# UNIVERSITY OF MISSOURI
# INTERCAMPUS FACULTY COUNCIL

## EVALUATING CLASSROOM-BASED, ONLINE, BLENDED AND LABORATORY TEACHING INTERACTIONS


Task Force Members

- Jennifer Fellabaum-Toston – MU
- Diane Filion – UMKC
- Andy Goodman – UMSL
- Larry Gragg – S&T
- Steve Graham – UM System
- Jon McGinnis – UMSL
- Isabel Montes Gutierrez – UM System
- Carrie Nicholson – UM System
- Candace Schlein - UMKC


June 18, 2018

# TASK FORCE MEMBERS

**Jennifer Fellabaum-Toston**

Assistant Teaching Professor, ELPA

Associate Director of the Statewide Cooperative EdD Program

Program Coordinator, Higher Education

University of Missouri - Columbia

**Diane Filion**

Interim Vice Provost for Faculty Affairs
Professor of Psychology

University of Missouri - Kansas City

**Andy Goodman**

Associate Professor

Director of the Center for Teaching and Learning

Associate Provost for Professional Development

University of Missouri - St. Louis

**Larry Gragg**

Curators' Distinguished Teaching Professor Emeritus

Chair, Center for Advancing Faculty Excellence

Missouri University of Science & Technology

**Steve Graham**

Senior Associate Vice President for Academic Affairs

University of Missouri System

**Jon McGinnis**

Professor and Chair of Philosophy

University of Missouri – St. Louis

**Isabel Montes Gutierrez**

Graduate Assistant

University of Missouri System

**Carrie Nicholson**

Research Consultant II

University of Missouri System

**Candace Schlein**

Associate Professor

Interim Program Director for Graduate Student Development

Interim Program Director for the Faculty Center for Excellence in Teaching (FaCET)

University of Missouri – Kansas City

# 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 (often referred to as student evaluations of teaching but research shows a more accurate terminology is student ratings), one of the most common strategies for evaluating teaching, has come under increasing scrutiny as numerous studies have suggested there is potential rating bias (e.g., gender, ethnicity, course structure, and experience of instructor). Furthermore, faculty complain of low response rates and students express concerns that the feedback is not utilized (Boring, Ottoboni, & Stark, 2016; Davis, Hirschberg, Lye & Johnston, 2007; Flaherty, 2018; Linse, 2017; Mitchell, & Martin, 2018; York University, 2002).

After several discussions, the University of Missouri Intercampus Faculty Council (IFC) created a task force to examine strategies to evaluate teaching effectiveness with the primary goal of improving teaching effectiveness and student learning. Although there are many components that may be considered under the category of teaching in tenure and promotion in this paper the reference to teaching is referring to classroom, online, blended and laboratory instruction. The task force is comprised of at least one faculty member from each of the UM campuses and is staffed by the UM Office of Academic Affairs.

The committee members sought input from IFC members as well as provosts and vice provosts in outlining the charge for the committee. From these discussions and from the work of previous IFC task force on faculty workload and post-tenure review the following issues emerged.

- With the rising cost of college, growing student debt, and a heightened focus on preparing students for employment, there is an increased focus on “student success” in college. Effective teaching is an important component of student success.
- With the use of online student ratings, the response rates have been dropping to the point that neither the faculty nor the students have confidence in the results.
- Despite the issues with student ratings of teaching, they are often used for promotion and tenure decisions – sometimes as the primary source of teaching evaluation.
- At research-focused universities, department chairs and faculty feel they have limited time to launch extensive teaching evaluation efforts.
- Only a few academic units use a comprehensive multiple measure system to evaluate teaching effectiveness.

- 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 instruments institutions should be able to: 1) select the aspects that are most important, according to their educational vision and policy, thereby developing student rating instruments that are consistent with their own preferences; and 2) all stakeholders (i.e., administrators, instructors, and students) should be involved in the definition of these characteristics (Penny, 2003; Spooren, Brockx, & Mortelmans, 2013). Utilizing a core set of questions allows for comparisons across disciplines and provides an avenue for cross-campus comparisons (e.g., 4 Likert-type and 2 open-ended questions), but additional customized questions should be added to best fit the campus, department, and/or course.

## NECESSARY COMPONENTS FOR STUDENT RATINGS OF TEACHING

To obtain the best results from student ratings of teaching data the evaluation instruments should include four different components: 1) questions about the course, 2) questions about the instructor, 3) questions about how the student prepared for the course, and perhaps most important, 4) students should be strongly encouraged to leave comments on the course (Goodman, 2015). In order to elicit the most useful information for improving teaching effectiveness, all of these components should have a clear distinction from one another so that students, faculty, and administrators know which element they are evaluating. See Lindahl and Unger (2010) for proposed changes that would begin to address what Fish (2005) referred to as "casual cruelty" in student comments on teaching evaluations.

## IMPROVING RESPONSE RATES

With an increasing number of survey requests going out to students today, those that arrive last might find the students prone to survey fatigue, consequently contributing to lower response rates (Groves, Presser, & Dipko, 2004). Many campus offices and departments administer surveys to students throughout the semester, but student ratings of teaching are often the last ones students receive. How students receive the evaluations is also an important consideration as there is variation in online and paper-based response rates (e.g., Stanny & Arruda, 2017). Overall response rates should be 50% in larger courses and 75-80% in smaller courses (e.g., graduate) or higher for the course. Lower response rates may lead to additional concerns of reliability (e.g., Kalender, 2015), including nonresponse bias (Bacon, Johnson, & Stewart, 2016).

A variety of strategies are available to increase response rates including:

- incentives (points, note card for exam, or treats) - this particular strategy is sometimes controversial and should be left to the discretion of the faculty member,
- dedicated time in class for completion,
- encouragement by the faculty (e.g., if faculty members show genuine interest in student ratings, students will be more motivated to participate)
- reminders with clear instructions concerning participation (via email, LMS, or in class), and
- informing the students how the evaluations will be used to improve instruction (highlighting their important role as raters) (Spooren et al., 2013).

Students will gain additional motivation to complete the questionnaires once they see their comments are taken seriously and incorporated into the lecturers' future material. Even when you have to explain what you are not changing and why, this opens communication lines and allows them to see that their feedback is valued.

## MULTIPLE MEASURES APPROACH

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 imperative that faculty understand the purpose of teaching evaluations. Additionally, they need access to professional developmental tools that will allow them to continue their professional growth and improvement as teachers.

Student ratings of teaching alone do not lead to better teaching. For this reason student ratings should be embedded within a more holistic approach to the evaluation of teaching. This involves

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 (Research Corporation for Science Advancement, 2015).

Evaluating teaching to promote a climate that fosters and rewards better teaching requires long-term commitment and shaping the culture in the department and college. Since most departments have not adopted a comprehensive approach using multiple measures, this paper includes a number of different alternatives that can be deployed. Below is a menu of strategies starting with the simpler options ranging to the more complex. The appendix to this document includes resources to aid in the implementation of these strategies.

- Course development or improvement
    - Attend a training or workshop to help with course design or improvement with the focus on engaging students in significant learning experiences (Goodman, n.d.).
    - Talk with an instructional designer to determine other ways to improve the course design.
    - Sit in on other classes with similar subject matter to learn from each other.

- Self-reflection and self-improvement
    - Attend a training or workshop around instructor personal improvement (e.g., digital fluency workshop at UMSL; ET@MO workshops at MU).
    - Self-evaluation - 82% of four-year colleges and universities reported using self-evaluations to measure teaching performance (Carnegie Foundation for the Advancement of Teaching, 1994). Tools such as a structured form and guiding questions are suggested (Seldin, 1999b); these forms can be used on a per course basis each semester to provide a structured reflection. These results could be incorporated into the Teaching Dossier, described below.
    - Video for self-reflection - instructors should be encouraged to systemically evaluate the behaviors observed using a rating scale or checklist (Brinko, 1993; Perlberg, 1983; Seldin, 1998). This would be even more helpful if an instructor can receive feedback from a peer, mentor, or consultant (Braskamp & Ory, 1994). If faculty are really committed to improving their teaching, a video is one of the best sources of

evidence for formative decisions, interpreted either alone or, preferably, with peer input (Berk, 2005).

- Malouff, Reid, Wilkes, and Emmerton (2015) outline a 14-step process for improving teaching through goal setting (step 1), self-evaluation of the course (step 2), reflection on the students' evaluations (steps 3-10), peer review (step 11), and developing an action plan (steps 12-14).

- Peer review of teaching materials requires a different type of scale to rate the quality of the course syllabus, instructional plans, texts, reading assignments, handouts, homework, and tests/projects (Braskamp & Ory, 1994).
- Student Classroom Assessment Techniques (CATs) – Formative classroom assessment can help us identify the effects of our teaching on learning. This is a timely way to help instructors identify gaps between what they teach and what students learn and enable them to adjust their teaching to make learning more efficient and effective. A few examples of these assessments are: 1) one-minute papers, 2) one-sentence summaries, 3) critical incident questionnaires, 4) focus groups, and 5) mid-year mini surveys. Use of CATs promotes reflective practice. It is important to balance the positive and negative comments and try to link negative commentary to issues of student learning. New users of classroom assessment techniques might find it helpful to discuss the critical comments with an experienced colleague (York University, 2002). See Angelo and Cross (1993) for a list of 50 CATs that instructors may find useful.

- Peer observation of teaching - requires a rating scale covering instructor's content knowledge, delivery, teaching methods, and learning activities (Berk, 2009; Berk, Naumann, & Appling, 2004).

    - To create the best outcomes the instructor and observer should meet prior to the class to discuss the objectives and strategies of the class, materials to be used, and to clarify expectations of the observation. Then, a post-observation meeting allows an opportunity for constructive feedback and assistance in the development of a plan for improvement.

    - One of the most valuable forms of observation is peer-pairing where two instructors provide each other with feedback on their teaching on a rotating basis, each evaluating the other for a period of time. Each learns from the other and may learn as much in the observing role as when being observed (York University, 2002).

- Student interviews

    - Quality control circles - The instructional version of the "circle" involves assembling a group of volunteer students to meet regularly (e.g., bi-weekly) to discuss teaching strategies, identify any areas of concern, and find ways to continuously improve. The

members of the circle can also report the results of the meeting to create open communication. This method can be extremely effective for making changes in instruction. However, faculty must be open to student comments and be willing to work on continuous improvement (Berk, 2005).

- Classroom group interviews - Classroom group interviews involves the entire class, but is conducted by someone other than the instructor, usually a colleague in the same department, a graduate TA, or a faculty development or student services professional. The interviewer uses a structured questionnaire to probe the strengths and weaknesses of the course and teaching activities (Braskamp & Ory, 1994).

- 360 degree assessment - Five of the best sources you could use are teacher mentor, peer ratings and video with self/peer, student ratings, student interviews, and self-ratings. By having all five of these sources one can fairly accurately determine their strengths and opportunities for improvement. Data from the different sources may be collected at different times during the semester (Berk, 2009).

- Teaching dossier - A teaching dossier can provide the framework for a systematic program of reflective analysis and peer collaboration leading to improvement of teaching and student learning. This gives the instructor a platform to describe their teaching philosophy, review their teaching goals and objectives, assess the effectiveness of their classroom practice and the strategies they use to animate their pedagogical values, and identify areas of strength and opportunities for improvement. A dossier should present a selection of information organized in a way that gives a comprehensive and accurate summary of teaching activities and effectiveness (York University, 2002).

## CLOSING THOUGHTS

The faculty must perceive that effective teaching is valued as much as other areas (i.e., research) and that it is rewarded (Morehead & Shedd, 1997). Understanding the roles and expectations of instructors in regard to teaching and learning is a central issue in establishing evaluation criteria (Johnson & Ryan, 2000). Defining expectations of faculty in relation to institutional and unit goals will require ongoing efforts by faculty and administrators to clearly identify and mutually develop expectations. An important element is that the effects can be different depending on the nature of the faculty involved. For example, a large ratio of NTT faculty with 80% teaching can be impacted differently than the more traditional T/TT faculty where teaching may only be 40% of their work.

Institutions should expect faculty members within a scholarly unit to collaborate in the creation of a mentoring community committed to the continuous improvement of teaching and responsible student learning (Cross, 1991). Within such a mentoring community, faculty members, individually and collectively, experience a mandate for collegial conversation, a peer discussion of the appropriateness of various evaluation criteria for specific kinds of evaluations within their particular discipline.

## REFERENCES


Angelo, T. A., & Cross, K. P. (1993). *Classroom assessment techniques: A handbook for college teachers (2nd ed.).* San Francisco, CA: Jossey-Bass.

Appling, S. E., Naumann, P. L., & Berk, R. A. (2001). Using a faculty evaluation triad to achieve evidence-based teaching. *Nursing Education Perspectives, 22*(5), 247.

Bacon, D. R., Johnson, C. J., & Stewart, K. A. (2016). Nonresponse bias in student evaluations of teaching. *Marketing Education Review, 26*(2), 93-104. doi: 10.1080/10528008.2016.1166442

Basow, S. A. (1994, June). *Student ratings of professors are not gender blind*. Paper presented at the meeting of the Society of Teaching and Learning in Higher Education, Vancouver, British Columbia.

Berk, R. A. (2005). Survey of 12 strategies to measure teaching effectiveness. *International Journal of Teaching and Learning in Higher Education, 17*(1), 48-62.

Berk, R. A. (2009). Using the 360 multisource feedback model to evaluate teaching and professionalism. *Medical Teacher, 31*(12), 1073-1080.

Berk, R. A., Naumann, P. L., & Appling, S. E. (2004). Beyond student ratings: Peer observation of classroom and clinical teaching. *International Journal of Nursing Education Scholarship, 1*(1), 1-26.

Berk, R.A. (2014). Should student outcomes be used to evaluate teaching? *Journal of Faculty Development 28(2), 87-96*

Boring, A., Ottoboni, K., & Stark, P. B. (2016). Student evaluations of teaching (mostly) do not measure teaching effectiveness. *ScienceOpen Research, 10*, 1-11.

Brandenburg, D. C., Slinde, J. A., & Batista, E. E. (1977). Student ratings of instruction: Validity and normative interpretations. *Research in Higher Education 7*(1) 67-78.

Braskamp, L. A., & Ory, J. C. (1994). *Assessing faculty work: Enhancing individual and institutional performance.* San Francisco, CA: Jossey-Bass.

Brinko, K. T. (1993). The practice of giving feedback to improve teaching: What is effective? *The Journal of Higher Education, 64*(5), 574-593.

Cavanagh, R. R. (1996). Formative and summative evaluation in the faculty peer review of teaching. *Innovative Higher Education, 20*(4), 235-240.

Centra, J. A. (1978). Using student assessments to improve performance and vitality. In W. R. Kirschling (Ed.), *Evaluating faculty performance and vitalilty* (pp. 31-49). San Francisco, CA: Jossey-Bass.

Cross, K. P. (1991). College teaching: What do we know about it? *Innovative Higher Education, 16*(1), 7-25.

Davies, M., Hirschberg, J., Lye, J., & Johnston, C.G. (2007). Systematic influences on teaching evaluations: The case for caution. *Australian Economic Papers*, 46(1), 18-38.

Feldman, K. A. (1978). Course characteristics and college students' ratings of their teachers: What we know and what we don't. *Research in Higher Education, 9*(3) 199-242.

Fish, S. (2005, February 4). Who's in charge here? *Chronicle of Higher Education*. Retrieved from https://www.chronicle.com/article/Whos-In-Charge-Here-/45097

Flaherty, C. (2016, September 21). Zero correlation between evaluations and learning, *Inside Higher Ed*. Retrieved from https://www.insidehighered.com/news/2016/09/21/new-study-could-be-another-nail-coffin-validity-student-evaluations-teaching

Flaherty, C. (2018, March 14). Same course, Different ratings. *Inside Higher Ed*. Retrieved from https://www.insidehighered.com/news/2018/03/14/study-says-students-rate-men-more-highly-women-even-when-theyre-teaching-identical

Franklin, J., Theall, M., & Ludlow, L. (1991, April). *Trade inflation and student ratings: A closer look.* Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

Gibbs, G. (1995). *Assessing student-centered courses*. Oxford, UK: Oxford Bookes University.

Goodman, A. (2015). *Increasing response rates with online course evals*. Unpublished report.

Goodman, A. (n.d). *Ten steps to success*. Retrieved from https://www.umsl.edu/services/ctl/faculty/tensteps.html

Groves, R. M., Presser, S., & Dipko, S. (2004). The role of topic interest in survey participation decisions. *Public Opinion Quarterly, 68*(1), 2-31.

Hamermesh, D. S., & Parker, A. (2005). Beauty in the classroom: Instructors' pulchritude and putative pedagogical productivity. *Economics of Education Review, 24*, 369-376. doi: 10.1016/j.econedurev.2004.07.013

Johnson, T. D., & Ryan, K. E. (2000). A comprehensive approach to the evaluation of college teaching. *New Directions for Teaching and Learning, 2000*(83), 109-123.

Kalender, I. (2015). Reliability-related issues in the context of student evaluations of teaching in higher education. *International Journal of Higher Education, 4*(3), 44-56. doi: 10.5430/ijhe.v4n3p44

Kember, D., & Leung, D.Y.P. (2011). Disciplinary differences in student ratings of teaching quality. *Research in Higher Education, 52*(3), 278-299.

Koblitz, N. (1990). Are student ratings unfair to women? *Newsletter of the Association for Women in Mathematics, 20*(5), 17-19.

Lindahl, M. W., & Unger, M. L. (2010). Cruelty in student teaching evaluations. *College Teaching, 58*, 71-76. doi: 10.1080/87567550903253643

Linse, A. (2017) . Interpreting and using student ratings data: Guidance for faculty serving as administrators and on evaluation committees. *Studies in Educational Evaluation, 54,*94-106. doi: 10.1016/j.stueduc.2016.12.004

MacNeil, L., Driscoll, A., & Hunt, A. N. (2014). What's in a name: Exposing gender bias in student ratings of teaching. *Innovative Higher Education, 40*(4), 291-303. doi: 10.1007/s10755-014-9313-4

Malouff, J. M., Reid, J., Wilkes, J., & Emmerton, A. J. (2015). Using the results of teaching evaluations to improve teaching: A case study of a new systematic process. *College Teaching, 63*(1), 3-7. doi: 10.1080/87567555.2014.956681

Miles, P., & House, D. (2015). The tail wagging the dog: An overdue examination of student teaching evaluations. *International Journal of Higher Education, 4*(2), 16-126. doi: 10.5430/ijhe.v4n2p116

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-stereotypes-about-women-faculty

Research Corporation for Science Advancement. (2015). Searching for better approaches: Effective evaluation of teaching and learning in STEM. Retrieved from http://rescorp.org/gdresources/publications/effectivebook.pdf

Seldin, P. (1998). *The teaching portfolio*. Paper presented at the American Council on Education, Department Chairs Seminar, San Diego, CA.

Seldin, P. (1999a). Current practices–good and bad–nationally. In P. Eldin & Associates (Eds.), *Changing practices in evaluating teaching: A practical guide to improved faculty performance and promotion/tenure decisions* (pp. 1-24). Bolton, MA: Anker.

Seldin, P. (1999b). Self-evaluation: What works? what doesn´t? . In P. Eldin & Associates (Eds.), *Changing practices in evaluating teaching: A practical guide to improved faculty performance and promotion/tenure decisions* (pp. 97-115). Bolton, MA: Anker.

Smith, B., & Johnson-Bailey, J. (2012). Student ratings of teaching effectiveness: Implications for non-white women in the academy. *Negro Educational Review, 62/63*(1-4), 115-140.

Spooren, P., Brockx, B., & Mortelmans, D. (2013). On the validity of student evaluation of teaching: The state of the art. *Review of Educational Research, 83*(4), 598-642.

Stanny, C. J., & Arruda, J. E. (2017). A comparison of student evaluations of teaching with online and paper-based administration. *Scholarship of Teaching and Learning in Psychology, 3*(3), 198-207. doi: 10.1037/stl0000087

Stark, P. B., & Freishtat, R. (2014). An evaluation of course evaluations. *ScienceOpen.* doi: 10.14293/S2199-1006.1.SOR-EDU.AOFRQA.v1

Carnegie Foundation for the Advancement of Teaching. (1994). *National survey on the reexamination of faculty roles and rewards.* Princeton, NJ: Author.

Tobin, R. G. (2017). Too early for physics? Effect of class meeting time on student evaluations of teaching introductory physics. *The Physics Teacher, 55*, 276-279. doi: 10.1119/1.4981033

Uttl, B., White, C.A., & Gonzalez, D.W. (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. doi: 10.1016/j.stueduc.2016.08.007

Webb, J., & McEnerney, K. (1995). The view from the back of the classroom: A faculty-based peer observation program. *Journal on Excellence in College Teaching, 6*(3), 145-160.

Weimer, M. (2015, September 16). *Student rating reminders and suggestions*. Retrieved from https://www.facultyfocus.com/articles/teaching-professor-blog/student-ratings-reminders-and-suggestions/

Wieman, C. (2015). A better way to evaluate undergraduate teaching. *Change: The Magazine of Higher Learning, 47*(1), 6-15.

York University. (2002). *Senate committee on teaching and learning´s guide to teaching assessment & evaluation*. Retrieved from http://secretariat.info.yorku.ca/files/tevguide.pdf

**Other references reviewed not cited above**

Adams, M. J. D. (2010, November). *Who doesn't respond and why? An analysis of nonresponse to online student evaluation of teaching*. Paper presented at the Annual Conference of the Association for the Study of Higher Education Indianapolis, IN.

Berk, R. A. (2014). Should student outcomes be used to evaluate teaching? *The Journal of Faculty Development, 28*(2), 87-96.

Buskist, C., & Hogan, J. (2010). She needs a haircut and a new pair of shoes: Handling those pesky course evaluations. *Journal of Effective Teaching, 10*(1), 51-56.

Carrell, S. E., & West, J. E. (2010). Does professor quality matter? Evidence from random assignment of students to professors. *Journal of Political Economy, 118*(3), 409-432.

Nasser, F., & Fresko, B. (2002). Faculty views of student evaluation of college teaching. *Assessment & Evaluation in Higher Education, 27*(2), 187-198.

Onwuegbuzie, A. J., Witcher, A. E., Collins, K. M., Filer, J. D., Wiedmaier, C. D., & Moore, C. W. (2007). Students' perceptions of characteristics of effective college teachers: A validity study of a teaching evaluation form using a mixed-methods analysis. *American Educational Research Journal, 44*(1), 113-160.

Skowronek, J. S., Friesen, B. K., & Masonjones, H. (2011). Developing a statistically valid AND practically useful student evaluation instrument. *International Journal for the Scholarship of Teaching and Learning, 5*(1), Article 11. doi: 10.20429/ijsotl.2011.050111

Smith, M. K., Jones, F. H., Gilbert, S. L., & Wieman, C. E. (2013). The classroom observation protocol for undergraduate STEM (COPUS): A new instrument to characterize university STEM classroom practices. *CBE-Life Sciences Education, 12*(4), 618-627.

Spooren, P., Mortelmans, D., & Thijssen, P. (2012). 'Content'versus 'style': acquiescence in student evaluation of teaching? *British Educational Research Journal, 38*(1), 3-21.

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

- Student evaluations
  - Mid-semester feedback (Three sample forms below)
    - Form A
    - Form B
    - Form C
- Course development or improvement
  - Course Design Institute is held at UMSL, open to all campuses
  - Course improvement flowchart – page 16 within the document
  - ET@MO is at MU and has instructional designers and some funding opportunities for redesign
- Self-reflection and Self-improvement
  - Postsecondary Instructional Practices Survey
  - Teaching Practices Inventory
  - Video self-analysis
    - Self-guided module for analyzing videos and rubric
    - How video self-analysis helped others
- Peer review of teaching materials
  - Syllabus rubric
  - Peer syllabus review form
- Classroom Assessment
- Peer observation of teaching
  - Classroom observation protocol
  - Observation form examples
    - Course instruction review form
    - Peer observation of teaching record
- Peer review
  - Guide for peer evaluators
  - CAFNR's peer review of teaching
- Student interviews
  - Quality control circles
  - Classroom group interviews
- 360 Degree Assessment
- Teaching Dossier
- Analyzing course evaluations
  - Making sense of course evaluations
  - Written comments
  - Qualitative example
- Campus contacts
  - MU Teaching for Learning Center
  - UMKC Faculty Center for Excellence in Teaching
  - S&T Center for Advancing Faculty Excellence
  - UMSL Center for Teaching and Learning