

**65th Session of the North Carolina State University
Faculty Senate
Academic Policy Committee (APC)
Meeting Minutes
January 15, 2019**

Present: RaJade Berry-James (co-chair), Richard Kotek (co-chair), Natalie Cooke, Beth Fath, Jeremiah Feducia, Kerry Havner, Helmut Hergeth, Robert Hayes, Doug Pearce

Absent: Deniz Eseryel, Min Liu

Summary of Discussion: We reviewed the proposed revisions to REG 05.20.10 Evaluation of Teaching and recommend the following changes to the proposed policy (see changes in **italics**):

I. DEFINITIONS

Revise 2.3: “For reviews of teaching, the peer reviewers may be colleagues of ~~any~~ **equal or higher rank** mutually agreed upon by the faculty member and the department head or academic administrator, pending resolution by the Dean if there is disagreement.”

II. SCOPE

Revise 3.2 “Faculty or a departmental designate must explain to students how ~~course~~ **university evaluation instruments** will be administered for their class.”

Revise 3.3 “Peer review of teaching is to be conducted for all ~~faculty~~ **people** with *teaching assignments* with 0.75 or greater FTE. Peer review should follow the schedule below.”

Revise 3.3.1 “Assistant Professors in both tenure- and ~~non-tenure~~ **professional** track positions should have a minimum of three peer reviews included in their dossiers for consideration of promotion to Associate Professor or consideration of the award of permanent tenure.”

Revise 3.3.4 “Peer review of **professional** ~~non-tenure~~ track faculty with 0.75 FTE or greater must be completed annual for the first three years of employment in a **professional** ~~non-tenure~~ track faculty appointment and a minimum of once every five years afterward.”

Revise 3.5 Title to read: “Other exemptions must be approved by the **Executive Vice Chancellor and Provost**”

III. INSTRUMENTS FOR EVALUATION

Revise 4.1.1.3 Title to read: "... approval of the *Executive Vice Chancellor and Provost*"

Add 4.1.2 Should this section also include specific core questions related to "safety hazards" in lab/field trip as well as a set of questions assessing/evaluating study abroad experiences or other non standard educational learning settings?

Revise 4.1.2.4 Title to read: "... approval of the *Executive Vice Chancellor and Provost*"

Revise 4.2 Peer Review Instrument "Documenting peer reviews of teaching effectiveness is a departmental responsibility. The use of the Peer Review Summary Template is strongly encouraged. If a department should choose to develop its own instrument, it ~~must~~ *should* contain the general requirements specified in 4.2.2."

Revise 4.2.2 "The peer review instrument(s) should include categories similar to those identified in the Peer Review Summary Template (4.2): Teaching methods, Teaching effectiveness, Student instructor interaction, Teaching materials, Areas of strength, Opportunities for improvement *and Overall evaluation*."

Revise 4.2.4 "*The peer review should be shared with the instructor and* the instructor should have an opportunity to respond to the peer reviewer's assessment of teaching."

IV. PEER REVIEWS

Lingering Questions: Section 5.1 Guidelines for Conducting Peer Reviews of Teaching is unclear >> Committee members had several questions regarding the proposed revisions in 5.1.1 - 5.1.5. Was the proposed revision merely a suggested guideline or an official mandate?

Recommendation: *Given the substantial change from the mandated "observation procedures" to the "proposed guidelines for peer review", committee members strongly recommend that the proposed revisions to REG05.20.10 Evaluation of Teaching be shared with the full Faculty Senate to ensure questions about faculty burden, compliance effort and process to improve teaching can be adequately discussed across disciplines and colleges.*

V. STUDENT COURSE EVALUATIONS

At NC State, student evaluations of teaching (SET) are currently used to assess self-improvement and make personnel decisions. In addition to substantiated concerns of bias, numerical averages of labels are erroneously used as an objective measure of teaching effectiveness in retention, promotion and tenure decisions. There is growing

evidence that the practice of using SET in the tenure and promotion process is problematic largely due to research documenting age, race/ethnicity and gender bias, nonacademic factors, and statistical complications involving response rates, nonresponse bias and small sample size. As such, members of the academic policy committee recommend that the validity of student course evaluations used for personnel decision must be seriously addressed before the proposed revision of the REG 05.20.10 Evaluation of Teaching is approved.

Below are peer-reviewed studies that affirm bias found in SETs and confirm persistent problems when using SET averages as objective measures for personnel decisions.

Beecham, R. (2009). Teaching quality and student satisfaction: Nexus or simulacrum? *London Review of Education*, 7, 135–146.10.1080/14748460902990336

Abstract: *Quality has clear meanings when associated with specific phenomena. The relative nature of quality in commercial contexts, specifically in relation to manufactured goods, has been studied. In service industries there is no satisfactory indicator of quality. The Likert-scale questionnaire does not allow the researcher to distinguish between spontaneous and constructed responses. The penetration of the premises of human capital theory and services marketing theory in the higher education sector has introduced inapplicable assumptions about and instruments of quality control. A study of one Australian university's re-definition of itself as a market-oriented institution suggests that long-standing criticisms of the use of student questionnaires as a measure of teaching quality are well-founded.*

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

Abstract: *Twelve potential sources of evidence to measure teaching effectiveness are critically reviewed: (a) student ratings, (b) peer ratings, (c) self-evaluation, (d) videos, (e) student interviews, (f) alumni ratings, (g) employer ratings, (h) administrator ratings, (i) teaching scholarship, (j) teaching awards, (k) learning outcome measures, and (l) teaching portfolios. National standards are presented to guide the definition and measurement of effective teaching. A unified conceptualization of teaching effectiveness is proposed to use multiple sources of evidence, such as student ratings, peer ratings, and self-evaluation, to provide an accurate and reliable base for formative and summative decisions. Multiple sources build on the strengths of all sources, while compensating for the weaknesses in any single source. This triangulation of sources is recommended in view of the complexity of measuring the act of teaching and the variety of direct and indirect sources and tools used to produce the evidence.*

Boring, A., Ottoboni, K., & Stark, P. B. (2016). Student evaluations of teaching (mostly) do not measure teaching effectiveness. Retrieved from Science Open Research. doi:10.14293/S2199-1006.1.SOR-EDU.AETBZC.v1

Abstract: *Student evaluations of teaching (SET) are widely used in academic personnel decisions as a measure of teaching effectiveness. We show: . SET are biased against female*

instructors by an amount that is large and statistically significant. . The bias affects how students rate even putatively objective aspects of teaching, such as how promptly assignments are graded. . The bias varies by discipline and by student gender, among other things. . It is not possible to adjust for the bias, because it depends on so many factors. . SET are more sensitive to students' gender bias and grade expectations than they are to teaching effectiveness. . Gender biases can be large enough to cause more effective instructors to get lower SET than less effective instructors. These findings are based on nonparametric statistical tests applied to two datasets: 23,001 SET of 379 instructors by 4,423 students in six mandatory first-year courses in a five year natural experiment at a French university, and 43 SET for four sections of an online course in a randomized, controlled, blind experiment at a US university.

Braga, M., Paccagnella, M., & Pellizzari, M. (2014). Evaluating students' evaluations of professors. *Economics of Education Review*, 41, 71–88. [10.1016/j.econedurev.2014.04.002](https://doi.org/10.1016/j.econedurev.2014.04.002)

Abstract: *This paper contrasts measures of teacher effectiveness with the students' evaluations for the same teachers using administrative data from Bocconi University. The effectiveness measures are estimated by comparing the performance in follow-on coursework of students who are randomly assigned to teachers. We find that teacher quality matters substantially and that our measure of effectiveness is negatively correlated with the students' evaluations of professors. A simple theory rationalizes this result under the assumption that students evaluate professors based on their realized utility, an assumption that is supported by additional evidence that the evaluations respond to meteorological conditions.*

Hoefler, P., Yurkiewicz, J., & Byrne, J. C. (2012). The association between students' evaluation of teaching and grades. *Decision Sciences Journal of Innovative Education*, 10, 447–459. [doi:10.1111/j.1540-4609.2012.00345.x](https://doi.org/10.1111/j.1540-4609.2012.00345.x)

Abstract: *The results of a student evaluation of teaching instrument are analyzed for a semester of classes at a large collegiate business school, at both the graduate and undergraduate academic level. In particular, we concentrate on the relationship between responses to the instrument and student grades. The results show interesting differences, suggesting dissimilarities, especially for gender, academic level, and field. Our research suggests that academic administrators and others need to be cautious when they use such data in evaluating teaching.*

Hornstein, H. A. & Fai, H. Edmond Law (Reviewing Editor) (2017) Student evaluations of teaching are an inadequate assessment tool for evaluating faculty performance, *Cogent Education*, 4:1, DOI: [10.1080/2331186X.2017.1304016](https://doi.org/10.1080/2331186X.2017.1304016)

Abstract: *Literature is examined to support the contention that student evaluations of teaching (SET) should not be used for summative evaluation of university faculty. Recommendations for alternatives to SET are provided.*

MacNell, L., Driscoll, A. & Hunt, A. N. (2015). What's in a Name: Exposing Gender Bias in Student Ratings of Teaching. *Innovative Higher Education*, 4, 291-303.

Abstract: *Student ratings of teaching play a significant role in career outcomes for higher education instructors. Although instructor gender has been shown to play an important role in influencing student ratings, the extent and nature of that role remains contested. While difficult to separate gender from teaching practices in person, it is possible to disguise an instructor's gender identity online. In our experiment, assistant instructors in an online class each operated under two different gender identities. Students rated the male identity significantly higher than the female identity, regardless of the instructor's actual gender, demonstrating gender bias. Given the vital role that student ratings play in academic career trajectories, this finding warrants considerable attention.*

Marsh, H. W., & Roche, L. A. (1997). Making students' evaluations of teaching effectiveness effective: The critical issues of validity, bias, and utility. *American Psychologist*, 52(11), 1187-1197. <http://dx.doi.org/10.1037/0003-066X.52.11.1187>

Abstract: This article reviews research indicating that, under appropriate conditions, students' evaluations of teaching (SETs) are (a) multidimensional; (b) reliable and stable; (c) primarily a function of the instructor who teaches a course rather than the course that is taught; (d) relatively valid against a variety of indicators of effective teaching; (e) relatively unaffected by a variety of variables hypothesized as potential biases (e.g., grading leniency, class size, workload, prior subject interest); and (f) useful in improving teaching effectiveness when SETs are coupled with appropriate consultation. The authors recommend rejecting a narrow criterion-related approach to validity and adopting a broad construct-validation approach, recognizing that effective teaching and SETs that reflect teaching effectiveness are multidimensional; no single criterion of effective teaching is sufficient; and tentative interpretations of relations with validity criteria and potential biases should be evaluated critically in different contexts, in relation to multiple criteria of effective teaching, theory, and existing knowledge.

Mitchell, K., & Martin, J. (2018). Gender Bias in Student Evaluations. *PS: Political Science & Politics*, 51(3), 648-652. doi: 10.1017/S10490965`8000001X

Abstract: *Many universities use student evaluations of teachers (SETs) as part of consideration for tenure, compensation, and other employment decisions. However, in doing so, they may be engaging in discriminatory practices against female academics. This study further explores the relationship between gender and SETs described by MacNell, Driscoll, and Hunt (2015) by using both content analysis in student-evaluation comments and quantitative analysis of students' ordinal scoring of their instructors. The authors show that the language students use in evaluations regarding male professors is significantly different than language used in evaluating female professors. They also show that a male instructor administering an identical online course as a female instructor receives higher ordinal scores in teaching evaluations, even when questions are not instructor-specific. Findings suggest that the relationship between gender and teaching evaluations may indicate that the use of evaluations in employment decisions is discriminatory against women.*

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, 598–642.10.3102/0034654313496870

Abstract: *This article provides an extensive overview of the recent literature on student evaluation of teaching (SET) in higher education. The review is based on the SET meta-validation model, drawing upon research reports published in peer-reviewed journals since 2000. Through the lens of validity, we consider both the more traditional research themes in the field of SET (i.e., the dimensionality debate, the ‘bias’ question, and questionnaire design) and some recent trends in SET research, such as online SET and bias investigations into additional teacher personal characteristics. The review provides a clear idea of the state of the art with regard to research on SET, thus allowing researchers to formulate suggestions for future research. It is argued that SET remains a current yet delicate topic in higher education, as well as in education research. Many stakeholders are not convinced of the usefulness and validity of SET for both formative and summative purposes. Research on SET has thus far failed to provide clear answers to several critical questions concerning the validity of SET.*

Stark, P. B., & Freishtat, R. (2014). *An evaluation of course evaluations*. doi:10.14293/S2199-1006.1.SQR-EDU.AOFRQA.v1Stark.

Abstract: *Student ratings of teaching have been used, studied, and debated for almost a century. This article examines student ratings of teaching from a statistical perspective. The common practice of relying on averages of student teaching evaluation scores as the primary measure of teaching effectiveness for promotion and tenure decisions should be abandoned for substantive and statistical reasons: There is strong evidence that student responses to questions of “effectiveness” do not measure teaching effectiveness. Response rates and response variability matter. And comparing averages of categorical responses, even if the categories are represented by numbers, makes little sense. 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.*

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, <https://doi.org/10.1016/j.stueduc.2016.08.007>

Abstract: *Student evaluation of teaching (SET) ratings are used to evaluate faculty's teaching effectiveness based on a widespread belief that students learn more from highly rated professors. The key evidence cited in support of this belief are meta-analyses of multisection studies showing small-to-moderate correlations between SET ratings and student achievement (e.g., [Cohen, 1980](#), [Cohen, 1981](#); [Feldman, 1989](#)). We re-analyzed previously published meta-analyses of the multisection studies and found that their findings were an artifact of small sample sized studies and publication bias. Whereas the small sample sized studies showed large and moderate correlation, the large sample sized studies showed no or only minimal correlation between SET ratings and learning. Our up-to-date meta-analysis of all multisection studies revealed no significant correlations between the SET ratings and learning. These findings suggest that*

institutions focused on student learning and career success may want to abandon SET ratings as a measure of faculty's teaching effectiveness.

Wright, S. L., & Jenkins-Guarnieri, M. A. (2012). Student evaluations of teaching: Combining the meta-analyses and demonstrating further evidence for effective use. *Assessment & Evaluation in Higher Education*, 37, 683–699.10.1080/02602938.2011.563279

Abstract: *There is a plethora of research on student evaluations of teaching (SETs) regarding their validity, susceptibility to bias, practical use and effective implementation. Given that there is not one study summarising all these domains of research, a comprehensive overview of SETs was conducted by combining all prior meta-analyses related to SETs. Eleven meta-analyses were identified, and nine meta-analyses covering 193 studies were included in the analysis, which yielded a small-to-medium overall weighted mean effect size ($r = .26$) between SETs and the variables studied. Findings suggest that SETs appear to be valid, have practical use that is largely free from gender bias and are most effective when implemented with consultation strategies. Research, teaching and policy implications are discussed.*

Wright, R. E. (2006). Student evaluations of faculty: Concerns raised in the literature, and possible solutions. *College Student Journal*, 40, 417–422.

Abstract: *Student evaluations of instruction have long been used to evaluate the teaching performance of instructors. However, despite the widespread use of data from student evaluations for the purpose of determining faculty teaching effectiveness, a review of the literature in the area indicates that issues concerning the validity and usefulness of such evaluations remain unresolved. Given the pervasiveness of usage of such evaluations, substantial changes in the system are unlikely to be implemented across the country. After reviewing key problem areas that have been identified in student evaluations of faculty, this paper suggests some possible methods to increase the validity of teaching evaluations without major changes to the current systems of evaluation of faculty. Such changes could benefit both college students and faculty by increasing the usefulness of student evaluations of faculty.*

Yunker, P. J., & Yunker, J. A. (2003). Are student evaluations of teaching valid? Evidence from an analytical business core course. *Journal of Education for Business*, 78, 313–317.10.1080/08832320309598619

Abstract: *In this study, the authors examined the relationship between student achievement (as measured by student grade earned in Intermediate Accounting I) and student evaluations (as measured by the mean class evaluation of the student's instructor in Introductory Accounting II). Controlling for ability with three variables—student grade in Introductory Accounting II, student grade point average, and student ACT score—the authors found a statistically significant negative relationship between student evaluations and student achievement. Thus, this research strengthens some faculty members' concern that widespread and significant application of student teacher evaluations in faculty performance evaluation may be partially responsible for the interrelated phenomena of grade inflation and content debasement.*

Zipser, N. & Mincieli, L. (2018). Administrative and structural changes in student evaluations of teaching and their effects on overall instructor scores, *Assessment & Evaluation in Higher Education*, 43:6, 995-1008, DOI: [10.1080/02602938.2018.1425368](https://doi.org/10.1080/02602938.2018.1425368)

Abstract: *Using nine years of student evaluation of teaching (SET) data from a large US research university, we examine whether changes to the SET instrument have a substantial impact on overall instructor scores. Our study exploits four distinct natural experiments that arose when the SET instrument was changed. To maximise power, we compare the same course/instructor before and after each of the four changes occurred. We find that switching from in-class, paper course evaluations to online evaluations generates an average change of -0.14 points on a five-point scale, or 0.25 standard deviations (SDs) in the overall instructor ratings. Changing labelling of the scale and the wording of the overall instructor question generates another decrease in the average rating: -0.15 of a point (0.27 SDs). In contrast, extending the evaluation period to include the final examination and offering an incentive (early grade release) for completing the evaluations do not have a statistically significant effect on the overall instructor rating. The cumulative impact of these individual changes is -0.29 points (0.52 SDs). This large decrease shows that SET scores are not comparable over time when instruments change. Therefore, administrators should measure and account for such changes when using historical benchmarks for evaluative purposes (e.g. appointments and compensation).*

Respectfully Submitted,

Jade M. Berry-James, PhD
Academic & Policy Committee (Co-chair)
Minutes Dated: 3.27.2019 Last Rev.