

9 January 2019

Dear Professor Simmons, Professor Fisk, Professor Crossley and Mr Nairn,

NTEU Submission: Quantitative Benchmarks

The NTEU UNSW branch makes this submission on the proposed Quantitative Benchmarks based on extensive feedback from our members and scholarship on the construction, application and effects of individualised quantitative metrics in higher education. Nationally, the NTEU has long contributed to policy development, debate and industrial best practice in the design and application of performance evaluation frameworks. We hope this expertise will assist the University to develop a collaborative approach to meet our mutual goals of a fair, equitable and supportive research and teaching environment. We set out alternative principles to guide the development of these goals in Appendix A.

NTEU members consider that the proposed Quantitative Benchmarks are in conflict with workplace rights in relation to workload, performance management, academic freedom and consultation commitments. If the University proceeds to implement these proposals, we cannot see how current consultation commitments, academic freedom commitments, academic workload clause obligations, professional development and performance management obligations will not be contravened by the University. We make detailed submissions on the industrial implications of these proposals later in the submission, and provide the opportunity for University representatives to avoid potential legal proceedings, and indeed to avoid allegations of University management representatives knowingly contravening the terms of our enterprise agreement.

In spite of the short feedback period, that overlapped with the end of year shutdown, our members resolutely urge that the proposed framework be abandoned and with it, the counter-productive, top-down approach employed in its development to date. Over 600 members of the University community have signed a petition calling for UNSW to work together with staff to collegially develop an alternative vision to the Quantitative Benchmarks for supporting Academic research and teaching. This commitment to democratic, participatory governance underpins all of the UNSW values that staff are expected to embrace and we would welcome a similar demonstration of these values from University management.

We welcome the opportunity to meet with you to discuss these matters, once you have considered our submissions.

Yours sincerely

Dr Sarah Gregson
President

Dr Vicki Sentas
Vice President (Academic)

NTEU UNSW Submission to the Proposed Quantitative Benchmarks

Executive Summary

The NTEU supports the pursuit of advancing research and teaching, consistent with the public purpose of the University. The key mechanism for facilitating strong Academic cultures must be with the provision of collegial, supportive, non-coercive research and teaching environments consistent with workplace rights. The proposed Quantitative Benchmarks seriously undermine this approach and present risks for productive and dynamic research cultures, academic freedom, democratic governance, safe and sustainable workloads and staff wellbeing.

These proposals have elicited widespread distress and opposition from staff. They come after successive changes that have unnecessarily increased managerial controls, intensified workloads and damaged staff morale. These include the new myCareer framework that has increased performance review from once to three times a year; the controversial BORIS research ranking system and its new teaching equivalent SET; and successive workplace change proposals that have resulted in job losses for professional staff. Academic and professional staff have made enormous sacrifices in preparation for the widely unpopular Trimester teaching arrangement, often at great personal costs. Yet, we collectively still produce the quality outputs and impacts that give the University its esteem and reputation. The University has given no outline of any problems to which these proposals are a purported solution. Nor is there any evidence that the proposals are capable of supporting or improving Academic culture or extrinsically motivating 'stronger' performance.

The proposals are at odds with research findings and best practice internationally on how to support and evaluate teaching and research. The benchmarks are unachievable, unfair and inequitable. They lack transparency in design and breach key provisions of the enterprise agreement including workloads, performance management, consultation, and academic freedom. The proposals have elicited justifiable dismay that UNSW is not on the right path to support staff to produce great research and teaching. Worryingly, the proposals contribute to a deeply felt concern that the University does not trust Academics to determine their own research objectives, methodologies and decisions on where and how to publish or apply for grants. Nor do the proposals recognise the lack of Academic control over the award of funding grants or that student satisfaction surveys bear little relation to actual teaching quality and effort.

If implemented, the proposals will foreseeably affect research direction and generate perverse dynamics in order that Academics meet the targets to demonstrate they are successful. As a consequence, Academic freedom will be undermined and research processes reshaped. Quantitative measurement as a means of imposing greater levels of institutional managerialism over research outputs, risks eroding relations of trust, staff morale and institutional reputation.

Recommendations

We recommend that UNSW:

1. Not proceed to implement any performance framework based on individual quantitative targets. Instead, the recently introduced qualitative Academic Expectations Framework and myCareer should be given time to be implemented before being evaluated in 2020, including through a survey of staff. The purpose and design of the evaluation should be in consultation with Academic staff and the NTEU.
2. Commit to best practice principles on research and teaching evaluation at UNSW as outlined in *Appendix A* of this submission.
3. Sign the *San Francisco Declaration on Research Assessment (DORA)*, joining 856¹ Universities, societies and scholarly journals internationally in the commitment “to improve the ways in which the output of scientific research is evaluated”.²

If after considering all the feedback from the UNSW community, the University proceeds to implement the framework of Quantitative Benchmarks, we recommend that UNSW pause to:

4. Initiate a bottom up, democratic and collegially developed consultation process at the School and Faculty level over the course of 2019 that enables all staff to participate in the development and decision making of any disciplinary appropriate framework. Any new process and proposal must be consistent with all commitments in the enterprise agreement.

¹ As of 8 January 2019. <https://sfdora.org/signers/>

² *San Francisco Declaration on Research Assessment (DORA)* <https://sfdora.org/read/>

Introduction

This submission is made in four parts:

- Overall concerns with quantitative metrics
- Concerns with the proposed indicators
- The parties' obligations under the Academic Staff Enterprise Agreement 2015 (the 'agreement')
- Principles for supporting Academic work and evaluating performance (Appendix A).

1. Overall Concerns with Quantitative Metrics

The quantitative benchmarks comprise seven interrelated metrics, across three broad categories:

- Research Quality and Quantity: Cat 1 Funding (\$), Cat 2-4 Funding (\$), number of quality outputs, FWCI, HDR Completions
- Teaching Quality: Student Satisfaction (%)
- Teaching Quantity: Teaching Volume (per annum, per week, etc.)

We understand UNSW's stated aim is to use these metrics as aspirational measures for 'strong' performance, as a reference point for career development conversations as part of the MyCareer process, and for promotion. The documentation provided to staff indicates a number of explicit and implicit objectives – the 'continuous improvement' of Academic performance, reflected in an increase in research quantum and quality, as a measure of research productivity. We presume the proposals are intended to increase UNSW's global rankings.

UNSW has not provided evidence to support its contention that the proposed new framework will achieve these goals. Whether these goals are indeed, the appropriate drivers for frameworks assessing individual academic performance is highly controversial. Global rankings have been found to be based on 'inaccurate data and arbitrary indicators'.³ More fundamentally, the goal of using individual career development and performance review systems to meet institutional objectives to rise in global rankings is in our submission, unethical. These practices have been subject to criticism: "Some HEIs effectively outsource crucial aspects of performance management to league table providers, by incorporating their rank position as a specific organisational target."⁴ UNSW's recent moves to incentivise behaviour through payments to individuals who are published in *Nature* and *Science* and for citation counts, reflects another strategy to influence its position in global rankings.

³ Hicks, D et al (2015) "The Leiden Manifesto for Research Metrics", *Nature*, vol 520, 22 April 429-431.

⁴ Wilsdon, J (2015) et al *The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management*, 73.

The use of quantitative metrics is designed to shape Academic endeavour towards competitive, market behaviour that is anathema to collaborative, collegial work. There is compelling and substantial evidence that individualised quantitative metrics have had a counterproductive effect on research quality, academic freedom, staff morale, and workplace health and safety. The controversial use of quantitative indicators and their negative impacts have resulted in the development of a number of well regarded international statements, manifestos and best practice reports.⁵ Whilst there is some qualified support in the literature for carefully designed quantitative indicators to supplement primary qualitative assessments, there is little evidence supporting a causal connection between quantitative indicators and research improvements.

In sum, the evidence is that quantitative indicators and 'outcome based performance management' can generate the following deleterious effects:

i) Impact on research practices

The key problem is that the Quantitative Benchmarks cannot account for the complexity and heterogeneity of scholarly knowledge production. In short, the established system effect is that numeric targets become the instrumental goal, both for managers and for staff, skewing research practices towards meeting target indicators. There are grave risks that the metrics will not only misrecognise core Academic work, but will reshape this work. Wilsdon et al in their foundational study *The Metric Tide*, suggest:

“Some of the most precious qualities of academic culture resist simple quantification, and individual indicators can struggle to do justice to the richness and plurality of our research. Too often, poorly designed evaluation criteria are dominating minds, distorting behaviour and determining careers.”⁶

The key distorting effect of numeric targets on academic cultures emerges when researchers select instrumental forms of research design and publication outputs. For example, 'goal displacement' describes the strategic behavioural response by researchers individually and collectively, whereby scoring high on the assessment criteria and measures becomes a goal in itself, rather than a means to assess whether an objective was reached.⁷ Scholars point to the transformation of the research process itself in response to metrics, such as the production of risk avoidance in selecting research topics and methodologies.⁸

⁵ DORA op cite 2; Hicks, op cite 3 ; Wilsdon, J (2015) et al *The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management*.

⁶ Wilsdon, op cite 4.

⁷ Muller R and de Rijcke S, (2017) “Exploring the epistemic impacts of academic performance indicators in the life sciences”, *Research Evaluation* 26(3) 157.

⁸ Wouters, P et al (2015) *The Metric Tide: Literature Review: Supplementary Report I to the Independent Review of the Role of Metrics in Research Assessment and Management*: 30-31.

Innovation is stunted because the qualities and social processes of academic inquiry that escape metrification become undervalued. Further, the mismatch between academic work and the metrics result in 'task reduction' where particular types of work are abandoned.⁹ In particular the work required for generating social impacts and public goods, including applied research is jeopardised through metric based evaluation. In this way, quantitative indicators function in practice as performance criteria against designated targets, rather than based on a consistent evaluation logic, in the way that qualitative criteria can. A commitment to meeting the goals of the 2025 strategy ought to begin with evaluating the current managerial systems to support research and teaching already in place (see **Recommendation 1**).

ii) Adverse effects on staff health and wellbeing:

Our members report that in the majority of Schools, the Benchmarks require an increase in workload through increased output and funding requirements to demonstrate strong performance. Our members anticipate the negative impacts of increased workload required to meet these increased output and funding expectations. This is in an environment where workloads have already increased substantially over the last few years. The 'more is better' approach explicit in the proposals are in conflict with UNSW policies on work/life balance, stress management, preventing bullying and promoting well being. Studies identify that systems driven by constantly improving performance are premised on intentionally generating anxiety as an imposed form of motivation.¹⁰ The negative effects of individual competition on psychological and physical health, on collaboration, and on productivity, are among the reasons why individual metrics have been abandoned in management practice in other sectors.

iii) "Destroys the trust between managers and staff necessary for optimal performance".¹¹

'Audit based control' erodes traditional forms of trust-based governance and decreases academic autonomy, and along with it, the capabilities for innovative research and teaching practices.¹² This is a particular risk in the design of the Quantitative Benchmarks. Underlying the proposals is the rationale that Academics are in need of external motivation in the form of quantitative targets. However, the literature indicates that Academics are more likely to be intrinsically motivated, connect with different types of 'incentives' incapable of measurement and flourish in high trust environments.¹³ The unachievable nature of the targets widens the gap between management and staff. The benchmarks misrecognise and

⁹ Laudel, G and Glaser, J. (2006) 'Tensions between Evaluations and Communication Practices', *Journal of Higher Education Policy and Management* 28(3): 289–95 at 294.

¹⁰ Hall and Bowels 2016; Morrish, L et al (2017) "Academic Identities in the Managed University: Neoliberalism and Resistance at Newcastle University, UK" *Australian Universities Review*, vol 59, no 2.

¹¹ Lowe T, and Wilson, R. (2015) "Playing the Game of Outcome-Based Performance Management (OBPM). Is Gamesmanship Inevitable? Evidence From Theory and Practice." *Social Policy and Administration*.

¹² Hoecht, A (2006) "Quality Assurance in UK Higher Education: Issues of Trust, Control, Professional Autonomy and Accountability", *Higher Education* 51(4), 541-563.

¹³ Lowe and Wilson op cite 11; Hoecht op cite 12.

devalue what matters in knowledge production, and risk demotivating and alienating Academics.

iv) Impairs transparency, accountability, equity and diversity

The NTEU requested the methodology and data underpinning the proposals, which UNSW declined to provide.¹⁴ The lack of transparency in the development of the proposals reflects broader concerns shared by researchers in the field of performance indicators:

“...there is a need for greater transparency in the construction and use of indicators, particularly for university rankings and league tables. Those involved in research assessment and management should behave responsibly, considering and pre-empting negative consequences wherever possible, particularly in terms of equality and diversity”.¹⁵

The shift towards numeric targets discriminates against researchers who deploy objectives, methods and conceptual frames that do not lend themselves instrumentally to meeting the logic implicit in each indicator. The benchmarks will particularly disadvantage those from groups who due to diverse structural and systemic factors do not have a high-output research profile or difficulty in maintaining one. The incorporation of ROPE considerations into the framework will not mitigate the systemic impacts of quantitative performance management, outlined above.

2. Concerns with the proposed specific indicators

2.1 Publication quantity and quality

Numeric publication targets have been widely critiqued as generating negative effects on diverse aspects of research culture. Evidence of any relationship between quantitative indicators and research quality is limited.¹⁶ However, the evidence suggests that setting high bibliometric targets compromises quality due to the displacement effect of high numbers.¹⁷

Our members overwhelmingly report concerns that the proposals represent an increase in expectation of publication outputs. Our members advise that across the majority of Schools, the proposed new quantity of publications are excessive, unachievable and contrary to disciplinary norms. In most Schools, the numeric targets do not reflect disciplinary practice, custom and shared understanding of the contextual, relational and process dynamics of knowledge production better evaluated through qualitative frameworks or peer review. Our members advise that the amount of publications required for ‘strong performance’ would undoubtedly deter longer, more complex projects and methodologies, unorthodox research topics, or publishing in the most appropriate journals

¹⁴ UNSW letter 20 November 2018.

¹⁵ Wilsdon, op cite 4: ix.

¹⁶ Ibid, ix.

¹⁷ Colwell R, et al (2012) *Informing Research Choices: Indicators and Judgment*. The Expert Panel on Science Performance and Research Funding, Ottawa: Council of Canadian Academics.

of research communities to which Academics freely and expertly choose to belong. Our members' feedback cautions that in order to meet the targets for being 'good', Academic choice regarding methodology, design and publication options is already being curtailed and that this will be exacerbated by the proposed performance management system. Many of our members anticipate that 'faster' and 'easier' or 'descriptive' work will become the publication default in order to meet the numeric targets.

2.2 Citation impact metrics

We understand that the University seeks to generate a high number of *quality* publications. Quality is understood by measuring individual publications through the Field Weighted Citation Index (FWCI), and the SJR (SCImago Journal Rank). In brief, the FWCI compares citations relative to the world average for the field subject, publication type and year. The SJR ranks journals included in SCOPUS and generates journal level citation data based on calculating the proportion of publications that belong to the top 10% or the top 1% of the most highly cited in their field.¹⁸

These are only two of a diverse range of citation impact metrics available. The rationale for selecting these particular indicators ought to be set out in the context of long standing debates on the meaning of particular citation measures. Even amongst experts who support the careful use of citation metrics as supplement to holistic evaluation, "best practice uses multiple indicators to provide a more robust and pluralistic picture."¹⁹ We anticipate the University may argue that the FWCI and the SJR represent more nuanced calibrations (known as 'normalised field variants') of the widely debated journal impact factor. We note that the favouring of normalised indicators based on highly cited publications, are central to calculating some of the most influential *University* rankings, including the SCImago Institutions Rankings.²⁰ We briefly outline why using citation impact indicators as proxies for 'quality' are inappropriate in the performance evaluation of individuals.

In practice, the object of these measures is to influence academic research and publication practices in order to, i) publish in particular journals and ii) to encourage publications capable of generating high FWCI scores. The instrument becomes the goal. We submit these objectives are flawed, do not have scientific merit and are inappropriate 'incentives' for academic research.

DORA cautions that "citation-based metrics should not be interpreted as a direct measure of research quality".²¹ Quality is understood by bibliometricians as a multidimensional concept. The particular aspects of quality constructed and prioritised in evaluations vary across different fields, and even within fields, dependent on diverse research objectives.²²

¹⁸ SCImago, (n.d.). <http://www.scimagojr.com>; Wouters, op cite 8: 17, 39.

¹⁹ Hicks, op cite 3.

²⁰ Wouters, op cite 8: 17.

²¹ <http://www.metrics-toolkit.org/field-normalized-citation-impact/>

²² Wouters op cite 7: viii.

In contrast to a multidimensional, context specific approach, UNSW states in the proposal documentation that the FWCI “is appropriate for citation based disciplines”. The University gives no recognition of how the FWCI as a measure would operate in relation to non-citation based disciplines. UNSW’s reliance on both the FWCI and the SJR index will discriminate against ‘non-citation’ disciplines where books are central forms of knowledge, and those disciplines where key journals are not indexed in SCOPUS. FWCI also only measures those publications captured by the SCOPUS database. Critically, the SJR indicator operates on the basis that:

“...in the calculation of citation impact indicators for journals more weight should be given to citations from high-impact sources, such as citations from Nature and Science, than to citations from low-impact sources, for instance from a relatively unknown national journal that receives hardly any citations itself.”²³

The intellectual freedom to publish in journals appropriate for one’s research, rather than in journals ranked highly by SCOPUS, is at risk when citation impact metrics are used to evaluate Academic performance.

Elsevier, the company that owns SCOPUS, acknowledges that “Field-Weighted Citation Impact calculates an average value, and these types of calculations are strongly influenced by outlying publications in a small data set.”²⁴ Skewed distributions are unhelpful as a basis for benchmarking individuals. Critically, the SJR is a journal level indicator that is sought to be used as a substitute for individual publication level citations. Wouters found that “..many bibliometric experts reject the use of journal level indicators for assessing individual publications”:

“The most important argument against this practice is that the citation impact of a journal offers only a weak predictor of the citation impact of individual publications in the journal. This is because the distribution of citations over the publications in a journal tends to be highly skewed, with for instance 20% of the publications receiving 60% of the citations. The average number of citations of the publications in a journal is therefore determined mainly by a small proportion of highly cited publications, and most publications in a journal have a citation impact that is substantially below the citation impact of the journal as a whole. Hence, the citation impact of a journal is not representative of the citation impact of a typical publication in the journal. This argument against the use of journal-level indicators for evaluating individual publications has received widespread support in the bibliometric literature. The work by Seglen (1992, 1994, 1997) on this topic has been especially influential. The inventor of the impact factor also warns against the use of this indicator for evaluating individual publications (Garfield, 1996b, 2006).”²⁵

²³ Wilsdon, op cite 4: 50.

²⁴ Elsevier, *Research Metrics Guidebook* at 47.

https://www.elsevier.com/__data/assets/pdf_file/0020/53327/ELSV-13013-Elsevier-Research-Metrics-Book-r5-Web.pdf

²⁵ Wouters op cite 7: 28.

A widely cited solution to address the recognised flaws in citation counts/impact measures is the substantive use of peer review through qualitative indicators so that assessment of individuals *as individuals* is based on the scientific content and context of their work.

DORA recommends that Institutions:

“4. Be explicit about the criteria used to reach hiring, tenure, and promotion decisions, clearly highlighting, especially for early-stage investigators, that the scientific content of a paper is much more important than publication metrics or the identity of the journal in which it was published”.

5. For the purposes of research assessment, consider the value and impact of all research outputs (including datasets and software) in addition to research publications, and consider a broad range of impact measures including qualitative indicators of research impact, such as influence on policy and practice.”

UNSW already has set out a range of qualitative indicators in its relatively new Academic Expectations Framework (AEF). It would be prudent to allow both the AEF and the new myCareer system time to function and then to evaluate these new regimes, in order to determine their effectiveness against mutually agreeable goals, before introducing yet another highly contested framework or ‘moving goalposts’ likely to do more harm.

2.3 Individual research income

Individual, annual, funding targets are a highly inappropriate evaluation tool for Academic research. This metric does not provide for an evaluation of quality research. Given the highly subjective, often arbitrary sets of circumstances that lead to the grant of funds, it is more a measure of luck. The known difficulties in winning grant funding makes the figures set out across Schools, unachievable. Funding pools are limited and ARC success rates are low in both HASS and STEM. More fundamentally, the indicator discriminates against research which is structurally excluded from the goals of funding bodies, especially when political interference is allowed to dictate ‘palatable’ research.

The research income indicator encourages goal displacement and task reduction. Pursuing winnable grants undermines the doing of important research that may not readily meet funders’ priorities and may have the effect of reducing time spent on long term research engagement that builds capacity for future funding. The individual research income indicator erodes cultures of collaboration and collegiality. Many of our members noted the emphasis on winning money would discourage Professors from including ECRs on their teams at the risk of reducing individual quantum. Perversely, in some schools, individuals would be required to win grants equivalent in amount to their annual yearly income to demonstrate strong performance.

2.4 Teaching metrics

The Harvard Business School, which is otherwise in favour of performance metrics, acknowledges three pitfalls of measuring performance that are “sources of inescapable conflict” between employer and employee:

1. **The Uncontrollability Problem.** *Most performance measures ... are the result of factors that are both controllable and uncontrollable. Almost all performance measures contain some level of uncontrollability.*
2. **The Alignment Problem.** *Today’s jobs typically require many tasks; some tasks are easy to measure, but others are not.*
3. **The Interdependency Problem.** *Most outcomes are the result of work by many people, sometimes working together and other times working independently. Isolating the contribution of individual employees within these outcomes is nearly impossible.²⁶*

The proposed Quantitative Benchmarks for teaching introduce two metrics, *Student Satisfaction* (and derived measures T1 and T2) and *Volume of Teaching*, that suffer from many uncontrollability, alignment and interdependency problems.

Student Satisfaction

A core objective of good teaching should be student learning with the goal of contributing to a well-educated and employable graduate. The student satisfaction metric ignores student learning entirely, and is instead akin to a popularity poll (or the number of “likes” a person has on Facebook). There is good evidence that there is no correlation between student learning and student evaluations of teaching.²⁷ Using this as the primary measure of quality teaching appears seriously misaligned with the University's objectives, and its adoption as the primary metric appears to have been justified simply on the basis that it is easy to calculate.

As student satisfaction is the only metric provided to assess teaching quality (*Volume of Teaching* is one measure of work load, not quality), this has the consequence of encouraging staff to “play the game” of increasing student satisfaction at the expense of educational outcomes. Because student satisfaction is aggregated across all courses in a term, this will discourage teachers from innovating, experimenting and taking risks. The short-term consequence of innovation always involves teething problems that can only be overcome in subsequent iterations. However, a low student satisfaction score in one term will raise flags about the staff member's teaching quality.

²⁶ Luecke, R. & Hall, B.J. (2006) *Performance Management*, Harvard Business School Press, 24-5.

²⁷ Uttl, B., White, C.A. & Wong Gonzalez, D. (2017) “Meta-analysis of faculty's teaching effectiveness: Student evaluation of teaching ratings and student learning are not related”, *Studies in Educational Evaluation*, 54: 22-42; Lee, L., Connolly, M., Dancy, M., Henderson, C., & Christensen, W. (2018) “A comparison of student evaluations of instruction vs. students' conceptual learning gains” *American Journal of Physics*, 86(7): 531-535.

Most seriously, use of this metric as the only evaluative metric in the Quantitative Benchmarks contravenes the Enterprise Agreement (Clause 27.0(c)) which states “On its own student feedback cannot be used as the total measure of teaching performance.”

Student satisfaction also suffers from problems of control and interdependency, which inevitably lead to biased results. This metric is therefore fundamentally unfair. First, factors **outside the control** of the individual that influence results include:

- type of course (core/elective)
- student body (undergraduate/postgraduate, local/international)
- delivery mode (face-to-face/web stream/fully online)
- level of study (introductory/advanced)
- class size (small/large)
- quality of resources (e.g. Moodle, lab equipment)
- cohort level of discipline knowledge (depends on prior education)
- cohort level of English ability (depends on UNSW admissions policy)
- timetable (day/evening)
- academic discipline

Personal attributes of an individual, which are obviously outside their control, have been found to bias the results even further²⁸ and are hence discriminatory metrics. These attributes include:

- gender (bias against women)
- ethnicity (bias against minorities)
- age (bias against older people)
- employment status (part-time or casual staff are less available to students)

Second, **interdependency** is a major concern for the validity of survey results for individual teachers. Many UNSW courses, and especially large classes, are taught by teaching teams, and as noted generally by the Harvard Business School, it is “nearly impossible” to isolate individual contributions. This applies especially to the lecturer-in-charge, who is responsible for much work

²⁸ Mitchell, K.M and Martin, J (2018) ‘Gender bias in student evaluations’ 51 *PS: Political Science & Politics* 648; Basow, S and Silberg, N (1987) ‘Student evaluations of college professors: Are female and male professors rated differently?’ 79 *Journal of Educational Psychology* 308; Reid, L (2010) ‘The role of perceived race and gender in the evaluation of college teaching on RateMyProfessors.com’ 3 *Journal of Diversity in Higher Education* 137; Ambady, N and Rosenthal, R (1993) ‘Half a Minute: Predicting Teacher Evaluations from Thin Slices of Nonverbal Behavior and Physical Attractiveness’ 64 *Journal of Personality and Social Psychology* 431; Arbuckle J and Williams, B (2003) ‘Students’ Perceptions of Expressiveness: Age and Gender Effects on Teacher Evaluations’ 49 *Sex Roles* 507.

typically invisible to students such as preparation of tutorial/lab materials, setting assignments and marking rubrics, coordination of teaching and support staff, handling student enquiries and special considerations, etc. Survey responses may fail to take into account this essential work.

The survey results for an individual lecturer's teaching will be influenced by:

- how well students learnt prerequisite material (depending on previous courses)
- the quality of other lecturers (when lectures build upon previous lectures in a course)
- the quality and nature of material delivered by the teacher but not written by them
- the quality of tutors/demonstrators (whose content overlaps with lecture material)
- the quality of feedback provided by tutors/demonstrators
- the timeliness of marking from tutors/demonstrators

Finally, the choice of 90% as the benchmark to *define* “good” performance is arbitrary (no justification for this definition has been provided following a request for this information), and as highlighted above, inevitably leads to unfair staff evaluations, where some staff are more easily able to meet the benchmark than others due to the many uncontrollable factors, biases and interdependency of factors outlined above.

Volume of Teaching

This simple metric does not measure quality of teaching *at all*, so this metric cannot be used to define “strong” or “good” performance. The metric is defined in the documentation as follows: “Measurement of teaching volume is determined by School workload models”. The Enterprise Agreement (Clause 24.2(c)) “*identifies a transparent correlation between the measure applied and the hours of work generated by each relevant academic activity*”. The metric therefore reflects only the *amount of time* a staff member spends in teaching activities, which in no way relates to evaluating performance. In addition, the enterprise agreement requires that teaching loads be distributed equitably and transparently. If this is not the case, management action is required to remedy any potential misapplication of the enterprise agreement – the allocation itself cannot be viewed as an indication of staff willingness or otherwise to take on a particular workload. In short, academic staff simply teach what they are allocated. However, if some staff are given high teaching workloads, it should be taken into account when assessing *research* performance.

The T1 and T2 Indices

These metrics rank staff across the University based on student satisfaction and on number of survey responses. The problems of relying on student satisfaction have already been addressed. Ranking staff on this measure exacerbates these problems. The metric is now no longer standards-based, a requirement in UNSW student assessment, but is instead normative. When this ranking is done across the diverse factors outside staff control (see above) any contribution of the individual teacher is lost in the noise. The same argument applies to the T2 index, where the number of survey responses depends primarily on class size, which is a factor outside the teacher's control.

Both of these measures rely on student satisfaction with teaching, which ignores the key role of course convenors, curriculum and assessment designers, developers of digital content, all of which are key contributors to the digital aspect of Strategy 2025 and all of which are core contributors to student learning. Assumptions that digitalisation will have a universally positive effect on student evaluations of teaching performance must also be questioned, as student comments in evaluations regularly indicate that they have many positive associations with the amount of one-on-one interaction they can have with academic staff.

Furthermore, as many courses are likely to be very similar in size, both indices are highly sensitive to small variations in survey responses and response rates. The use of rankings will amplify these very small differences, leading to a false impression of individual teaching performance. Moreover, the combination of the T1 and T2 indices by taking their average is meaningless: there is no justification provided for the validity of this procedure, and no discussion of what this number (the mean of two ranks) signifies. Furthermore, the small variations that result in misleading T1 and T2 indices will be amplified even more when the T1 and T2 rankings are combined.

3. Proposed quantitative benchmarks and the parties obligations under the Academic Staff Enterprise Agreement 2015 (the ‘agreement’)

The NTEU wrote to the University on 14 November 2018, outlining concerns regarding these proposed benchmarks, seeking relevant information from the University in order that staff could make informed consideration of the University proposals. The University’s letter of 20 November 2018 declined to provide the data sought by the NTEU, and further did not provide the feedback or information that the University asserts was the basis of further substantial revision, asserted to have been received through “engagement across all Faculties”.²⁹ We consider that employees and their representatives (Clause 8 of the agreement) are entitled to be provided with the asserted information and input that was the basis of this revision, in order to make informed contributions to the consultation process. We continue to seek the provision of this relevant information, as per Clause 3.3 of the agreement.

The NTEU expects that after receiving feedback from University staff on 9 January 2019, the University will of course require a period in which to genuinely consider this feedback before advising staff of next steps. Taking time to genuinely consider the variety and significance of the no doubt substantial feedback is required by the definition of consultation in our current agreement. Clause 3.3 Definitions at (c) provides for a definition of consultation: *The University will provide relevant information (orally or in writing), the affected parties will confer, and the views expressed will be taken into account before a final decision is made by the University.*

NTEU members hold concerns regarding the lack of examination or consideration generally of current provisions of the agreement in the development of Quantitative Benchmarks, and

²⁹ UNSW letter 20 November 2018 at p2.

seek further information in relation to professional development and intellectual freedom that we will outline below.

3.1 Proposed quantitative benchmarks and employee entitlements – advice to staff and their representatives

Members are concerned at the impact these proposed benchmarks will have on workload allocation, professional development and performance management, due to the advice from the University to staff and their unions through documents so far.

For example, in the University's letter of 20 November 2018, the University has advised the NTEU that: *The Academic Benchmarks are intended to be used as a reference point to enable academic staff and their supervisors to develop standard, development and/or stretch goals depending on an individual's current performance level as part of myCareer review process. The Academic Benchmarks do not represent a minimum or standard level of research performance within an academic discipline.*

In the above advice to the NTEU, the University does not confirm that these proposed Academic Benchmarks do not measure or provide a benchmark or level for 'strong performance'. Indeed, UNSW management has advised in the 'Quantitative Benchmark Sharepoint published 14 October 2018', that: *'Quantitative Benchmarks are not intended to be a **single tool** that defines academic performance, but as an **indicator of what strong performance looks like** in a particular school or discipline. Quantitative Benchmarks provide academics with a clear reference point to set aspirational targets and have productive MyCareer conversations about their teaching and research. However, meeting (or not meeting) these benchmarks **does not, in isolation**, define the quality of an academic's performance or their career trajectory.'* (our emphasis in **bold**).

There appears to be contradictory messages from the University in the advice to the NTEU on 20 November 2018, versus the information in this document of 14 October 2018, in that the University information to employees asserts that Quantitative Benchmarks will in fact be an indicator of what strong performance looks like, a reference point for aspirational targets, rather than as the University suggests to the NTEU, a reference point for the development of 'standard, development and/or stretch goals.'

Further, the failure to refute Quantitative Benchmarks as a standard of 'strong performance' in the letter to the NTEU of 20 November 2018 appears to be explained by the University's position that it will be a tool, not a single tool, but a component of academic performance, and will not exist in isolation, but will be a component of academic performance and career trajectory. Indeed, the University appears to be clearly advising that the intention of the University would be to use Quantitative Benchmarks as a factor in determining what strong performance looks like, which will contribute to determining both performance and promotion criteria for all Academic employees.

Yet as is apparent, the aspirational targets require employees to increase output, and logically, increase their workload.

This is evidenced in the information to staff with the published date of 14 October 2018 where at page 3 of this document, employees are advised:

The Quantitative Benchmarks provided are for a full-time (1 FTE) academic with a 40:40:20 workload across research, teaching and SEGIL. For consistency, benchmarks should be adjusted proportionately based on factors / considerations including:

- **FTE:** Benchmarks will be scaled based on the academic's average FTE over the course of the year (excluding FWCI and student satisfaction which remain constant).
- **Teaching, Research and SEGIL:** For academics not operating with a 40:40:20 workload, adjustments will be made to align benchmarks with their workload. This will require proportionate increases and decreases to benchmarks, as agreed with the academic's Head of School.
- **Extended Leave:** For academics who have experienced extended periods of time away from the University, adjustments will be made to benchmarks based on individual circumstances. This should be discussed and agreed with the academic's Head of School.

It is clear that employees' workload will be impacted, including increases in workload. Yet the clause which regulates workload is not cited in this section to employees, and no indications are given as to how workloads will be adjusted in compliance with the workload clause in our agreement. We also note that a further publishing date for information to staff was indicated on Sharepoint on 26 November 2018. However we understand the only change in this document from the document published on 14 October 2018 appears to be the extension of feedback to 9 January 2018.

Quantitative benchmarks that require performing a workload which is not reasonable, will not comply with our shared commitments at Clause 24.1 (a) that: *a reasonable level and equitable distribution of workload for academic staff recognising the diversity of the University and the range of activities undertaken by academic staff in the course of a year.* Members are greatly concerned that the University will contravene these provisions if the University proceeds on the basis of proposing employees work above a reasonable level in order to achieve 'strong performance' standards, which will influence employees' performance and promotion prospects.

The objective of the proposed quantitative benchmarks is that employees aspire to 'strong performance' indicators by increasing output, as a consequence exceeding reasonable workloads and exceeding current workload limitations, which in each School or Faculty is required to be measured in a manner: *that identifies a transparent correlation between the measure applied and the hours of work generated.* NTEU members are concerned that the proposed benchmarks actively encourage employees to work above current transparent

measures which generate a limited hours of work, and instead will require working above this hours limit in order to 'exceed expectations'.

There appears to be no consideration that the parties to the Agreement are required to adhere to a workload formula which: *must contain a quantifiable maximum on required workload measured in hours and subject to the provisions of 24.2 (f) a quantifiable maximum on teaching contact hours.* If Academic employees are currently required to perform a quantifiable maximum on required workload measured in hours in order to meet satisfactory performance expectations, how will it be possible to achieve 'strong performance' without working above these limits?

How does the University propose to reconcile the proposed targets with both current Academic workload limitations, and indeed Academic employees' right to enforceable workload maximums? NTEU members are concerned that proposed benchmarks actively seek that members exceed current workload limitations to achieve the 'exceeding expectations' category, and specifically propose to contravene current workload limitations.

The University document 'Quantitative Benchmarks Sharepoint published 14 October 2018' makes clear that Quantitative Benchmarks will be a factor in determining what strong performance looks like, which will contribute to determining both performance and promotion criteria for all Academic employees. This will not only mean that an application of Quantitative Benchmarks would not take into consideration the current agreement provisions of Clause 27 and 28 and the relevant classification standards for Academic employees, but would also not consider the application of current commitments contained in the proposed enterprise agreement.

The NTEU is concerned that if the University was to attempt to proceed with the application of Quantitative Benchmarks, the University would be proposing to contravene the relevant clauses of our agreement as outlined above.

NTEU members are concerned at the impact proposed benchmarks will have on Professional Development commitments to employees. Clause 27.0 (b) provides that in an annual meeting... *The major areas for discussion between the employee and the supervisor will include: (i) career planning and development, including accessing research grants, scheduling Special Studies Program Leave and promotion prospects; (ii) the strategic plans of the School and the Faculty (iii) allocated duties within the School, including teaching and administrative duties; (iv) leave planning; (v) support needed by the employee from the School/Faculty and the University in order to achieve personal career goals.*

The impact of proposed benchmarks on career planning and development, access to Study Leave and promotion prospects, allocated duties within the School, and the impact upon support provided by the School or Faculty in order to achieve personal career goals is clearly affected by the proposal to make Quantitative Benchmarks a factor in determining what strong performance looks like, which will contribute to determining both performance and promotion criteria for all Academic employees.

NTEU members are concerned at the impact of these proposed benchmarks on the commitment made to employees at Clause 23.0 (ii), which recognises the right of an employee to: *pursue critical and open inquiry, publish, research and, consistent with the University's academic processes, freely discuss, teach, assess and develop curricula.*

.....

Appendix A: Principles to Guide Professional Development, Performance Evaluation and Improvement

Any framework for career development, performance evaluation/management and improvement at UNSW, should be designed and implemented according to the following principles:³⁰

DESIGN AND CONTENT

Staff negotiated.

- It must be developed through adequate and meaningful consultation at an institutional level with the NTEU, and at a local workplace level through a transparent process that includes meaningful consultation with all academic staff.

Consistent with the enterprise agreement and all other legal requirements binding the University.

- It should operate within the current framework of enterprise agreements.

Where the imposition of systems of research performance measurement impact upon the character, workload, or autonomy of academic staff, this is a matter that is fundamentally relevant to the employment relationship, and if it is used as a basis for the management of staff performance, it is a matter relevant to codification in the collective agreement.

Based upon a credible methodology and design.

- The process and methodology must be transparent, accountable and dependent upon staff consent;
- Any instrument must be subject to independent review and assessed to account for distortions and unintended consequences;
- It must take account of research inputs as well as outputs with an appropriate balance between inputs and outputs. Output targets must not be central to performance management;
- Peer review and qualitative criteria should be prioritised as the primary basis for assessment and given weight over metric-based alternatives;
- Narrow quantitative indicators like publication count, journal impact factors and citation counts should not be given weight.

Context of research and teaching performance

- The context of an individual's research methodology and objectives should be central to evaluating performance;
- Research/teaching performance should be understood and assessed in the context of other criteria and workload, such as teaching/research, service, impact, engagement;

³⁰ These principles are drawn from: *The Metric Tide*, DORA, the *Leiden Manifesto*; the UNSW Staff Enterprise Agreement (Academic) and the NTEU's study: Kwok, J (2013) *Impact of ERA Research Assessment on University Behaviour and their Staff*, NTEU, April 2013. <www.erawatch.org.au>

- It must preserve the teaching/research nexus (including for Education Focused scholarship and teaching, as recognised in CI24 of the 2018 Enterprise Agreement). Teaching and scholarship must continue to be linked in the process of engaging with new knowledge for university staff and students;

Ethical commitments and non-discrimination

- It must be consistent with the principles of intellectual freedom of inquiry.
- It must explicitly preserve academic freedom and prevent institutions from using performance measurement as a means of shutting down perceived 'non-strategic' research interests of staff;
- It must not produce divisive competition nor impinge on the collegial environment that is essential for the development and rigorous testing of new knowledge;
- It must not disadvantage groups that traditionally have not had a high research profile or experience difficulty in maintaining one;
- It must protect and build Indigenous research;
- It must ensure that early career researchers (ECRs) are not disadvantaged;

Evaluating Teaching

- Student satisfaction surveys must not be used in performance evaluations. They are a subjective representation of student experience, not academic performance.

PERFORMANCE REVIEW AND MANAGEMENT

Fair, transparent process and clear criteria

- Procedural fairness and natural justice must apply to the design of performance criteria as well as the application of criteria and all aspects of decision making;
- Satisfactory criteria ought to be both explicitly defined, and fit for purpose (the vast majority i.e. at least 95% of staff must already satisfy these requirements).
- Clear distinctions between satisfactory and good/outstanding criteria;
- No retrospective performance criteria;
- Performance criteria should be sensitive to career progress and research opportunity and assessed in relation to personal circumstances;
- Performance improvement ought to be conducted with reference to known minimum standards;
- Able to detail the professional development that will assist staff who have not achieved satisfactory performance expectations;
- Should acknowledge that in many work units, all staff will be performing satisfactorily and that no staff member who is meeting expectations/satisfactory, should be put on performance improvement plans;
- BORIS and SEt should not be used as *defacto* performance criteria;
- Data on performance should be verified and checked with staff before any formal process of performance improvement is initiated