Summative Peer Review of Teaching

Crisp, Geoffrey

PVCE Portfolio, University of New South Wales, Australia.

Abstract

We describe the introduction of a summative peer review of teaching process at the institutional level for the purpose of providing additional, independent evidence of the quality of teaching for teaching awards and academic promotion. This paper will describe the introduction of a formal processes at two universities where the peer review reports are used for decision making purposes. We describe why it is important to separate formative peer review of teaching for professional development and self-improvement purposes from summative peer review for high stakes decision making purposes.

Keywords: peer review of teaching, summative peer review of teaching.

Introduction

The criteria and standards used for academic promotion processes within a university is a formal statement by the institution on what is valued; it is the message that academics understand when it comes to knowing what their institution will reward and the promotion criteria will determine how most academics will allocate their limited time and resources. Academics often feel that research is privileged over teaching and/or service when it comes to promotion because it is considered easier to define comparative quantitative measures of quality in research, despite reservations about the current research metrics (Visser-Wijnveen, *et al*, 2014). Peer review in research is seen as independent evidence by experts in the field on the quality of the reviewee's work. It is possible to have independent peer reviewed evidence of the quality of teaching but for face to face teaching activities that are observed it is not possible to have blind peer review unless the reviewers are physically separated from the reviewee and students.

Formative peer feedback for professional development and improving an aspect of one's teaching is now quite common in universities (Bell, 2012), but the use of peer review of teaching for summative or decision making purposes is not as common and continues to be resisted by many academics (Iqbal, 2013).

The most commonly used form of feedback on teaching comes from student surveys which are undertaken routinely at most institutions across the world. Students can provide feedback on their experiences of the academic's teaching and this is commonly used as evidence in teaching award applications and academic promotion (Smithson, *et al*, 2015). However, students provide evidence of their experience and their perceptions of the quality of the teaching and the academic; students normally do not evaluate the academic. Evaluation implies expert knowledge and understanding, it assumes the reviewer is appropriately qualified to evaluate against criteria that are clearly understand by both the reviewer and the reviewee. As important as student feedback is in the university quality cycle, we must be cognizant of its purpose – to provide students with an opportunity to reflect and inform the institution on their experience of the teaching.

This paper describes the introduction of a formal, summative peer review of teaching process at two universities and the lessons learnt from the large scale introduction of the process. The summative peer review of teaching fills a current gap in the quality cycle in many universities as it provides a more formal, structured process of independent evidence against specific criteria and attempts to minimize personal opinion of teaching quality.

Summative peer review of teaching process

The genesis of the methodology for this summative peer review process was an Office for Learning and Teaching national project that sought a process for peer review of teaching

for academic promotion (OLT, 2006). The whole of university summative peer review of face to face teaching was introduced initially at a large, comprehensive institution, RMIT University in Australia. RMIT had previously adopted a successful formative process of peer feedback of teaching called peer partnerships (Chester, *et al*, 2013). This process was voluntary and collegial with individual academics choosing their peer partner and working with them to mutually agree on the aspects of teaching to be reviewed. Importantly, the reports from the peer partnership process belonged to the reviewee and were not required for academic promotion or teaching awards but could be used as evidence of a commitment to continuing professional development.

A formal decision was made to have two distinct peer feedback processes with separate names, peer partnerships for the formative process and peer review for the new summative process. Both processes were important and had a crucial part to play in the quality cycle of the university, but they served different purposes; it is important to label processes clearly so that all the participants are aware of the outcomes from the activity and what can be expected to happen with the peer review reports.

At RMIT the development of the summative peer review documentation followed a lengthy consultation process involving a working group with representation from students, Human Resources, and academics. When draft documentation was developed this was sent to a wider group of academic stakeholders for feedback. Through an iterative process the documentation and the details of the methodology were refined and an implementation plan for the review of face to face teaching was approved. The RMIT documentation consisted of nine core dimensions of teaching (the criteria) and they are based on literature precedence for active learning and the promotion of student engagement (RMIT, 2017). The peer review report consists of both "quantitative" and "qualitative" components. The "quantitative" section is not a numeric scale but rather an indication of the volume of evidence observed during a single session of face to face teaching; no apparent examples, some examples, many examples and extensive examples. Any type of face to face session could be observed, including lectures, tutorials, studios, workshops, team teaching, seminars, laboratory classes and two peer reviewers were present at the same session. One peer reviewer was a broad discipline expert and the other was a specialist in learning and teaching. The "qualitative" component relates to the apparent effectiveness of the examples in the particular context being observed; effectiveness not clear, effective, very effective, exceptionally effective. It is made very clear to the reviewers during the training sessions that they are not there to provide a personal opinion of the quality of the teaching, but rather as an independent observer documenting what they have seen for this particular session against the specified criteria.

The appointment of appropriate peer reviewers is an important part of the overall process as both the institution and the reviewees must have confidence in the chosen reviewers. This is

where the formative and summative processes are quite different. In a formative process it is common for the reviewee to choose their reviewer and the dimensions of teaching to be reviewed. For this summative process the reviewee cannot choose their reviewers, but does have a right of veto over their nominated reviewers if there is a conflict of interest. Peer reviewers were chosen based on their known evidence for scholarship in learning and teaching, publications and grants in learning and teaching, the receipt of teaching awards or teaching fellowships or having held positional responsibility for learning and teaching within the institution. The names of the approved peer reviewers were publicly available on the institutional web site and being nominated as a peer reviewer was a measure of esteem.

Potential peer reviewers had to participate in a training workshop where a series of videos of different teaching situations were analysed for instances of the stated dimensions of teaching and whether the examples appeared to be effective from the students' perspective. As expected, there were often significant differences amongst the academics on the examples and what constituted effectiveness. The purpose of the workshop was to have an open and honest discussion on these differences and to move academics towards a consensus on what contextual evidence and effectiveness looks like. A minimum of two videos and often three were required before broad consensus was reached. The selection of appropriate peer reviewers and the training process for both reviewers and reviewees proved critical to the acceptance of the overall process. The workshops for the peer reviewers lasts two to three hours and consensus is usually reached within this time. Very occasionally a potential reviewer pulls out of the process if they do not agree with their colleague's judgements. Peer reviewers are not expected to agree exactly since each reviewer sees the teaching activity through their own lens. Reviewers do come to understand that they are not applying a personal judgement about whether this is an appropriate way to teach. Peer reviewers do not give formative feedback to the reviewee as this would undermine the purpose of summative peer review for decision making and begin to mix the formative and summative processes. Peer reviewers do not make any judgement about whether a reviewee should receive a teaching award or be promoted. The reviewer is providing independent evidence that they observed a teacher do particular things and it appeared to be effective or not from the students' perspective.

Approved peer reviewers were expected to complete at least two peer reviews a semester and a minimum of two reviews annually and attend an update session once every two years. At RMIT up to 120 peer reviewers were active in the system and around 170 peer reviews annually were conducted when the system was fully operational. There was also a process for peer reviewers to be removed from the register if their reviews continually differed from their peers over a period of time.

An important part of the independence of the process was that peer reviewers could not review a colleague from their own school. This meant that reviewers were not content

experts and the strength of the process rests on both this independence and the fact that reviewers are not "biased" by how they think teaching should be conducted in a particular discipline. The reviewee liaises with the two reviewers to determine which session will be observed – the reviewee has complete choice over the session to be reviewed. Only one session has to be observed unless there is an unforeseen disruption to the teaching session in which case a new session is reviewed. If the two reviewers differ markedly in their reports then the central administering group seeks a third reviewer who independently reviews another teaching session of the reviewee. This happens each year with an average of three review sessions having to go to a third reviewer. In these cases, all the peer review reports are submitted to the relevant decision-making panel.

A slightly revised set of documentation was introduced at the University of New South Wales (UNSW). The major changes involved reducing the number of dimensions of teaching from nine to eight and reducing the reviewer selection boxes from four down to three (UNSW, 2017). These changes were made on the basis of feedback from academics at UNSW and observations on the use of the four selection boxes at RMIT. The selection, training and reporting process retained the same features as introduced at RMIT. At UNSW there are now over 70 trained peer reviewers and the process is being introduced over a two-year period.

The definition of what constitutes effective teaching in the context of the review session has been discussed widely as part of the implementation process at UNSW. For the purposes of the summative peer review, effective teaching means that students are actively engaged in a process that enhances their learning during the session being observed.

We have found that a mandatory pre-observation meeting between reviewee and the two reviewers is required so that the reviewee can briefly outline the types of students who will be at the session, the context for the session and whether any of the dimensions of teaching will not be used for the particular session to be observed. At UNSW we have stated that a minimum of six of the eight dimensions must be observed. The main dimension not used by some reviewees is that related to actively using links between research, industry or professional practice and teaching. There is no implied hierarchy in the order of the dimensions and we have found that reviewees will usually demonstrate a preference for some dimensions over others in their teaching. There is the option to have a post review meeting if there has been some unexpected disruption during the session reviewed and the reviewers and reviewee can discuss whether this was serious enough to warrant a second opportunity for the reviewee to be reviewed. No formative feedback is given although the reviewee receives the copies of the review reports. We do not allow reviewees to request a second review session on the grounds that they could have done a better performance; only unforeseen disruptions trigger a second review.

The summative review process under the conditions described in this paper will provide reports that are different to those generated under a formative process. The reviewers in the summative process are independent of the outcomes sought by the reviewee, whereas in the formative process the reviewers have been sought out by the reviewee and form a trusted relationship within which to provide suggestions for improvement in the reviewee's teaching. Some universities have combined the two processes so that the same protocols, reviewers and documentation are used for both formative and summative peer review. One reason for this approach is efficiency, since the same reports can be used multiple times, and this reduces the workload on both reviewers and reviewees. However, we thought that a single process with two different purposes could lead to confusion for all stakeholders, including the decision-making panels. Academic promotion panels have been concerned with the use of peer review reports because they are often conducted under voluntary conditions where reviewees are able to choose their own reviewers and where the reviewer is making a personal judgement about how the reviewee could improve their teaching (Thomas, *et al*,2014). The process described in this paper makes it very clear the purpose of the peer review and the conditions under which the reports are generated. The promotion panel can have confidence that the reviewer is an independent observer and is not making subjective judgements or has a personal interest in the success of the reviewee in their application.

We have found that having two peer review reports, one from a learning and teaching expert and one from a broad discipline expert is important to ensure no inherent bias is introduced into the process. There is still concern from some academics that peer reviewers who do not have expert discipline knowledge will not be able to make a valid judgement about their teaching. Over the several hundred reviews conducted at RMIT and UNSW this has not been observed and our peer reviewers have expressed confidence in being able to judge the effectiveness of the teaching when using the dimensions specified in the template. It is true that new reviewers are sometimes apprehensive about whether they will be able to determine the effectiveness of examples observed during the session, but after one or two reviews this apprehension disappears. In feedback sessions with reviewers they have indicated that the training session using the videos is a crucial component of the process as it allows them to align their approach to the peer review with the observation of evidence against the dimensions. We have found the alignment of reviewer and reviewee in terms of their own teaching methodology is more important than the alignment of discipline area. So, if we have academics teaching predominantly online, we assign reviewers who have experience in this mode of delivery. Likewise, we attempt to match reviewers who are familiar with team teaching or studio teaching where this is the format of the session to be observed, although the availability of specific reviewers can limit this approach.

We have an annual workshop and debrief session for reviewers and reviewees so that they can provide advice on any improvements to the process and discuss how the training might be more effective. Peer reviewers have routinely described the act of observing other academics teach as a form of professional development for themselves and that taking part in this process has improved their own teaching. Being a peer reviewer is a form of professional development in its own right as the peer reviewer is engaging with the dimensions of teaching and observing how effective particular approaches to teaching are in enacting these dimensions. The peer reviewers have commented that they have adapted some of the approaches of the reviewee to their own teaching. So, although we have stated that peer review was for decision making purposes and not for professional development, a consequential outcome of the process is an improvement in teaching practice.

An extension to the summative peer review process at UNSW has been the development of a template for the summative peer review of online teaching. This is still in its early stages and will be trialed in the coming semester. We have not yet revised the original documentation from the OLT project for the summative peer review of curriculum documentation (OLT, 2006). Many promotion applications include evidence of impact at the curriculum level in addition to quality classroom delivery practices. We are working further on adapting the OLT project documentation on evidencing quality curriculum design and assessment tasks to further complement our use of peer review of classroom practices.

Conclusions

Universities are required to demonstrate that they have a quality assurance process in place and the criteria and evidence used for the academic promotion process is a key part of this activity. Research metrics have been relatively stable over many years even if refinements are applied in different countries. Expert peer reviewed scholarly output in highly ranked journals, citations and peer reviewed external competitive grants are the main currencies used to measure the quality of research.

Common metrics for describing the quality of teaching in promotion applications have been less universally accepted, except for the use of student feedback. Designing and implementing a more structured process for the collection of independent evidence of the effectiveness of teaching provides one step in the process of creating a more generally acceptable measure of teaching quality. This paper has described only one part of the process, that of a teaching session that can be observed by others. There is still a need to fully test the documentation and processes for the peer review of online programs and teaching and the peer review of curriculum design and assessment.

References

Alderman, L., Towers, S. & Bannah, S. (2012). Student feedback systems in higher education: a focused literature review and environmental scan. *Quality in Higher Education*, 18(3), 261-280.

Bell, M. (2012). *Peer Observation Partnerships in Higher Education*. 2 Edn, Higher Educational Research and Development Society of Australasia, Milperra.

Chester, A., Clarke, A., Wingrove, D. & Denny, B. (2013). *Talking about teaching: empowering staff through peer partnerships*, in Andrea Nolan (ed.) Proceedings of the 10th Enhancement and Innovation in Higher Education conference, Glasgow, United Kingdom, 11-13 June 2013, pp. 568-575.

Iqbal, I. (2013). Academics' resistance to summative peer review of teaching: questionable rewards and the importance of student evaluations. *Teaching in Higher Education*, 18, 557-569.

OLT (2006). *Peer Review of Teaching for Promotion Purposes: a project to develop and implement a pilot program of Peer Review of Teaching at four Australian Universities*. Retrieved from http://www.olt.gov.au/project-develop-implement-pilot-program-unsw-2006.

RMIT (2017). *Peer Review of Teaching*. Retrieved from http://www.rmit.edu.au/teaching/peer-review

Smithson, J., Birks, M., Harrison, G., Nair, C.S, Hitchins, M. (2015). Benchmarking for the effective use of student evaluation data. *Quality Assurance in Education*, 23(1), 20-29.

Thomas, S., Chie, Q.T., Abraham, M., Raj, S..J. & Beh, L-S. (2014). A Qualitative Review of Literature on Peer Review of Teaching in Higher Education. An Application of the SWOT Framework. *Review of Educational Research*. 84(1), 112-159.

UNSW (2017). *Peer Review of Teaching*. Retrieved from https://teaching.unsw.edu.au/peerreview.

Van Vugt, M., Hogan, R., & Kaiser, R.B. (2008). Leadership, followership, and evolution: Some lessons from the past. *American Psychologist*, 63(3), 182-196.

Visser-Wijnveen, G.J., Stes, A. & Peter Van Petegem, P. (2014), Clustering teachers' motivations for teaching. *Teaching in Higher Education*, 19(6), 644-656.