Task Force Members

Jennifer Fellabaum-Toston – MU

Diane Filion – UMKC

June 18, 2018

IFC STATEMENT ON EVALUATING CLASSROOM-BASED, ONLINE, BLENDED AND LABORATORY TEACHING INTERACTIONS

INTRODUCTION

Identifying appropriate strategies to effectively evaluate college teaching has been an issue nationally and has received more attention recently at the University of Missouri. The use of Student Ratings of Teaching (

National standards for measuring recommend using data from multiple sources.

Student ratings of teaching are a necessary, but not sufficient, strategy in assessing the quality of teaching.

CRITICAL ISSUES IN EVALUATING TEACHING

WHAT THE LITERATURE TELLS US

A critical element in encouraging quality teaching is defining the expectations for effective teaching and explaining how it contributes to the institutional goals. This requires ongoing efforts by both faculty and administrators to communicate high expectations for teaching and to reward faculty who achieve that level. The expectations and goals for faculty related to teaching and learning must be clear. The essential elements when establishing the evaluation criteria include: 1) evaluations that are of optimum use in faculty development, 2) appropriate use of the evaluation results, and 3) assurance there is alignment between evaluation and development efforts and the departmental and institutional goals.

A persistent issue when determining what sources of information to gather is to ask what purpose the evaluations serving; teaching evaluations can either be “formative” or “summative.” Formative data is collected with the sole purpose of providing feedback for development whereas summative data is collected for evaluation purposes. Because of the tension created by trying to address both of these (Morehead & Shedd, 1997), it may be necessary to employ two separate evaluation systems (Cavanagh, 1996). The key is to determine the combination of sources that will be used and how each of these should be used formative, summative or both.

Students, faculty, and administrators generally agree that quality teaching: 1) establishes a positive learning environment; 2) motivates student engagement; 3) provides appropriate challenges; 4) is responsive to students’ learning needs; and 5) is fair in evaluating students’ learning (Berk, 2005). Historically, Student Ratings of Teaching have been the primary measure for teaching effectiveness (Seldin, 1999a). Research has shown that the student ratings of teaching tool is one important measure of student perception, but is not sufficient to fully assess and improve the quality of teaching (Berk, 2014). There are also concerns, especially by those institutions focusing on student learning outcomes, that the student ratings are not related to learning outcomes (Flaherty, 2016; Uttl, White & Gonzalez, 2017).

There is considerable evidence of the rating bias with student ratings and the potential for bias needs to be taken into account when both designing student ratings of teaching and analyzing their results. Studies suggest: women were rated lower than men (Basow, 1994; Koblitz, 1990; MacNeil, Driscoll, & Hunt, 2014; Mitchell & Martin, 2018; Morgan et al., 2016); faculty of color received lower ratings than Caucasian faculty (Hamermesh & Parker, 2005; Smith & Johnson-Bailey, 2012); novice faculty were rated lower than the experienced (Centra, 1978); graduate students were rated lower than ranked faculty (Brandenburg, 1977); faculty in STEM disciplines were rated lower than those in

the humanities (Kember & Leung, 2011); and medium or large section courses received lower ratings than small section courses (Feldman, 1978; Franklin et al., 1991; Miles & House, 2015). Some studies have even found that the content of a course may influence evaluation results (e.g.., quantitative courses studied by Uttl & Smibert, 2017) as well as variables such as the timing of the course (e.g., early mornings for an introductory college physics class by Tobin, 2017). It is not easy to adjust for these biases, because students draw upon multiple factors when completing evaluations. Indeed, Boring, Ottoboni, and Stark (2016) argue that student ratings are more sensitive to students' gender bias and grade expectations than they are to teaching effectiveness (Ray, 2018).

National standards for measuring teaching effectiveness recommend using data from multiple sources. An ideal approach is to create a triangulation strategy by using at least three sources of data. Triangulating the sources of information balances the strengths and weaknesses of each measure and provides a more accurate reflection of teaching effectiveness. A variety of methods used across the nation are described later in the paper.

IMPROVING STUDENT RATINGS OF TEACHING

Student ratings of teaching do allow instructors to learn from those in their classes, and is often the only way to hear directly from those in the course. We concur with the following statement from Stark and Freishtat (2014), "student ratings of teaching are valuable when they ask the right questions, report response rates and score distributions, and are balanced by a variety of other sources and methods to evaluate teaching" (p. 2). Student ratings of teaching should not ask students how much they have learned in the course because people are poor at evaluating their own learning and it is difficult for students to know what they do not know. Lastly, it is difficult for a student to judge the effectiveness of any instructional practice except by comparing it with others that they have already experienced (Wieman, 2015). When utilizing student ratings average scores should not be used, instead those reviewing this data should look at the distributions (Linse, 2017).

RATING QUESTIONS

Creating or modifying the instrument(s) used in evaluating teaching needs to begin with discussions among faculty and administration to determine what qualities are essential to being an "effective teacher" across all disciplines. These multidisciplinary considerations should be based on experience and grounded in supporting research and literature. Creating a shared definition is an essential first step in evaluating quality teaching (Gibbs, 1995). When developing or modifying student ratings of teaching

NECESSARY COMPONENTS FOR STUDENT RATINGS OF TEACHING

To obtain the best results from

National standards for measuring teaching effectiveness recommend using data from multiple sources. The best approach is to create a triangulation strategy, using three or more sources of evidence, this allows the strengths and weaknesses of each source to balance each other out (Appling, Naumann, & Berk, 2001). They can also provide a more accurate, reliable, and comprehensive picture of teaching effectiveness (Berk, 2005). When possible one should use strategies to gain feedback from students, peers, and self-evaluation to create a comprehensive evaluation.

While most departments do not employ multiple strategies, they all seem to agree that just using student ratings does not provide the information needed to evaluate teaching effectiveness or provide the information needed for promotion and tenure decisions. Weimer (2015) said it best when he summarized the problem this way:

> …feedback on end-of-course rating instruments offers a view of your teaching. It's not a 360-degree panorama, but rather something closer to the view provided by a small window. And if the instrument isn't very good, it's like looking through a dirty window. For years, most ratings experts have advised institutions and individuals to collect data from multiple sources and in different ways. We don't need just one view. We should be looking at our teaching from every window in the house. (Work for a realistic perspective on the results section, para. 5)

Teaching is a scholarly activity, and to prepare for a course requires several elements. Faculty must review the literature, select resources, create content outline, prepare a syllabus, design learning activities, integrate instructional technology, and construct evaluation measures (Webb & McEnerney, 1995). If teaching performance is to be recognized and rewarded as scholarship, teaching should be judged by the same high standards applied to other forms of scholarship: peer review.

Peer review of teaching is composed of two activities: peer observation of in-class teaching performance and peer review of the written documents used in a course. Both forms of peer review should be included in a comprehensive system, where possible. Peer ratings of teaching performance and materials is the most complementary to student ratings. It covers those aspects of teaching that students are not in a position to evaluate. However, peer ratings should not be used for personnel decisions (Braskamp & Ory, 1994). There are differing definitions of peers depending on the institution, these could include those within a department, college, school, teaching and learning specialists or other peers that the department and faculty agree upon.

RECOMMENDATIONS: ENHANCED STRATEGIES FOR EVALUATING TEACHING

In order to create a positive climate that is conducive to improving teaching effectiveness, it is imd inclrov, itpat tING(B 541rovrsRE1ye wte tihe

determining which combination of sources (three or more) should be used for both continued improvement and growth and which will be used to evaluate the achievement of baseline standards.

Whatever methods are chosen it is imperative to define the use of these methods and to appropriately design, execute, and report the results. The accuracy of faculty evaluation decisions hinges on the integrity of the process and the reliability and validity of the evidence you collect (Braskamp & Ory, 1994). Begin with the end goal of improving teaching and learning in mind and then develop the strategies that will most effectively achieve the goal (

Mitchell, K. M. W., & Martin, J. (2018, March 6). Gender bias in student evaluations. *PS: Political Science & Politics.* doi: 10/1017/S104909651800001X

Morehead, J. W., & Shedd, P. J. (1997). Utilizing summative evaluation through external peer review of teaching. *Innovative Higher Education, 22*(1), 37-44.

Morgan, H. K., Purkiss, J. A., Porter, A. C., Lypson, M. L., Santen, S. A., Christner, J. G., …Hammoud, M. M. (2016). Student evaluation of faculty physicians: Gender differences In teaching evaluations. *Journal of Women's Health, 25*(5), 453-456. doi: 10.1089/jwh.2015.5475

Penny, A. R. (2003). Changing the agenda for research into students' views about university teaching: Four shortcomings of SRT research. *Teaching in Higher Education, 8*(3), 399-411.

Perlberg, A. (1983). When professors confront themselves: Towards a theoretical conceptualization of video self-confrontation in higher education. *Higher Education, 12*(6), 633-663.

Ray, V. (2018, February 9). Is gender bias an intended feature of teaching e?.valuations, *Inside Higher Ed*. Retrieved from https://www.insidehighered.com/advice/2018/02/09/teaching-evaluations-are-often-used-confirm-worst-stereotypesereotyTT2 1 v.ng 004 Tc 0.006rnaTj -0.01 Tc 0.0.0.4 (

Webb, J., & McEnerney, K. (1995). The view

Evaluating Teaching Resource Appendix
(Underlined items are hyperlinks to resources)

Student evaluations

- Mid-semester feedback (Three sample forms below)
  -