evidence explains, in words, changes, improvements and impact of, for example, a peer review of a teaching approach or the development of learning materials. Similarly, quantitative evidence provides numerical data and may substantiate claims about improvement, growth, or usefulness over time. By way of illustration, statistics tracking may indicate the number of times a resource is downloaded by students on a Learning Management System (LMS). Where the quantitative evidence consists of surveys of relevant stakeholders, the number of respondents gives a greater insight into the significance of the results. [50] When interpreting the results, law academics should appreciate any limitations such as potential biases, subjectivity of self-reporting respondents and lack of self-awareness by the volunteer respondents. To help mitigate the shortcomings of quantitative evidence, a wide range of measures of teaching effectiveness should be utilised, including ProT.

ProT has been promoted as an alternative teaching effectiveness tool to SETs for over a decade. [51] Peer evaluations in classes were underscored as a salient characteristic of the gold standard of teaching effectiveness in *Ryerson*.[52] Presumably, the peer evaluation in classes could have taken place in the new learning environment, for example, Zoom, when some law classes rapidly moved from face-to-face to remote learning during the COVID-19 pandemic. However, the authors advocate that peer review should not be limited to what happens in classes and could canvass the effectiveness of learning materials, innovation, authentic experience, curriculum, assessment, leadership roles and mentoring. [53] ProT may consist of solicited, including via structured peer review processes, or unsolicited qualitative comments from academics within a law school or institution, including academics

---

[50] Vicci YJ Lau, 'How to Encourage Student Voice: Obtaining Effective Feedback from Law Students in Course Evaluation' (2019) 29(1) *Legal Education Review* 1, 2.

[51] Philip Morgan, 'The Course Improvement Flowchart: A Description of a Tool and Process for the Evaluation of University Teaching' (2008) 5(2) *Journal of University Teaching and Learning Practice* 1, 7-8. The literature identifies various models of PRoT including formative, diagnostic, summative, evaluative and developmental: Alexandra L Johnston, Chi Baik and Andrea Chester, 'Peer Review of Teaching in Australian Higher Education: A Systematic Review' (2022) 41(2) *Higher Education Research and Development* 390, 392. The existing literature discusses the advantages and disadvantages of participating in PRoT programs: Peter Grainger et al, 'Peer Review of Teaching (PRoT) in Higher Education – A Practitioner's Reflection' (2016) 17(5) *Reflective Practice* 523. For an example of a PRoT program used in the disciplines of law and criminology see Dominique Mortiz et al, 'Designing a Peer-led Approach to Teaching Review and Enhancement in Academia' (2021) 18(1) *Journal of University Teaching and Learning Practice* Article 7.

[52] *Ryerson* (n 3)

[53] 'Indicative Standards by Promotional Level: Professor (E), Australian University Teaching & Criteria & Standards' (Webpage) <http://uniteachingcriteria.edu.au/framework/career-promotions/indicative-standards-promotional-level/new-professor-e/>.

with experience in the same field of law, early career academics, sessional staff who commonly also work in practice and academics in other disciplines. Peers could be external, recognised experts in teaching and learning, and may be from the academic's own institution, from another institution nationally or internationally.[54] Within the legal academy, other peers who may offer qualitative comments about teaching effectiveness include CALD, Discipline Scholars in Law, LEAD, and the Law Admissions Consultative Committee and equivalent bodies in other jurisdictions.

Additional examples of qualitative peer recognition include the adoption of another academic's teaching practices or resources and sharing exemplars with appropriate acknowledgements on websites. Contributing to the scholarship of teaching through peer-reviewed journal articles can generate quantitative evidence from peers such as the number of views, downloads, citations, Google Scholar h-index and i10-index, as well as qualitative evidence such as comments from peer reviewers and an editorial board of a journal. Recognition by peers is valued in teaching and learning application endeavours, but a contemporary approach to teaching effectiveness encourages applicants to transcend the boundaries of academia.

A teaching portfolio could provide evidence of teaching effectiveness from a wide range of peers in the broader community. For example, industry partners could provide feedback on whether the learning outcomes are aligned with industry needs, whether industry experience is being integrated into learning materials and whether the program is preparing graduates for professional practice.[55] In a legal education context, qualitative comments may be derived from the Australian Law Reform Commission, members of the judiciary, legal practitioners from community legal services or law firms, barristers, in-house counsel for government agencies and publishers.

In addition to peer review of monograph or other publications designed for students in the form of, for example, book reviews and other qualitative comments, publishers should have the capacity to provide quantitative data to inform assessment of an academic's impact on teaching through uptake of published works designed for students[56] through on the number of new book sales; which institutions prescribe a book; and where the book has been sold internationally. Further quantitative data on student borrowings from general or high use collections can be gathered from institutional libraries. Such data would, for instance, assist in a case being made against the AAUT Citation criterion 'development of curricula, resources or services that reflect a command of the field'.[57]

---

[54]   Ibid.

[55]   Ibid.

[56]   These may include traditional textbooks, 'how to' guides, work published on open sources, such as Creative Commons.

[57]   Australian Awards for University Teaching, 'Citations for Outstanding Contributions to Student Learning: 2022 Citation Nomination Instructions', *Universities Australia* (Web Page) <https://www.universitiesaustralia.edu.au/policy-submissions/teaching-learning-funding/australian-awards-for-university-teaching/>.

The decision in *Ryerson* emphasised that the gold standard for measuring teaching effectiveness was predicated on the vigilant assessment of a teaching portfolio (dossier).[58] The authors agree that a teaching portfolio that incorporates a variety of evidence, including peer review of aspects of their teaching practice, is best practice in evaluating teaching and urge law academics to utilise a wide range of qualitative and quantitative evidence from multiple stakeholders, inspired by the preceding examples, when preparing applications in relation to their teaching and learning endeavours.

VI CONCLUSION

University teachers cannot be replaced by machines, or generative artificial intelligence.[59] Technology is an aid and not the master in a student's learning journey. University educators are scholars and should be reflective pedagogical practitioners. A university learning experience involves learning engagement between teachers (educators), students (learners) and peer-to-peer opportunities. These learning contexts are not standardised and are differentiated in subtle and nuanced ways. While being able to accurately measure what is done is important, determinations in relation to academic teaching effectiveness must be able to embrace this complexity as part of the teaching and learning narrative that necessarily informs teacher-student interactions.[60]

Authentic tertiary educators are thus more appropriately viewed as facilitators and guide in their students' learning journeys where different individual teachers construct different and varied learning pathways for their students influenced by their own educational stories, experience and scholarship. Measuring and documenting this is challenging but that is not a reason to adhere to mechanisms which bear little correlation to modern academic teaching and learning reality. SETs provide a limited data set recording student perceptions of teaching and a subject at a point in time that must be carefully analysed with due regard for their limitations. However, SETs lack both the agility to adapt quickly to changes wrought by events such as the COVID-19 pandemic and the nuance needed for a tertiary educator to

---

[58] *Ryerson* (n 3).

[59] See, eg, Michael Nancarrow, 'Subversion Through Technology: The Changing Landscape of Property Law Learning and Why I'm Not an Algorithm' (Legal Education Research Conference, Faculty of Law, University of New South Wales, 2019). On the emerging literature on the impact of generative artificial intelligence in the tertiary sector, see, eg, Weng Marc Lim et al `Generative AI and the Future of Education: Ragnarök or Reformation? A Paradoxical Perspective from Management Educators' (2023) 21 (2) *The International Journal of Management Education*: 100790. The Tertiary Education and Quality Assurance Agency (TEQSA) is collating a range of resources on artificial intelligence, see 'Artificial intelligence' *TEQSA* (Web Page, 4 April 2023) <https://www.teqsa.gov.au/guides-resources/higher-education-good-practice-hub/artificial-intelligence>.

[60] Cf Lorraine Daston and Peter Galison, *Objectivity* (Zone Books, New York, 2010), Fig 4.5 'Newton Domesticated' and Figs 7.20-21 Transport II (Image, Nature)' as representations of the themes discussed in this section.

engage in reflective practice and iterative development of their teaching practice and to provide evidence of the quality of their teaching. The stakeholders providing teaching effectiveness evidence should come from both, within and outside, the institution and discipline. LEAD recommends the adoption of a broad approach to teaching effectiveness evidence, consistent with a teaching portfolio and premised on *Ryerson*.[61] According to *Ryerson*,[62] a teaching portfolio is integral to the gold standard for measuring teaching effectiveness. A teaching portfolio should include the PRoT in classes, as well as peer review in other, more diverse forms. This approach is to be encouraged for multiple purposes, both internal to an institution, such as confirmation and promotion processes, and for external recognition, such as the newly developed national teaching and learning awards for the discipline of law, the Australian Legal Education Awards, which ran for the first time in 2021.

Going forward, LEAD aims to encourage law academics to engage more in PRoT, at a school and institutional level. Additionally, LEAD aims to engender a sense of empowerment in law academics by enabling cross-institutional PRoT in the law discipline, and support colleagues to develop and implement ways in which students can be exposed to structured learning opportunities on how to provide effective constructive feedback on their learning.

Firstly, LEAD will develop a national, fit-for-purpose PRoT program that is safe, collegial and sustainable. This program will be voluntary and allow individual academics to seek review of any aspect of their teaching practice on which they wish to receive feedback. Importantly, the program will allow law academics who do not have access to a formal peer review program at their own institution, or who seek feedback from peers with expertise not available in their own law schools, to seek peer review. Ideally, Australian law schools will acknowledge and recognise the service and workloads of the peer reviewers and peer reviewees. A trial of the LEAD PRoT program with 10 Australian law schools is expected to commence in 2023. The LEAD Co-convenors will reflect on the trial's outcomes, implement changes and aim to expand the program more broadly, if there is sufficient interest.

Secondly, LEAD intends to develop a community of practice whereby law academics participate in regular dialogues to share their teaching and peer review practices and grow this knowledge in the law discipline.[63] Ground rules will be co-created amongst participants and reinforced as necessary to ensure that the inter-institutional discussions are safe, collegial and sustainable. Further, LEAD will seek the support for this initiative from the Council of Australian Law Deans and, thereby, the Australian law schools.

---

[61]   *Ryerson* (n 3).
[62]   Ibid.
[63]   See generally, Jacquie McDonald, *Community, Domain, Practice: Facilitator Catch Cry for Revitalising Learning and Teaching through Communities of Practice*, (ALTC Teaching Fellowship Final Report, Australian Government Office for Learning & Teaching, 2014).

Thirdly, LEAD aspires to develop a multi-media resource that guides students on how to provide, and receive, constructive feedback as a professional skill they can begin to hone while studying.[64] This initiative is imperative, given the lack of appetite to abandon the use of SETs by university executives in the present neoliberal context.

LEAD's proposed initiatives have the potential to strengthen a law academic's teaching portfolio through broadening the scope and depth of available methods of measuring teaching effectiveness and avoiding over-reliance on SETs. A combination of multiple measures, tailored to the learning aims of the subject and teacher, offer the best opportunities for achieving best practice in measuring for teaching effectiveness.

---

[64] Guides on constructive feedback include Douglas Stone and Sheila Heen, *Thanks for the Feedback: The Science and Art of Receiving Feedback Well* (Penguin Books, 2015).

APPENDIX A: LEGAL EDUCATION ASSOCIATE DEANS
NETWORK (LEAD) RECOMMENDATIONS ON BEST
PRACTICE IN STUDENT EVALUATIONS OF TEACHING
DURING COVD-19

The following recommendations are based on consultation with the LEAD network. We note that there is a variety of views, and that particular concerns may vary from institution to institution. The recommendations are intended to help frame discussion on this issue and provide guidance on best practice that is broadly supported by LEAD members.

A   *LEAD supports the suspension of SETS during COVID-19*

LEAD supports the suspension of Student Evaluations of Teaching (SETs) for the duration of the impact of COVID-19 on the delivery of previously face-face law programs that are now being delivered remotely. LEAD encourages law academics, including sessional academics, to be resourceful in identifying opportunities for scholarship in these challenges times, including through collecting and recording information about learning, teaching and assessment strategies developed in response to the unprecedented changes as a result of COVID-19.

Where institutions, or law schools, do seek to evaluate the remote delivery of teaching during COVID-19, we encourage law schools to work with teachers to develop tailored feedback tools that will meet the specific purposes of the evaluation of each subject/course/unit, which may include short polls to obtain student feedback at key points during semester; discussion fora (e.g. anonymous discussion boards on the subject's Learning Management System (LMS)) and invitations to students to provide responsible, reasonable and constructive feedback to teachers in the spirit of a learning community. These suggestions are not exhaustive.

B   *Where SETs are administered during COVID-19*

LEAD acknowledges that institutions may seek to use SETs for data collection purposes rather than developing COVID-19-specific evaluation instruments. Where institutions do elect to continue the use of SETs, LEAD *strongly recommends* these are *not* used for performance evaluation or for confirmation or promotion purposes without the explicit agreement of the academic/teacher who is the subject of the SET. Where SETs are used with the permission of the relevant teacher, LEAD's position is that SETs conducted during the COVID-19 period should be clearly identified as such, in order to guard against inappropriate comparisons to prior or subsequent SETs conducted within the same subject/course/unit or about the same teacher and to ensure that the COVID-19 SET results are interpreted in context.