Engineering Educators Bringing the World Together
2025 ASEE Annual Conference & Exposition
Palais des congrès de Montréal, Montréal, QC • June 22–25, 2025 ASEE
Paper ID #48366

WIP: Establishing Peer Observation of Engineering Teaching (POET) as a More Effective Means of Evaluating Teaching

Dr. Stacy Tantum, Duke University

Stacy Tantum received the B.S. in Electrical Engineering and Economics from Tufts University in 1994, and the M.S. and Ph.D. degrees in Electrical Engineering from Duke University in 1996 and 1998, respectively. She is currently an Associate Professor of the Practice in the Department of Electrical & Computer Engineering at Duke University

Dr. Sophia T Santillan, Duke University
Dr. Elizabeth Kathleen Bucholz, Duke University

Dr. Bucholz is an Associate Professor of the Practice for the Department of Biomedical Engineering at Duke University and is the Director of Undergraduate Studies for the Department of Biomedical Engineering in the Pratt School of Engineering. She graduated with her Ph.D. in Biomedical Engineering from Duke University in 2008 and has been teaching at Duke since 2010. She teaches classes such as Signals and Systems, Modern Diagnostic Imaging Systems, Freshmen Design and Communication, a Magnetic Resonance Imaging class and a graduate version of Signals and Systems.

Dr. Genevieve M Lipp

Genevieve Lipp received a B.S.E. in mechanical engineering from Duke University in 2010 and a Ph.D. in 2014 with a focus on nonlinear dynamical systems. She now works in the Center for Instructional Technology and teaches dynamics at Duke University.

Benjamin Cooke, Duke University
Victoria Akin, Duke University

WIP: Establishing Peer Observation of Engineering Teaching (POET) as a More Effective Means of Evaluating Teaching

Abstract

This work in progress paper investigates a model for peer observation of engineering teaching (POET) and its potential use in regular evaluations for instructors. Applying lessons learned from an initial pilot program as well as perspectives from the literature, we outline an initiative for peer evaluation. We include options for data collection to assess the impact of such a program. We discuss broader implications of the work including creating an intellectual community to improve pedagogy, supporting new instructors, and addressing systemic biases in teaching evaluations. We prefer to present this work in a lightning talk.

Background and motivation

Faculty members devote significant time and energy to the pursuit of teaching excellence. Just as we give students the opportunity to practice, receive feedback, and improve their skills, faculty benefit from receiving helpful feedback to improve their teaching. At many institutions, the current primary teaching evaluation mechanism is student course evaluations, which are shown to contain biases. Research in this area shows that student teaching and course evaluations exhibit bias against instructors who: are women [13, 17], are nonnative English speakers [22], are racial minorities [16], or issue comparatively lower grades throughout the semester (regardless of student learning outcomes) [21, 6, 15]. Other studies have found that students tend to review instructors of the same sex more favorably, an outcome resulting in bias against the minority sex in a given field [14, 4]. Finally, reviewing course evaluations can have a negative impact on instructor morale. Faculty burnout has been shown to correlate with numerical student evaluation outcomes [11].

Despite significant evidence of bias in the literature, many institutions rely partially or fully upon student evaluation results to evaluate teaching quality in the context of decisions for promotion, reappointment, and annual merit-based salary increases for a given instructor [18]. The use of student evaluations of teaching for personnel decisions is well-documented [5, 8], and the prevalence of this administrative use is such that the American Sociological Association issued a statement in 2020 outlining the myriad reasons why the use of student evaluations of teaching to assess teaching efficacy is "problematic" and providing a set of evidence-based best practices to incorporate student perspectives in a holistic evaluation of instructor teaching effectiveness [2].

At our institution, Duke University, teaching and course evaluations ask students to rate instructor quality, among other course features, for each of their courses at the end of each semester on a 5-point Likert scale. Analysis of our engineering school's student course evaluation data shows that male students rate male professors more favorably than female professors, and female students rate female professors more favorably than male professors. The difference in rating distributions by female students, compared with those by male students, of female faculty is statistically significant (Chi-squared test, $p < 0.05$), and the same is true for ratings of male faculty

($p < 0.01$). These differences are consistent with literature [14, 4], and highlight the subjectivity of student course evaluations. Within our engineering school, approximately two thirds of the undergraduate students are male, and greater than two thirds of the faculty are male. The underlying bias in the evaluation data benefits male faculty disproportionately because male students rate male faculty higher than female faculty, and both male students and male faculty are overrepresented in school-wide aggregation and averaging of course evaluation ratings. Though these results demonstrate that a gender bias is present, there is no reason to believe that a similar inequitable trend does not exist across other demographic groups studied in the research. To that end, we were motivated to find equitable ways to give faculty feedback that are useful not only for continuous professional development as instructors, but also as summative evaluation for reappointment and promotion, as well as annual merit-based salary adjustments.

Our peer teaching observation program goals are to enable and encourage continuous improvement of instruction and the design of learning experiences; to provide constructive feedback to instructors and individuals who review teaching quality as part of assessment for reappointment, promotion, or annual merit-based salary adjustments; and to promote equity for all faculty, mitigating effects of the bias demonstrated in student teaching and course evaluations. The POET program involves regular group meetings (roughly every three weeks) to establish a community of practice. These meetings provide a place to share concerns, talk through obstacles, and move toward clarity. Observations together with group meetings may lead to increased self-reflection on teaching because participants are held accountable to convene and discuss pedagogy. Moreover, peer observation may lead to increased investment in teaching and could motivate instructors to prepare more thoroughly for class. Establishing a community of practice and building trust among members of the group may be critical to maintaining an environment in which instructors feel supported while receiving feedback.

This work in progress is an initial report on a faculty-initiated effort of a particular model for peer observations of teaching. In Spring and Fall 2024, we, six instructors, piloted a Peer Observation for Engineering Teaching (POET) community of practice to work toward goals of idea exchange and peer-to-peer feedback for faculty in and adjacent to engineering. Our community of practice model was informed by peer observation programs designed to strengthen both teaching and student learning at other universities [12, 7, 9]. In particular, we opted to avoid a checklist-style observation process to allow for each observation to focus on elements that the observed instructor believed they would find most helpful [19]. Instead, we leaned on guided observation prompts (provided in the appendix) to shape the observation and inform the post-observation conversation. Throughout the Spring 2024 and Fall 2024 semesters, we observed each other's classes, discussed what we saw, and shared feedback with each other. Our peer observation efforts are ongoing in Spring 2025, with a focus on offering the peer observation to junior faculty within engineering. The focus on junior faculty is motivated to provide this resource to our newer faculty who may have the most to gain from peer observation.

Methods

Many resources exist for developing peer teaching observation protocols, evaluating institutional or department-level teaching evaluation processes, and training faculty to objectively and/or constructively evaluate teaching through classroom visits [10, 3, 1, 20]. Starting in January of 2024, our POET group met roughly every three weeks for an hour, beginning with a review of many existing peer evaluation protocols. The group elected to begin our work with informal classroom visits with the goal of developing trust before implementing more structured evaluation protocols. Each participant performed two classroom observations of each of the other instructors

(schedule permitting). Prior to their classroom visit, observers corresponded via email with the instructor to be observed in order to discuss particular areas of focus the instructor requested. During these two semesters, observers visited lecture and discussion meetings. During these visited class meetings, some observers took notes of what they observed and of questions to ask the instructor in a later meeting; others used the set of guiding prompts.

In group meetings, observers provided verbal, in-person feedback to the observed instructor. The explicit goals of the feedback sessions were to offer observations and ask questions of the instructor about particular choices they made in the design of the observed class session and/or of their course as a whole. Observers did not express value judgments of pedagogical practices or offer unsolicited advice to address challenges. Within the POET group, we each have a unique teaching style that has developed from our own experiences and development. After these discussions, we agreed that good teaching can take many forms and that it's essential to develop a positive, authentic rapport with students, however each instructor does that well. Additionally, we felt the importance of developing and maintaining trust within the working group to better give and receive feedback about instruction.

An equitable observation, evaluation, and feedback process would allow individual instructors to teach to their own strengths, continue developing their teaching, and learn from colleagues about effective and engaging teaching practices. While bias is certainly not fully mitigated by a peer observation process, the additional perspective from a colleague removes biases related to an expected grade in the course, whether the course is required or elective, and other factors inherent to the student-instructor dynamic. Our emphasis on a set of guided observation prompts, rather than a checklist, and group conversation among observers and observee serve to shift the discussion away from what an individual may believe good teaching looks like (a potentially biased perspective) to an inclusive group exploration of the variety of forms good teaching may take.

Garnering the support of institutional leadership has been instrumental in the ongoing development of POET. While leadership was impressed by the extensive literature surrounding biases present in student evaluations of teaching, there were questions regarding the types and prevalence of bias that are present in student course evaluations at our institution. Thus, the most impactful element in our conversations with leadership was analysis of our own course evaluation data that demonstrate a statistically significant difference in instructor ratings between students rating instructors of the same sex and students rating instructors of the opposite sex – data demonstrating that sex-based biases in student evaluations of teaching are indeed present at our institution. Leadership also appreciated the opportunity peer observation presents to expand evaluation of teaching for administrative purposes (annual merit-based salary adjustments, reappointment, and promotion) beyond student perspectives gathered through end-of-semester course evaluations, though there were questions surrounding the level of effort that may be required for a formalized peer observation program, particularly for observers. This concern was addressed by agreeing to implement a small-scale pilot observing engineering faculty who are not POET participants to test the observation and feedback processes, as well as to gather information regarding the effort level for both observers and observees. Since receiving support from institutional leadership to continue developing this initiative, we have pivoted our efforts toward developing a formalized POET protocol that can be implemented at scale to expand our impact.

Anticipated results

Evaluation of the implementation and impacts of our peer observation framework is ongoing. During the Fall 2024 semester, we observed each other's courses, and two of us observed the courses of a colleague not participating in POET in preparation for their upcoming administrative review

process. POET pre-observation conversations typically focus on the instructor's learning objectives for the class meeting being observed, as well as for the semester, and what the instructor hopes to learn as a result of the observation. Typical post-observation conversations tend to focus on information exchange between the observer and the observee, with each learning from the other throughout the conversation and each gaining new insights as well as ideas for practices they may wish to incorporate into their own teaching. We believe that one of the great strengths of our observation process is the openness of the observers to learning from their observations.

We are currently working with our university's institutional review board (IRB) to develop assessment and reporting protocols that ensure re-identification risk for both students and faculty within course evaluation data is appropriately mitigated; we do not yet have IRB approval to share data. We aim to assess the POET program by collecting a variety of qualitative and quantitative data. Success will be evaluated, in part, through quantitative measures of POET activity including the number of classroom observation visits, the number of group meetings, and the number of individuals communicating interest in the program. Specific elements we are aiming to evaluate as markers of success for this initiative include:

- Are participants' senses of belonging enhanced by joining a community of faculty focused on our students' classroom experiences, thereby improving morale among participants?
- Are feelings of trustworthiness in the administrative processes reviewing faculty's teaching for reappointment and promotion enhanced?
- Do additional faculty opt to participate in POET, thereby expanding the community of faculty proactively engaged in continuous improvement of their teaching?
- Do participants opt to remain with POET across multiple academic years?
- Do observers find the peer cohorts as helpful as observees do, also enhancing their knowledge of pedagogy in practice from conversations in the group?
- Are participants' teaching practices influenced by their POET experiences?

Discussion and implications

POET has met several of its initial goals: forming a supportive community of pedagogically focused faculty, facilitating conversations about teaching approaches and methods, and providing input to teaching evaluation that complements student perspectives (counterbalancing potential student biases in course evaluations). Although the immediate impacts of POET are currently focused within the existing community of participants, we hope to influence the culture surrounding teaching and its evaluation at our institution. In its first year, POET has already made some progress with respect to shaping institutional perspectives. Most notably, institutional leadership has been receptive to this faculty-initiated effort, appreciative of the potential deleterious impacts of biases that have been shown to be present in student evaluations of teaching, and supportive of ongoing efforts to formalize observation processes that are suitable for all interested faculty.Our institution's strong tradition of faculty governance is such that we anticipate that, while faculty may be encouraged to participate in POET, it is unlikely to become a required activity. At present, we envision POET as a supportive network of experienced faculty, within which a culture of teaching observation, feedback, and discussion is normalized, that is available as a resource for all faculty, and especially junior faculty as they embark on their faculty careers.

Faculty seeking to develop a similar leadership-endorsed faculty-initiated effort at their own institutions are likely to encounter their own unique challenges and opportunities. We worked through the challenge of securing institutional leadership support by engaging in regular conversations to reach consensus on common goals and providing information that demonstrated the benefits and opportunities for both our faculty and our institution.

Appendix: Guided observation prompts

1. What are your first impressions?
2. How is the class structured? Are students expected to be prepared with/for anything?
3. Positive comments about the structure/ presentation/ overall classroom experience.
4. Other feedback about class structure.
5. What did you see that you would be interested in trying in your own class?
6. What is working well?
7. What do you want to ask the instructor? (*i.e.*, why did you do it this way? How would you have responded if xyz happened instead today?)
8. What class content was particularly interesting/engaging?
9. Other comments or questions.

References

[1] Resources for teaching evaluation guides. https://tinyurl.com/5ah3ehc3, 2021. Accessed on June 3, 2024.

[2] American Sociological Association. Statement on student evaluations of teaching. www.asanet.org/studentevaluations, 2020. Accessed on February 9, 2025.

[3] T. C. Andrews, P. Brickman, E. L. Dolan, and P. P. Lemons. Every tool in the toolbox: Pursuing multilevel institutional change in the DeLTA project. *Change: The Magazine of Higher Learning*, 53(2):25–32, 2021.

[4] O. R. Aragón, E. S. Pietri, and B. A. Powell. Gender bias in teaching evaluations: The causal role of department gender composition. *PNAS Psychological and Cognitive Sciences*, 120(4):e2118466120, 2023.

[5] A. Boring, K. Ottoboni, and P. B. Stark. Student evaluations of teaching (mostly) do not measure teaching effectiveness. *ScienceOpen Research*, 2016.

[6] D.E. Clayson, T.F. Frost, and M.J. Sheffet. Grades and the student evaluation of instruction: A test of the reciprocity effect. *Academy of Management Learning and Education*, 5:52–85, 2006.

[7] P. Deraney. "because more trust now": The role of peer observation of teaching in building a faculty community of practice. *Qualitative Research in Education*, 11(3):270–297, 2022.

[8] T. Heffernan. Sexism, racism, prejudice, and bias: A literature review and synthesis of research surrounding student evaluations of courses and teaching. *Assessment & Evaluation in Higher Education*, 47(1):144–154, 2022.

[9] G. D. Hendry, H. Georgiou, H. Lloyd, V. Tzioumis, S. Herkes, and M. D. Sharma. 'it's hard to grow when you're stuck on your own': Enhancing teaching through a peer observation and review of teaching program. *International Journal for Academic Development*, 26(1):54–68, 2021.

[10] S. Krishnan, J. Gehrtz, P.P. Lemons, E.L. Dolan, P. Brickman, and T.C. Andrews. Guides to advance teaching evaluation (GATEs): A resource for STEM departments planning robust and equitable evaluation practices. *CBE—Life Sciences Education*, 21(3):ar42, 2022.

[11] J. R. Lackritz. Exploring burnout among university faculty: Incidence, performance, and demographic issues. *Teaching and Teacher Education*, 20:713–729, 2004.

[12] L. Lowder, M. Atiqualla, D. Colebeck, S. Das, M. A. Karim, A. Khalid, R. Singh, and T. Utshig. Peer observation: Improvement of teaching effectiveness through class participation at a polytechnic university. *Journal of STEM Education: Innovations and Research*, 18(4):51–55, 2017.

[13] L. MacNell, A. Driscoll, and A.N. Hunt. What's in a name: Exposing gender bias in student ratings of teaching. *Innovative Higher Education*, 40:291–303, 2015.

[14] E. Martin. Power and authority in the classroom: Sexist stereotypes in teaching evaluations. *Journal of Women in Culture and Society*, 9:482–492, 1984.

[15] M. Millea and P.W. Grimes. Grade expectations and student evaluation of teaching. *College Student Journal*, 36:582–590, 2002.

[16] L.D. Reid. The role of perceived race and gender in the evaluation of college teaching on ratemyprofessors.com. *Diversity in Higher Education*, 3(3):137–152, 2010.

[17] B. Schmidt. Rate my professor visualization. https://benschmidt.org/2015/02/06/rate-my-professor/, 2015. Accessed on June 3, 2024.

[18] D.D. Shorter. Teaching evaluations are racist and sexist: It's time to put these flawed measures in their place. *The Chronicle of Higher Education*, 69(24):42+, 2023.

[19] S. Shortland. Feedback within peer observation: Continuing professional development and unexpected consequences. *Innovations in Education and Teaching International*, 47(3):295–304, 2010.

[20] M. K. Smith, F. H. M. Jones, S. L. Gilbert, and C. E. Wieman. The classroom observation protocol for undergraduate STEM (COPUS): A new instrument to characterize university stem classroom practices. *CBE: Life Sciences Education*, 12(4):25–32, 2013.

[21] W. Stroebe. Why good teaching evaluations may reward bad teaching: On grade inflation and other unintended consequences of student evaluations. *Perspectives on Psychological Science*, 11(6):800–816, 2016.

[22] N.C. Subtirelu. "she does have an accent but...": Race and language ideology in students' evaluations of mathematics instructors on ratemyprofessors.com. *Language in Society*, 44(1):35–62, 2015.