City Research Online

City, University of London Institutional Repository

This is the draft version of the paper.

This version of the publication may differ from the final published version.

COMMENTARY

Recommendations for University Module Evaluation Policies

V1. 6 June 2019

Dr Ernesto Priego

Centre for Human-Computer Interaction Design, City, University of London

Ernesto.priego.1@city.ac.uk

Abstract

This paper discusses the limitations of University Module Evaluation processes and shares a series of recommendations that could improve their design and implementation. The paper concludes that regardless of staff gender, age, academic position or ethnic background, no metric or quantitative indicator should be used without thoughtful, qualitative social and organisational context awareness and unconscious bias awareness. The paper concludes there is a need to eliminate the use of Module Evaluation metrics in appointment and promotion considerations.

Keywords

bias; Higher Education; module feedback; Student evaluation of teaching

Module Evaluation

"Module evaluation" refers to the process in which students feedback, assess and rate their academic studies and the quality of teaching on the module (in other countries "modules" might be known as courses or subjects). Below I discuss the limitations of Module Evaluation processes and offers a series of recommendations that I hope could improve their design and implementation.

On "Potential Bias"

Research has shown how internationally "potential bias" against gender and ethnic minorities is real. Holland has described how

"different treatment of male and female staff is increasingly well evidenced: some studies have found that students may rate the same online educators significantly higher if

perceived as male compared to female (MacNell, Driscoll, and Hunt 2015), while other studies have shown that students can make more requests of and expect a greater level of nurturing behaviour from females compared to males, penalising those who do not comply (El-Alayli, Hansen-Brown, and Ceynar 2018)" (Holland 2019).

Research has also suggested "that bias may decrease with better representation of minority groups in the university workforce" (Shepherd et al 2019). However, even if an institution, school or department has good staff representation of (some) minority groups in some areas, it would be important that a policy went beyond mandating support for staff from minority groups to prepare for promotion. The way to tackle bias is not necessarily by giving more guidance and support to minority staff, but by re-addressing the data collection tools and the assessment of the resulting indicators and its practical professional and psychological consequences for staff.

As discussed above the cause for lower scores might be related to the bias implicit in the evaluation exercise itself. Arguably, lower scores can in many cases be explained not by the lecturer's lack of skills or opportunities, but by other highly influential circumstances beyond the lecturer's control, such as cultural attitudes to specific minority groups, demographic composition of specific student cohorts, class size, state of facilities where staff teach, etc.

In my view Universities need policies that clearly state that ME scores should not to be used as unequivocal indicators of a member of staff's performance. The fact that the scores are often perceived by staff (correctly or incorrectly) to be used as evidence of one's performance, that those indicators will be used as evidence in promotion processes, can indeed be a deterrent for those members of staff to apply for promotion. It can also play a role in the demoralisation of staff.

On Student Staff Ratios (SSR), Increased Workloads and Context Awareness

University Module Evaluation policies could be improved by acknowledging that workload and Student Staff Ratios are perceived to have an effect on the student experience and therefore on ME scores.

Though there is a need for more recent and UK-based research regarding the impact of class size and SSR on ME, higher education scholars such as McDonald are clear that

"research testifies to the fact that student satisfaction is not entirely dependent on small class sizes, a view particularly popular in the 1970s and late twentieth century (Kokkelenberg et al.,2008). Having said that, recent literature (post-2000) on the issue is focused heavily on the detrimental impact raised SSRs has on students, teachers and teaching and learning in general. The Bradley Review of higher education in Australia was just one 'voice' amongst many in the international arena, arguing that raised SSRs are seriously damaging to students and teachers alike" (McDonald 2013).

Module Evaluation policies should take into account current settings in Higher Education in relation to student attitudes to educational practices, including expectations of students today, communication expectations established by VLEs, mobile Internet, email and social media.

Raised SSRs do create higher workload for lecturers and have required new workload models. Raised SSRs imply that lecturers may not be able to meet those expectations and demands, or be forced to stretch their personal resources to the maximum, endangering their wellbeing beyond all reasonable sustainability. As I discussed in my previous post (Priego 2019) the recent HEPI Report on Mental Health in Higher Education shows "a big increase in the number of university staff accessing counselling and occupational health services", with "new workload models" and "more directive approaches to performance management" as the two main factors behind this rise (Morish 2019).

Module Evaluation policies could do well to recognise that time is a finite resource, and that raised SSRs mean that a single lecturer will not be able to allocate the same amount of time to each student if there were lower SSRs. Raised SSRs also mean that institutions struggle to find enough appropriate rooms for lectures, which can also lead to lower scores as they impact negatively the student experience.

Who is being evaluated in multi-lecturer modules?

As part of context awareness, it is essential any interpretation of ME scores takes into account that various modules are delivered by a team of lecturers and often TAs and visiting lecturers. However, in practice the ME questionnaires are standardised and often outsourced and designed with individual session leaders in mind and generic settings that may not apply to the institution, school, department, module or session which is the setting and objective of the evaluation.

Regardless of clarification in the contrary, students often evaluate the lecturer they have in front of them that specific day in which they complete the questionnaires, not necessarily the whole team, and if they do the questionnaire's data collection design does not allow for distinguishing what member of staff students had in mind.

Hence module leaders of large modules can arguably be penalised doubly at least, first by leading complex modules taught to many students, and second by being assessed for the performance of a group of peers, not themselves alone. Any truly effective ME Policy would need to address the urgent need to periodically revise and update MEQ's design *in consultation with the academic staff that would be evaluated with those instruments*. Given who mandates the evaluations and their role in other assessment exercises such as rankings or league tables, a user-centred approach to designing module evaluation questionnaires/surveys seems sadly unlikely, but who knows.

Module Evaluation scores are more than just about staff performance

As we all know teaching is never disconnected from its infrastructural context. Room design, location, temperature, state of the equipment, illumination, level of comfort of the seats and tables, and importantly, the timing (stage in the teaching term, day of the week, time of the day, how many MEQs students have completed before, whether examinations or coursework deadlines are imminent or not) have a potential effect on the feedback given by students. ME policies would be more effective by acknowledging that academic staff do not

teach in a vacuum and that many factors that might affect negatively the evaluation scores may have in fact very little to do with a member of staff's actual professional performance.

Module Evaluation assessment done well

Members of staff potentially benefit from discussing their evaluation scores during appraisal sessions, where they can provide qualitative self-assessments of their own performance in relation to their academic practice teaching a module, get peer review and co-design strategies for professional development with their appraiser.

When done well, module evaluation scores and their discussion can help academics learn from what went well, what could go even better, what did not go as well (or went badly), interrogate the causes, and co-design strategies for improvement.

However, any assessment of module evaluation scores should be done in a way that takes into consideration a whole set of contextual issues around the way the data is collected. How to address this issue? Better designed data collection tools could address it, but it would also be much welcome if module evaluation policies stated that scores should never be taken verbatim as unequivocal indicators of an academic's performance.

Conclusions

University Module Evaluation policies should acknowledge module evaluation scores can be potentially useful for staff personal professional development, particularly if the the data collection mechanisms have been co-designed with staff with experience in the evaluated practice within the context of a specific institution, and the discussion takes place within productive, respectful, and sensitive appraisal sessions.

Policies should acknowledge that, as indicators, the evaluation scores never tell the whole story and, depending on the way the data is collected and quantified, the numbers can present an unreliable and potentially toxic picture. The objective of the evaluation should be to be a means to improve what can be improved within a specific context, not a measure of surveillance and repression that can potentially affect more negatively those who are already more likely to be victims of both conscious and unconscious bias or working within already-difficult circumstances.

Regardless of staff subject area, gender, age, academic position or ethnic background, no metric or quantitative indicator should be used without social and organisational context awareness and unconscious bias awareness.

To paraphrase the San Francisco Declaration on Research Assessment, I would argue there is a "need to eliminate the use of [Module Evaluation] metrics in funding, appointment, and promotion considerations" [DORA 2012-2018].

Competing Interests

The author has no competing interests to declare.

Author's Note

This article is shared in a personal capacity and does not represent the views of colleagues or employers. I share these ideas here as a means to contribute publicly to a larger scholarly dialogue which is not only inter-disciplinary but inter-institutional and international. An earlier version of this text was originally published as a post on the author's personal blog, dated 6 June 2019, available at https://epriego.blog/2019/06/06/thougths-university-module-evaluation/ [Accessed 6 June 2019].

References

Fan Y, Shepherd LJ, Slavich E, Waters D, Stone M, et al. (2019) Gender and cultural bias in student evaluations: Why representation matters. PLOS ONE 14(2): e0209749.https://doi.org/10.1371/journal.pone.0209749

Holland, E. P. (2019) Making sense of module feedback: accounting for individual behaviours in student evaluations of teaching, Assessment & Evaluation in Higher Education, 44:6, 961-972, DOI: 10.1080/02602938.2018.1556777

McDonald, G. (2013). "Does size matter? The impact of student-staff ratios". Journal of higher education policy and management (1360-080X), 35 (6), p. 652. http://0-www.tandfonline.com.wam.city.ac.uk/loi/cjhe20

Morish, L. (23 May 2019). *Pressure Vessels: The epidemic of poor mental health among higher education staff* , HEPI Occasional Paper 20. Available from https://www.hepi.ac.uk/2019/05/23/new-report-shows-big-increase-in-demand-for-mental-health-support-among-higher-education-staff/ [Accessed 6 June 2019].

Priego, E. (30/05/2019) Awareness, Engagement and Overload – The Roles We All Play. Available at https://epriego.blog/2019/05/30/awareness-engagement-and-overload-the-roles-we-play/ [Accessed 6 June 2019]

San Francisco Declaration on Research Assessment (2012-2018) https://sfdora.org/ [Accessed 6 June 2019]