

Ad hoc Committee on Teaching Evaluation Standards

*“The Faculty Senate shall undertake a study of best practices in utilizing methodologies for assessing teaching quality, to include peer-review, classroom visitation, mentoring, student evaluations, and any other methods which may be used in the process of evaluating faculty and in the determination of retention, promotion, and tenure decisions. The committee should examine best practices from other universities and solicit comments from the UNH community.”*

---

Rationale

Faculty in higher education across the country are seeking improved processes for assessment of teaching with primary focus on the purpose of professional development. In particular, there is movement away from using student end-of-semester ratings as the sole metric. This movement is due to dissatisfaction with the utility of the rating information for improving one’s teaching. Furthermore, emerging experimental evidence demonstrates that student ratings fail to measure teaching effectiveness and are subject to substantial implicit bias. Early-course assessment (similar to the UNH Midcourse Assessment Program), peer observation, and portfolio-based documentation are practices that are gaining substantial traction, and which give more voice for students and more insight for instructors. There is no uniformly accepted ideal model, and reports indicate that faculty at other institutions are engaged in similar introspective activity.

Motion

That the Faculty Senate support development of a teaching assessment process that embodies the goal of continual improvement for all faculty, and which provides earlier and more productive engagement of students in this process within each course. That process should include the recommended actions that follow.

Recommendations that would enact the spirit of this motion

- 1) That a faculty senate committee, working with the Center for Excellence and Innovation in Teaching and Learning, develop specific recommendations and guidance for instructors regarding early-semester course assessment including professional development on engaging students in that process.
- 2) That the Center for Excellence and Innovation in Teaching and Learning develop instructional recommendations for students on how to provide constructive professional feedback.

- 3) That a faculty senate committee, working with the Center for Excellence and Innovation in Teaching and Learning, develop specific recommendations and guidance regarding peer-assessment practices, and what documentation may or may not become part of P&T or employment decisions.
- 4) That a faculty senate committee, working with the Center for Excellence and Innovation in Teaching and Learning and the Provost's office, develop specific recommendations and guidance regarding documentation of teaching activity and effectiveness by portfolio, for inclusion in the normal evaluative process for faculty promotion, tenure, and continued employment .
- 5) That a faculty senate committee, working with the Center for Excellence and Innovation in Teaching and Learning and the Provost's office, articulate a set of components of quality teaching, building on the work of the Study Committee 2016-2017.
- 6) That end-of-course student input be called surveys and not evaluations.
- 7) That a faculty senate committee take an entirely fresh look at student end-of-course survey prompts to develop a set (numerical and narrative) that provides more insight regarding student perceptions of learning, instructor behaviors, and course learning conditions. In other words, engage students in commenting on what may have affected their learning.
- 8) That the Center for Excellence and Innovation in Teaching and Learning, working with faculty senate representatives, develop a workshop and set of resources regarding the literature on the validity, bias, and reliability of numerical course surveys, including use of historical UNH data. This workshop may be deployed for Deans, Chairs, and all faculty.

### Preamble to the Report

The Committee took a more proactive position regarding our charge “to study”. What we were learning about the state-of-the-art of teaching evaluation led us to the following principled foundation and vision for the improvement and assessment of teaching at UNH. We were driven to think bigger. To sustain the thrust of this thinking, we proposed the motion to the Faculty Senate, and suggested potential specific follow-up actions to enact that motion.

### Investigation Process

The Committee met fifteen times since last September. Despite our best efforts, a single time was not amenable to the schedules of everyone who initially requested to participate. In the fall, the meeting time was moved to try to accommodate more participants, unsuccessfully. In the spring, to make progress, an unsatisfactory compromise time was selected for a weekly meeting. We were able to carry through an agenda with a core of about five or six people who were able to participate for most of the fall and spring meetings. As chair, I want to especially acknowledge Marieka Brower-Berg, Karen Collins, Diane Freedman, Adele Marone, and John Turner (graduate student). Others were hampered by schedules, but contributed where they could to the discussion over a number of meetings: Susan Endrizzi, James Ramsay, Nena Stracuzzi, and Bryce Moser (undergraduate student).

Notes were generated from each meeting. Each meeting was also video-recorded with the acquiescence of all participants. Those recordings were intended for Committee members who missed or could not regularly attend the meeting times. These were not intended to be shared outside the Committee.

The process in the fall was exploratory. We started by sharing our personal motivations for participating and whatever expertise we would be able to contribute. Then we divided labor to sample what information existed from various external sources to get a sense of the state of the art of teaching evaluation. We explored what our own disciplines contributed to the knowledge base, what particular general higher education journals had to say, reports from faculty study groups at other higher education institutions or organizations, and recommendations from centers for teaching and learning. Our search was one of convenience. We did not complete a comprehensive literature review and analysis, but we did glean sufficient perspective that we feel we could ask good questions pertinent to the committee charge, and follow up on those lines of inquiry. By the end of the fall semester, we felt we understood the lay of the land and knew what specific things we wanted to investigate in the spring. We then dedicated individual meetings in the spring to particular topics, seeded by pertinent readings that we could find.

Every meeting contained substantive discussion. All voices had an opportunity to contribute. Committee members not at the table occasionally contributed, but the meat of the discussion and action emerged from the in-person meetings. The discussion was frank, respectful, and demonstrated a unanimous sense of purpose: to establish a principled foundation for the improvement and assessment of teaching at UNH.

To make sense of the findings, we present an alternative vision of how a multifaceted teaching assessment process could serve the larger purpose of improving teaching and learning. The vision is grounded in four substantive principles.

First, the University should be cultivating educators. To that end, the faculty evaluation process needs to move away from what is now a punitive perspective for not hitting some ill-defined and demonstrably-biased numerical standard, and toward a growth perspective. We need to move away from the assumption that an instructor knows, and can implement, and exhibit “excellence” from the start to the idea that all instructors can and should improve over time, emphasizing the importance of the process of teaching.

Second, faculty members should welcome the opportunity to demonstrate and to observe each other at work as part of the normal enterprise of quality teaching. Just as scholarly activity thrives and grows in the light, the same is true for teaching. Teaching is community property, not private practice (Shulman, 1993). The responsibility for supporting trajectories of improvement is an institutional and professional responsibility.

Third, students should gain more insight about teaching as a process, and that will enable them to be more cooperative, engaged, and successful learners. The current process does not provide this opportunity. Instructors should incorporate conversational opportunities early in the semester to help students understand how the course design, activities, and expectations support their learning. In other words, students should not just feel that teaching is something that is “done” to them. We need to understand how they perceive and receive what we ask them to do, and share with them our goals and rationale for why. We should represent it as a dialectic opportunity which raises their perspective about what it means to be accountable for their work and what it means to think hard about something.

Fourth, in order to justify a process for evaluation of teaching, the university and faculty should clearly articulate the components of quality teaching. Many voices in the literature talk about quality teaching but there are scant articles or reports that express exactly what that means. So our Committee took the time to do that. The table at the end summarizes what the Committee members agreed were nineteen essential components of teaching quality. We also categorized these into five areas: design, pedagogy, content, reflection, engagement. We argue that an instructor, who does more of any of the listed components, is engaging in higher

quality instructional work. All of these components are measurable. Evidence regarding each of these components may be developed as part of the improved teaching evaluation process we describe below.

### Vision

This description of an improved teaching development and evaluation process incorporates the Committee's perceptions of emerging "best practices" in higher education. Each of the components of this process would require discussion and development by the faculty in coordination with the university administration.

The Committee envisions a process of teaching assessment that focuses on formative development of teaching knowledge, skill, and practices for individual instructors at all levels of faculty. Each course is a "data point" in this process, contributing to a trajectory of activity that is documented in the form of a teaching portfolio. The portfolio becomes the evidentiary source for administrative evaluation at appropriate points. Mentors may assist new faculty in negotiating the process and engaging students in feedback.

At the beginning of the semester, each instructor is explicit with students regarding expectations and learning outcomes for each course they teach, and describes how course activities support their expectations and learning outcomes, starting with the syllabus. Early in the semester the instructor gives the students an opportunity to engage in assessment of course progress and their own learning. This happens about one-third of the way through a course (e.g. at about 4-5 weeks for a traditional semester course). The current MAPS process is one way to accomplish this, but the results must be handled quickly and shared with students. Then at the end of the course, a student survey is provided, which asks for reflection back to the early semester conversation. The dual assessment opportunities, along with the feedback and explicit discussion with the instructor, bring students into a deeper conversation about learning and in a way that is to their direct benefit during the course.

During the semester, faculty peers engage in observing classes and inspecting class materials as part of a conversation with the instructor regarding the teaching and learning in that course. The visit process consists of a preface meeting and a debriefing meeting so that what is observed can be positioned in terms of its purpose to the larger goals of the course. A likely structure is for two to three peers to participate in the review, and for review to involve a number of courses commensurate with that instructor's appointment. This process may be used more often for junior faculty and instructors, but should not be abandoned for senior faculty. Guidelines would need to be developed for the process.

At the end of the semester, students would again respond to questions that are relevant to their perspective as daily participants in a class, with a focus more on behaviors of the instructor or description of class norms and away from evaluative judgments. The end-of-semester survey (not “evaluation”) given to students has both numerical and narrative features. Numerical questions help identify for the instructor behaviors or classroom conditions that enhance or detract from learning. Narrative questions probe details about how course structure or events or assessments supported learning. Narrative questions are essential to extracting deeper insight regarding the students’ experience with the course.

Instructors maintain a teaching portfolio, updated with each semester. The instructor uses the peer review comments and student input from both early and end-of semester to engage in reflection about the course and student learning, and that reflection becomes part of their teaching portfolio. Evidence accumulates in this portfolio for each course, and is available for review for purposes of evaluation.

The slide shows an illustration of the overall process.

### Findings

Our general finding is that, although there is a sense of what components should be included in “best practices” for teaching evaluation, there is little reliable research-based guidance. There is, however, plenty of opinion. We highlight and elaborate on specific findings by the Committee. Findings are numbered to allow easy reference when talking about them.

1) *All of the literature we found presents its primary purpose as supporting the goal of improving teaching and learning. None of the literature focused primarily on supporting employment and P&T decision-making.*

Irrespective of what methods of teaching assessment were being discussed, the claim was made that these methods were being used as a means for improving teaching and learning. Many institutional documents acknowledged that this information may feed into P&T and appointment-renewal processes, but we saw nothing that provided new insight as to how this might best be accomplished.

2) *There is movement away from the sole use of student end-of-course surveys for evaluation. There is increasing evidence that they are uni-dimensional, fail to measure what we hope they would measure, provide weak leverage for supporting positive changes in teaching, and are known to be subject to implicit bias.*

Arguments at the level of opinion continue regarding the value of student input. However, there are recent studies that cast serious doubt on the validity and reliability of student surveys as a measure of teaching effectiveness. The concern is especially directed at administrative decision-making. Uttl, White, and Gonzalez (2016) in a meta-analysis of all prior research on student ratings, demonstrates in a strong way that student ratings are unrelated to teaching effectiveness, and an artifact of small samples sizes and publication bias. Carrell and West (2008) in a study at the Air Force Academy, where the curriculum, assessments, goals, and expectations are closely similar across many sections of the course, have shown that less experienced instructors get higher student ratings and award higher grades to students in the current semester than do more experienced instructors, but students did better in subsequent semesters if they had the more experienced and tougher-grading instructors the previous semester. Braga, Paccagnella, & Pellizzari (2014) found similar results at an Italian school of business and public policy where randomized enrollment was involved. They also showed that students evaluate instructors more negatively on rainy and cold days. MacNell, Driscoll, & Hunt (2014) took advantage of an on-line class where a male instructor and female instructor each presented themselves as a male and female identity. Students rated instruction significantly lower if the instructor identity was female, regardless of the instructor's actual sex. The difference was equivalent to 0.5 on a 5 point scale. Boring, Ottoboni, and Stark (2016) analyzed ratings from thousands of students at a French university over five years where course surveys were mandatory. Results indicated a bias against female instructors' teaching, which carried over into unrelated questions such as how promptly assignments were graded.

3) *There is continuing movement in higher education to incorporate systematically a process of peer-based review of teaching.*

Chism (2007) has authored what appears to the "bible" on faculty peer review of teaching. It offers a thorough analysis of the challenges and balanced view of the process, as well as providing guidance and rubrics that could be used to support the process. Our committee deployed a survey to all department and program chairs to assess the extent to which peer review was already being used at UNH. With only ten replies (reasonable representation across colleges), we can't say how widespread the practice is but we can say that what exists now tends to be aimed at evaluative purposes and as a complement to student surveys. The processes reported seem to lack the desirable pre and post components of visits. Inspection of course materials seems to be rare. This attitude toward peer observation reflects an evaluative stance in which the observer or reviewer is the expert and is looking for how the observed person is deficient from some norm. As mentioned previously, since there is nothing developed explicitly to define what the quality components of teaching are, the observation is unguided. Reviewers in fact may learn something by conducting the observation. So a strong peer process acknowledges that the learning will go both ways if that conversation happens.

4) *There is continuing movement on other campuses to use the portfolio approach for documenting teaching practices over time.*

There is a lot of literature on portfolios, and potential electronic platforms that would allow for systematic development of these as documentation of an instructor's trajectory of development. Our experience with the Faculty Activity Report (FAR) suggests that this approach is unsuitable. The FAR is designed to count things and is extremely unwieldy for handling narrative or reflective commentary on teaching.

5) *Students should receive guidance or professional development on how to provide constructive feedback, this process should start early, and it should be informed by research literature on the subject.*

Some campuses are doing this (e.g. Michigan Center for Research on Learning and Teaching) where students are given general suggestions on tone and purpose, and specific examples of less and more helpful comments on instructor surveys. Current end-of-semester surveys are often criticized because students have limited experience and vision, and may provide vague comments. Yet little to nothing is done to help them do a better job in giving feedback that might improve their learning experiences. One place where this may be accomplished is the early-semester assessment followed by a conversation between instructor and students. Another insight comes from the work by Seymour (1997) on the Student Assessment of Learning Gains (SALG), which was developed as an assessment for STEM curriculum development assessment. Asking students how something affected their learning generated more useful responses than asking them what they liked or not about a course, which tends to generate complaints. Perry's work on epistemological development is also pertinent here: students and instructors don't have the same idea about what it means to know something. Thus, students will have difficulty understanding the why of a course unless (a) the instructor has a why and (b) a discussion about it happens.

6) *Faculty should receive guidance on how to engage students and peers in a discussion where they lay out their course design and pedagogic decisions.*

Some faculty may have evaluation anxiety and feel that peer visitors are there to criticize and students are just being given an opportunity to complain. A more generous sentiment is one in which we all would be happy to invite colleagues into our classrooms because we want to share what we are doing and why – just like we do with our scholarly work, and that students would value and learn from getting deeper insight regarding how one thinks about and organizes teaching within the challenges of that discipline.

7) *The end-of-semester on-line student survey has many uncontrollable factors that complicate employment decisions and P&T discussions. Instead of trying to fix the data after collection, we should improve the design by which it is collected.*

The current end-of-semester survey approach is intended to be a census of independent respondents who have a shared understanding of the questions being asked and an experience base that allows an informed judgment about teaching effectiveness (i.e. learning).

There is justifiable concern about student response rates, which has not been improved by the move to exclusively on-line format. The literature repeatedly points to response rates below 85% as a concern, and cites carrot and stick processes to try to improve response rates, such as bonus points or withheld grades. The concern with low response rates is the potential that extreme opinion-holders will be more motivated to respond, and that negative opinions may be more motivated than positive. We did not find studies specific to this issue, but we point to the tone of sites like Rate-My-Professor as evidence for non-representative response patterns. We are aware from student comments about the evaluation process that some students intend to “go after” certain instructors. We did not hear similar intentions to provide positive comments.

Responses are intended to be independent, but it is known that in the on-line environment students may respond cooperatively. Since students would have been talking about the course and instructor anyway, this does not seem to be induced uniquely by the on-line format.

Substantial discussion in the literature pertains to the nature of questions presented to students, including questions for which students lack the experience or expertise to provide an informed judgment. A lot of effort has gone into and continues to go into finding the “right” set of questions and wordings, and to provide nuance for different course venues (e.g. on-line). There was not much new except the acknowledgment that numerical ratings alone don’t provide any insight regarding why students are responding as they do. Alternative approaches (now being used in research investigations, e.g. Lund *et al.*, 2015 ) focus on behaviors and conditions, and avoid evaluative judgments. This has allowed reliable description and categorization of course structures and pedagogies. We believe a student numerical survey could be constructed that focused on describing the mechanical and physical characteristics of their class experience that would allow capturing of issues that students may not report in written comments (e.g. class duration, missed sessions, frequency of assessment opportunities, nature of feedback). Comments on some of these issues would triangulate with peer observation. We are aware of students who don’t bother to report these issues.

We saw very little consideration for course settings such as clinical, the arts, or other practice-oriented classes. This topic arose in our Committee since we had representatives from those communities.

8) *The large data base and historical record of student surveys at UNH should be used to provide reliable estimates and insights for decision-makers, should end-of-semester numerical survey data continue to be used.*

Our primary concern here is that faculty and administrators compare and talk about course evaluation results as if the overall rating question 14 tells the entire story of the course, and that this number is interpreted without sufficient context. In particular, given that any numerical measure, even if unbiased, will be subject to noise, it is disturbingly rare in P&T discussions to hear comments about the confidence interval to be expected based on historical precedent, or the expected magnitude of bias regarding class size or type of students enrolled. This raises the chance that discussions about teaching outcomes will focus on individual and idiosyncratic events or student comments rather than on the broader consideration of a person's teaching abilities. The literature is not helpful on this issue. We find that essentially no one discusses in a statistically justifiable way how to interpret numerical ratings for purposes of making judgments (i.e. P&T and re-appointment decisions). A few novel studies have proposed control chart approaches to monitoring progress of instructors or courses over time.

Using some of our own historical data, the Committee was able to establish the likely variability range for individual instructors in a particular course and for a group of instructors in a common course. A course standard deviation of 0.5 was found. This suggests that a normal range for this course is plus or minus 1 unit out of 5.

It is particularly disturbing that student survey ratings seem to carry the weight of employment evaluation. The Committee believes that, if numerical surveys are to be used, a more justifiable context needs to be established for claiming that someone has reached a certain bar, or whether there is evidence of improvement, or no improvement, or trending over time.

### Surprising Findings

- There are many studies that try to make sense of whether student end-of-semester ratings are valid indicators of teaching effectiveness for students in those courses. Most of the studies that are supportive of student rating conclude by indicating that they are useful but require cautious interpretation. In contrast, there are several definitive studies that demonstrate serious deficiencies that contradict this claim. Furthermore, clear evidence of bias exists.

- There are many opinion pieces on how to approach the challenge of “evaluation” of faculty via student ratings or other processes. They are all repetitive and fail to advance our understanding.
- There are many opinion pieces arguing that end-of-semester student surveys are valuable for instructors in improving teaching. We have read descriptions from people who explain how they use student surveys for this purpose. We have not found any literature that studies this claim in a systematic way.
- There is a lot of talk about improving teaching quality without articulating what those qualities are such that someone could work to improve them.
- There is precious little guidance for administrators on what to do with numerical student ratings that could lead to statistically defensible decision-making.
- Although people talk about divorcing formative goals (teaching improvement) from evaluative goals (appointment or merit decisions), no one does that because of the workload it creates.
- Given all of the effort on campuses to deploy approaches to gather student input regarding teaching, you would think we would have a better understanding.

Appendices:

Table: Components of Quality Teaching

Bibliography

A lot of literature on quality speaks to the “institution” or to a “major”. Current end-of-semester surveys pertain to “the instructor” and “the course”. We want to focus on the “course” as a unit. Too many other stakeholders come into play when the field of view is broadened. “The course” is something an individual instructor can do something about, and be responsible for.

Components of Teaching Quality

	Tentatively identified categories:	Design	Delivery Pedagogy	Content	Reflection	Engagmnt
1	Instructor makes course design explicit (for students and colleagues). Course is purposeful in developing student understanding.	X				
2	Instructor makes course goals (big picture) explicit (for students and colleagues)	X				
3	Instructor makes learning objectives (specific know and do verbs) explicit and measurable (for students and colleagues), with assessments that align with the objectives and course design. And with the realization that students will need frequent reminders and reinforcement.	X		X		
4	Instructor communicates concerning expectations for students, and provides timely and appropriate feedback regarding work products.		X	X		
5	Instructor considers principled implementation of evidence-based pedagogic practices.	X	X			
6	Instructor engages in conversation with students early (1/3 way) during a class regarding their perception and adaptation to course goals, design, objectives, and learning expectations.	X				X
7	For each course, instructor seeks and uses student, peer, and personal observations to reflect on course goals, outcomes, and structure, with an eye toward improving the learning experience. This is documented for personal development and as a source of promotion/employment evidence.				X	
8	Instructor demonstrates disciplinary expertise and state-of-the-art knowledge, including regarding learning challenges for students in that discipline (pedagogical content knowledge).			X		
9	Instructor enables students to reflect on their learning and articulate something about the content they have learned.			X	X	
10	Course content goals include how it integrates with other fields of study or application. (as opposed to narrow isolated and context-less content)			X		
11	Instructor conducts class in a way that makes efficient use of class time for learning.		X			

Components of Teaching Quality Feb 2017 Faculty Senate Ad Hoc Committee on Teaching Evaluation Standards

12	Instructor manages class sessions in a way so that all students have effective access to the information and experiences presented. This includes being able to see and hear, to process information, to be able to ask questions. This also includes showing up ahead of time, being ready to go on the clock and using the full period, and not randomly canceling classes.		X			X
13	Instructor makes him/herself available for individual student meetings (in person or electronic).					X
14	Instructor demonstrates passion for the subject.		X			X
15	Instructor develops a rapport and respectful relationship with students.					X
16	Instructor engages students in ways that promote the intellectual growth of the individual and in ways that meet their learning needs. (not driving people away)					X
17	Instructor approaches concepts to be learned through a variety of learning modalities.	X				
18	Instructor introduces students to the norms and protocols of the discipline.			X		
19	Instructor sets students up to become aware of professional performance expectations.			X		

References

- Baiduc, R., R., Linsenmeier, R. A., & Ruggeri, N. (2016). Mentored Discussion Of Teaching: An Introductory Teaching Development Program For Future STEM Faculty. *Innovative Higher Education, 41*, 237-254. doi:10.1007/s10755-015-9348-1
- Bartlett, A. (2005). “She seems nice”: Teaching evaluations and gender trouble. *Feminist Teacher, 15* (3), 195-202.
- Berk, R.A. (2006). *Thirteen strategies to measure college teaching: A consumer’s guide to rating scale construction, assessment, and decision making for faculty, administrators, and clinicians*. Sterling, VA: Stylus.
- Bernstein, D. J. (1996). A Departmental System For Balancing The Development And Evaluation Of College Teaching: A Commentary Of Cavanagh. *Innovative Higher Education, 20*(4), 241-247.
- Boring, A., Ottoboni, K., & Stark, P. B. (2016). Student Evaluations of Teaching (Mostly) Do Not Measure Teaching Effectiveness. *ScienceOpen Research*. 10.14293/S2199-1006.1.SOR-EDU.AETBZC.v1.
- Braga, M., Paccagnella, M., & Pellizzari, M. (2014). Evaluating Students’ Evaluations of Professors. *Economics of Education Review, 41*, 71-88.
- Brent, R., & Felder, R.M. (2004). A Protocol for Peer Review of Teaching. *Proc. Amer. Soc. Eng. Educ. Session 3530*.  
[http://www4.ncsu.edu/unity/lockers/users/f/felder/public/Papers/ASEE04\(Peer-Review\).pdf](http://www4.ncsu.edu/unity/lockers/users/f/felder/public/Papers/ASEE04(Peer-Review).pdf)

Report of Committee on Teaching Evaluation Standards, Appendix: Reading List, April 2017

Brown, B., & Ward-Griffin, C. (1994). The use of peer evaluation in promoting nursing faculty teaching effectiveness: A review of the literature. *Nurse Educ. Today* 14 (4), 299-305.

Carrell, S. E., & West, J. E. (2008). Does Professor Quality Matter? Evidence From Random Assignment Of Students To Professors. Working Paper 14081. National Bureau of Economic Research, Cambridge, MA.

Casey, R. J., Gentile, P., & Bigger, S. W. (1997). Teaching Appraisal In Higher Education: An Australian Perspective. *Higher Education*, 34(4), 459-482. doi:10.1023/a:1003042830109

Cavanagh, R. R. (1996). Formative And Summative Evaluation In The Faculty Peer Review Of Teaching. *Innovative Higher Education*, 20(4), 235-240.

Chism, N. V. N. (2004). Characteristics Of Effective Teaching In Higher Education: Between Definitional Despair And Certainty. *Journal of Excellence in College Teaching*, 15(3), 5.

Chism, N. V. N. (Ed.). (2007). *Peer review of teaching: A sourcebook*. (second ed.) Anker Publishing Company, Inc.

Cohen, P. A. (1981). Student Ratings Of Instruction And Student Achievement: A Meta-Analysis Of Multisection Validation Studies. *Rev. Educ. Research*, 51(3) 281-309.

Cohen, P. A. (1980). Effectiveness Of Student Rating Feedback For Improving College Instruction: A Meta-Analysis Of Findings. *Research in Higher Education*, 13, 321.

Cornell Teaching Evaluation Handbook, 1997

<https://www.cte.cornell.edu/documents/Teaching%20Evaluation%20Handbook.pdf>

Cosser, M. (1998). Towards The Design Of A System Of Peer Review Of Teaching For The Advancement Of The Individual Within The University. *Higher Education*, 35(2), 143-162. doi:10.1023/a:1003155703281

Dolmas, D., Wolfhagen, H., Scherpbier, A., & Van der Vleuten, C. (2003). Development Of An Instrument To Evaluate The Effectiveness Of Teachers In Guiding Small Groups. *Higher Education*, 46(4), 431-446. doi:10.1023/a:1027388806218

Draeger, J., del Prado Hill, P., & Mahler, R. (2015). Developing A Student Conception Of Academic Rigor. *Innovative Higher Education*, 40, 215-228. doi:10.1007/s10755-014-9308-1

Drew, S., & Klopper, C. (2014). Evaluating Faculty Pedagogic Practices To Inform Strategic Academic Professional Development: A Case Of Cases. *Higher Education*, 67(3), 349-367. doi:10.1007/s10734-013-9657-1

Dukes, R. L., & Victoria G. (1989). The Effects of Gender, Status, and Effective Teaching on the Evaluation of College Instruction. *Teaching Sociology* 17, 447-457.

Embry, M. (2015). Am I a Good Teacher? What Do My Evaluations Say? In *Quick Hits for Adjunct Faculty and Lecturers: Successful Strategies from Award-Winning Teachers*. Morgan, R. K., Olivares, K. T., & Becker, J. (eds.). University of Indiana Press, Bloomington, IN. 98-99.

Flaherty, C. (2016). Zero Correlation Between Evaluations And Learning. *Inside Higher Education*,

- Fraile, R., & Bosch-Morell, F. (2015). Considering Teaching History And Calculating Confidence Intervals In Student Evaluations Of Teaching Quality. *Higher Education*, 70(1), 55-72. doi:10.1007/s10734-014-9823-0
- Ghedin, E., & Aquario, D. (20078). Moving Towards Multidimensional Evaluation Of Teaching Higher Education: A Study Across Four Faculties. *Higher Education*, 56(5), 583. doi:10.1007/s10734-008-9112-x
- Gormally, C., Evans, M., & Brickman, P. (2014). Feedback about Teaching in Higher Ed: Neglected Opportunities to Promote Change. *CBE-Life Sciences Educ.* 13, 187-199.
- Hampton, S.E. & Reiser, R.A. (2004). Effects of a theory-based feedback and consultation process on instruction and learning in college classrooms. *Research in Higher Education*, 45 (5), 497-527.
- Hativa, N. Can Faculty Misinterpretation And Misuse Of Student Rating Results Lead To The "Dumbing Down" Of College Education? *Student ratings of instruction: Recognizing effective teaching* ()
- Husbands, C. T. (1997). Variations In Students' Evaluations Of Teachers' Lecturing In Different Courses On Which They Lecture: A Study At The London School Of Economics And Political Science. *Higher Education*, 33(1), 51-70. doi:10.1023/a:1002911620174
- Jaquett, C. M., VanMaaren, V. G., & Williams, R. L. (2016). Course Factors That Motivate Students To Submit End-Of-Course Evaluations. *Innovative Higher Education*, doi:10.1007/s10755-016-9368-5

Kell, C., & Annetts, S. (2009). Peer Review Of Teaching Embedded Practice Or Policy-Holding Complacency? *Innovations in Education and Teaching International*, 46(1), 61.

Kember, D., & Wong, A. (2000). Implications For Evaluation From A Study Of Students' Perceptions Of Good And Poor Teaching. *Higher Education*, 40(1), 69.  
doi:0.1023/a:1004068500314

Kite, M.E., Subedi, P.C., & Bryant-Lees, K.B. (2015). Students' perceptions of the teaching evaluation process. *Teaching of Psychology* 42 (4), 307-314.

Kuzmanovic, M., Savic, G., Popovic, M., & Martic, M. (2013). A New Approach To Evaluation Of University Teaching Considering Heterogeneity Of Students' Preferences. *Higher Education*, 66(2), 153. doi:10.1007/s10734-012-9596-2

Lyde, A. R., Grieshaber, D. C., & Byrns, G. (2016). Faculty Teaching Performance: Perceptions Of A Multi-Source Method For Evaluation (MME). *Journal of the Scholarship of Teaching and Learning*, 16(3), 82. doi:doi: 10.14434/josotl.v16i3.18145

MacNell, L., Driscoll, A., & Hunt, A. N. (2015). What's In A Name: Exposing Gender Bias In Student Ratings Of Teaching. *Innovative Higher Education*, 40, 291-303.  
doi:10.1007/s10755-014-9313-4

Malone, R.P., Van Ness, B.F., & Van Ness, R. A. (2005). Online and In-class Student Evaluations. *J. Financial Educ.* 31, 15-22.

McGowen, W.R. & Osgathorpe, R.T. (2011). Student and faculty perceptions of effects of midcourse evaluation. *To Improve the Academy*, 29, 160-172.

Report of Committee on Teaching Evaluation Standards, Appendix: Reading List, April 2017

Morehead, J. W., & Shedd, P. J. (1997). Utilizing Summative Evaluation Through External Peer Review Of Teaching. *Innovative Higher Education*, 22(1), 37-44.

Nikolaidis, Y., & Dimitriadis, S.G. (2014). On the student evaluation of university courses and faculty members' teaching performance. *European Journal of Operational Research*, 238, 199-207.

Obenchain, K.M., Abernathy, T.V., & Wiest, L.R. (2001). The Reliability of Students' Rating of Faculty Teaching Effectiveness. *College Teaching*, 49, 100-104.

Partington, P. (1998). Evaluating Teacher Quality In Higher Education. *Higher Education*, 35(1), 115. doi:10.1023/a:1002929712550

Pratt, D. D. (1997). Reconceptualizing The Evaluation Of Teaching In Higher Education. *Higher Education*, 34(1), 23. doi:10.1023/a:1003046127941

Quinlan, K. M. (1996). Involving Peers In The Evaluation And Improvement Of Teaching: A Menu Of Strategies. *Innovative Higher Education*, 20(4), 299-307.

Research Corporation for Science Advancement (2015). Searching for Better Approaches: Effective Evaluation of Teaching and Learning in STEM. Tucson, AZ: Research Corp.

Schindler, L., Puls-Elvidge, S., Welzant, H., & Crawford, L. (2015). Definitions of Quality in Higher Education: A Synthesis of the Literature. *Higher Learning Research Communications* 5(3), 3-13. <http://dx.doi.org/10.18870/hlrc.v5i3.244>.

Shulman, L.S. (1993). Teaching as Community Property: Putting an End to Pedagogical Solitude. *Change*, 25 (6), 6-7.

Spooren, P., Brockx, B., & Mortelmans, D. (2013). On the Validity of Student Evaluation of Teaching: The State of the Art. *Rev. Educ. Research*, 83 (4) 598-642

Stanford University Course Evaluation Committee Report. (2013).

[https://vptl.stanford.edu/sites/default/files/cec\\_report\\_dec\\_18\\_1.pdf](https://vptl.stanford.edu/sites/default/files/cec_report_dec_18_1.pdf).

Stupans, I., McGuren, T., & Babey, A. M. (2016). Student Evaluation Of Teaching: A Study Exploring Student Rating Instrument Free-Form Text Comments. *Innovative Higher Education*, 41, 33-42. doi:0.1007/s10755-015-9328-5

Teaching Quality in Higher Education: Literature Review and Qualitative Research. (2016).

Department for Business, Innovation & Skills.

<https://www.gov.uk/government/publications/teaching-quality-in-higher-education-literature-review-and-qualitative-research>. See also Higher Education Academy

([www.heacademy.ac.uk](http://www.heacademy.ac.uk)).

Tigelaar, D. E. H., Dolmans, D., Wolfhagen, I., & Van der Vleuten, C. P. M. (2004). The Development and Validation of a Framework for Teaching Competencies in Higher Education. *Higher Education*, 48(2), 253. doi:10.1023/B:HIGH.0000034318.74275.e4

Trujillo, J.M., DiVall, M.V., Barr, J., Gonyeau, M., Van Amburgh, J.A., Matthews, S. J., & Qualters, D. (2009). Development of a Peer Teaching-Assessment Program and a Peer Observation and Evaluation Tool, *Am. J. Pharm. Educ.* 72 (6), article 147.

Tucker, B. (2014). Student Evaluation Surveys: Anonymous Comments That Offend Or Are Unprofessional. *Higher Education*, 68(3), 347. doi:10.1007/s10734-014-9716-2

VanMaaren, V. G., Jaquett, C. M., & Williams, R. L. (2016). Factors Most Likely To Contribute To Positive Course Evaluations. *Innovative Higher Education*, doi:10.1007/s10755-016-9360-0

Way, D. (1997). *Teaching evaluation handbook*. Center for Teaching Excellence, Cornell University.

Whittington, L. A. (2001). Detecting Good Teaching. *Journal of Public Affairs in Education* 7(1). 5-8.