Summary Version: Potential Biases in Peer Reviews of Teaching
Center for Teaching and Learning

Jordan D. Troisi, Ph.D., Senior Associate Director of the CTL
jdtroisi@colby.edu

Background

During the Chairs and Program Directors meeting on October 1, 2020, I presented and hosted a short discussion on the value of peer review of teaching. During the Q/A period of that discussion, someone raised a question about the special challenges of faculty of color in tenure review processes. **This document is intended to extend that conversation and provide information about how biases may be present in evaluations of teaching** (especially pertaining to student evaluations of teaching, or SETs, and peer reviews of teaching).

At the outset, let me identify 3 important pieces of information:
1. I know of **no scientific studies** that have systematically and rigorously examined peer review of teaching. Thus, peer reviews of teaching may or may not be (a) valid measures of teaching effectiveness, (b) biased based on aspects of teacher identity (e.g., racial background, gender, sexual orientation, age).
2. Though there is a lot of discussion in the popular academic media on how teacher identity may be associated with student evaluations of teaching (SETs), **there is much more gray area and nuance in the research base than the headlines make it appear** (see references below). This is relevant if we wish to "borrow" ideas from how biases may influence peer review of teaching.
3. Based on my review of the research on biases in SETs, my own reflection is that bias has at least a small effect on SETs. There **may be parallel effects in peer reviews of teaching**.

Note: most of these ideas are inferential, and based on my understanding of social psychological processes associated with the major components of explicit and implicit attitudes (i.e., affective, behavioral, and cognitive components). More evidence is needed to confirm the ideas below.

How Peer Reviews of Teaching **might introduce bias** based on teacher identity	How Peer Reviews of Teaching **might refute/reduce bias** based on teacher identity
Peer reviewers view biased behavior in the classroom and have that inform their evaluation	←Peer reviewers, anticipating student bias in classroom behavior, might attempt to "correct" for it and construe the teaching to be more effective. Though this mitigates concerns associated with teacher identity, it is likely still a biased view of teaching
Peer reviewers, having seen course evaluations of the teacher, may perceive teaching in a biased way (this is a "confirmation bias" effect)	←Peer reviewers, anticipating student bias in course evaluations, might attempt to "correct" for it and construe the teaching to be more effective. Though this mitigates concerns associated with teacher identity, it is likely still a biased view of teaching
Peer reviewers have relationships with the teachers under evaluation. When these relationships are favorable, evaluations will likely be positive.	←Peer reviewers have relationships with the teachers under evaluation. When these relationships are unfavorable, evaluations will likely be negative. Of note, identity-based enclaves often exist, leading to friendship groups that are homogenous. The identity make-up of these groups may vary based on seniority and tenure status.
Peer reviewers may have specific ideas about teaching related to their own sub-discipline within their field.	Peer reviewers may seek to "correct" for the potential of bias in their overall evaluation of the faculty member under review.
Peer reviewers may have a significant preference for their own teaching methods. Thus, they may be biased based on "teaching approach," rather than people's identity.	People generally enact more biased judgments when they have less information to draw from. Pre-observation meetings with peer reviewers can establish the goals of the instructor, and thus, mitigate opportunities for bias. They leave less "open to interpretation" or inference.
Peer reviewers may harbor implicit or explicit biases toward the identity groups of the teacher. Of note, just like how students do not perceive themselves to be	Peer reviewers who adhere more to agreed upon observation forms and checklists should demonstrate less bias. These approaches leave less "open to interpretation" or inference.

biased, faculty members also likely do not perceive themselves to be biased (this is referred to as the "bias blind spot" phenomenon).	

Other recommendations, drawing from research on SETs

- Many factors lead to more or less favorable reviews of teachers: class size, student motivation, student effort, the discipline in question, etc. Peer reviewers should keep this in mind. Faculty members compiling their dossiers should keep these in mind also.
- It is likely that peer observations of teaching, like student evaluations of teaching, are not very strong predictors of student learning (which could be considered the "gold standard" of teaching effectiveness). One way of getting at direct measures of student learning would be to require standardized tests of student learning for all instructors as a measure of teacher performance, but I imagine there little appetite for this. Instead, because all measures are imperfect, perhaps the best approach is to "use as many of them as we can--recognizing that each has its own unique flaws" (see the summary at the end here).
- Quantitative scores on student ratings of teaching can be valuable when examining individual instructor growth over time. The same could be true of peer observations, but this would require multiple measurements over time, something that is not usually in place.

References/summaries of findings on potential bias in SETs

- https://docs.google.com/document/d/1t38eBhXHpy2PmeIj9F58Uqy5IYCJwCEbBwT4ZRyCLv8/edit?usp=sharing
- https://crlt.umich.edu/tstrategies/studentratingfaq
- https://cte.rice.edu/blogarchive/2015/07/09/studentevaluations
- https://cte.rice.edu/blogarchive/2018/2/20/studentratingsupdate

Gleason, Nancy W. and Sanger, Catherine S., "Guidelines for Peer Observation of Teaching: A Sourcebook for International Liberal Arts Learning" Centre for Teaching and Learning, Yale-NUS College, Singapore (September 2017)

Managing Implicit Bias

The Problem of Implicit Bias in Faculty Assessment

"Implicit bias refers to the attitudes or stereotypes that unconsciously affect our understanding, actions, and decisions. These biases include both favourable and unfavourable assessments, and are activated without one's awareness or intentional control. Individuals therefore do not consciously hide these biases, and they also do not necessarily

align with our declared beliefs. Instead, implicit biases are inaccessible to one's conscious mind. Early life experiences and the media are common origins of implicit associations."[6]

We call attention to implicit bias in this document not to suggest that faculty would intentionally observe each other or write about each other's teaching in a biased way. The term *implicit* bias is powerful because it highlights the fact that we are all vulnerable to unconsciously held ways of thinking and interpreting each other's behaviours. Having unconsciously held attitudes, heuristics, and even stereotypes does not make us bad people – it makes us people. But despite these ingrained tendencies, we have the capacity to bring more nuanced thinking to the fore when we need to. No one enters into a peer observation process intending to act from bias. But it can happen if we do not take the time to reflect on possible biases we hold and check in with ourselves during the peer observation process to make sure we are not describing or judging our colleague's teaching through a biased lens.

One of the reasons to add peer observations into our review and promotion process is to complement student teaching evaluations, which are frequently tainted by implicit bias. For example, in an experiment in which female assistant instructors in an online class each operated under two different gender identities, students consistently gave perceived male instructors higher ratings.[7] Other studies conducted in the US showed that racial minority faculty members receive lower teaching evaluations than do majority instructors, and non-native English speakers receive substantially lower ratings than do native English speakers. This is not a US-specific problem. Here in Singapore, there are concerns that local students continue to hold favourable attitudes towards white professors due more to their race than teaching skill.[8]

Implicit bias is not just a problem in student evaluations. ***Faculty can also act from implicit bias in their evaluations of each other***. This is documented in research on academic hiring and promotion practices, showing in particular that women and racial minorities can be subject to misperceptions that they are less competent.[9] For example, one study found that external reviewers and research committee members displayed an unconscious gender bias when reviewing curricula vitae to hire job applicants. Notably, in this study both men and women were more likely to vote to hire a male job applicant than a female job applicant with an identical record. Readers of both sexes reported that the male job applicant had adequate teaching, research, and service experience, whereas the female job

---

[6] Kirwan Institute for the Study of Race and Ethnicity, *2015–2016 Biennial Report* (Columbus: Ohio State University, 2017).

[7] MacNell, Lillian, Adam Driscoll, and Andrea N. Hun, "What's in a Name: Exposing Gender Bias in Student Ratings of Teaching," *Innovative Higher Education* 40, no.4 (2015): 291–303; Boring, Anne, "Gender Biases In Student Evaluations Of Teaching," *Journal of Public Economics* 145, (2017): 27–41; Wagner, Natascha, Matthias Rieger, and Katherine Voorvelt, "Gender, Ethnicity and Teaching Evaluations: Evidence From Mixed Teaching Teams," *Economics of Education Review* 54, (2016): 79–94.

[8] "Research suggests students are biased against female lecturers," *The Economist,* September 21, 2017, accessed February 26, 2018, from https://www.economist.com/news/science-and-technology/21729426-how-long-does-prejudice-last-research-suggests-students-are-biased-against; Hamermesh, Daniel S., and Amy Parker, "Beauty in the classroom: instructors' pulchritude and putative pedagogical productivity," *Economics of Education Review* 24, no. 4 (2005): 369–376; Lim, Linda, "Asian Students Still Idolize Western Professors," *Wall Street Journal*, July 27, 2017.

[9] For a review of research on gendered bias in academia, from hiring trends to treatment by colleagues to promotion – including how bias impacts recommendation letters, how we read cover letters and curriculum vitae, and how we assess teaching, research, and service – see Chenoweth, Erica, Page Fortna, Sara Mitchell, Burcu Savun, Jessica Weeks, and Kathleen Cunningham, "How to Get Tenure (If You're a Woman)," *Foreign Policy,* April 18, 2016, accessed February 26, 2018, from www.foreignpolicy.com/2016/04/19/how-to-get-tenure-if-youre-a-woman-academia-stephen-walt/; Trix, Frances, and Carolyn Psenka, "Exploring the color of glass: letters of recommendation for female and male medical faculty," *Discourse & Society* 14, no. 2 (2003): 191–220; Phelan, Julie E., Corinne A. Moss-Racusin, and Laurie A. Rudman, "Competent yet out in the cold: shifting criteria for hiring reflect backlash towards agentic women," *Psychology of Women Quarterly* 32, no. 4 (2008): 406–413; Maliniak, Daniel, Ryan Powers, and Barbara F. Walter, "The Gender Citation Gap in International Relations," *International Organization* 67, no. 4 (2013): 889–922; Shaw, Claire, "Stories of sexism in science: 'sorry about all the women in this laboratory'," *The Guardian,* June 12, 2015, accessed February 26, 2018, from https://www.theguardian.com/higher-education-network/2015/jun/12/stories-of-sexism-in-science-sorry-about-all-the-women-in-this-laboratory.

applicant with an identical record received less favourable assessment in these areas.[10] In another piece of refereed research, evaluators perceived competent women in clear leadership roles as less likeable, and colleagues were less likely to recommend them for hiring or promotion.[11] Another analysis, conducted on over 300 letters of recommendation for medical school faculty, "found that letters about women tend to be shorter and focus on personality traits. Letters about men typically feature superlatives about their aptitudes and abilities, and letters for men contained more standout adjectives, such as *superb, outstanding*, and *remarkable.*"[12] Another study has shown that while men are tenured at roughly the same rate regardless of whether they co-author or solo-author, women become less likely to receive tenure the more they co-author.[13]

While these examples have focused on biased evaluation of female faculty, faculty from traditionally marginalised ethnic groups as well as gender non-conforming, overweight, and less able-bodied faculty may be the subject of biased evaluation. Ageism, which can lead to biased assessments of younger and older faculty alike, is also something to guard against. Importantly, research shows that faculty often act from bias even when evaluating peers within their same identity groups (e.g. an Asian female observing an Asian female colleague can still interpret behaviour through biased gendered and racial frames.)

Put differently, we all need to confront and manage our implicit biases when conducting peer observations, and any other professional responsibility.

Minimizing Implicit Bias in Peer Observation

Though implicit biases are powerful and difficult to overcome, there are steps you can take to gain greater awareness of your own implicit biases and mitigate their effects. With effort, you can grow more aware of your own implicit biases and work to dismantle them.

First, before conducting an observation, ***take some time to reflect*** upon possible implicit biases, especially ones that will surface when observing this particular peer.
   a. Ask yourself difficult questions about the judgements you make about others. Do you find yourself more comfortable with some groups of people over others? What are you assuming about a person or group of people? Are those judgements based on stereotypes?
   b. Ask people you trust if they have noticed any particular biases. Have an open conversation about your patterns of behaviour with respect to different gender, racial, socio-economic, and other identities.

---

[10] Steinpreis, Rhea E., Katie A. Anders, and Dawn Ritzke, "The Impact of Gender on the Review of the Curricula Vitae of Job Applicants and Tenure Candidates: A National Empirical Study," *Sex Roles* 41, no. 7–8 (1999): 509–528.

[11] Eagly, A. H., and S. J. Karau. "Role congruity theory of prejudice toward female leaders," *Psychological Review* 109, no. 3 (2002): 573–598; Ridgeway, Cecilia L., "Gender, Status, and Leadership," *Journal of Social Issues* 57, no. 4 (2001): 637–655; Heilman, Madeline E., "The impact of situational factors on personnel decisions concerning women: Varying the se composition of the applicant pool," *Organizational Behavior and Human Performance* 26, no. 3 (1980): 386–395; Rutgers School of Arts and Sciences, "Implicit Bias," accessed February 26, 2018, from https://www.philosophy.rutgers.edu/graduate-program/climate/133-graduate/climate/529-climate-of-women-implicit-bias.

[12] Trix and Psenka, "Exploring the color of glass." See similar results in an even larger study by Madera, J. M., M. R. Hebl, and R. C. Martin, "Gender and Letters of Recommendation for Academia: Agentic and Communal Differences," *Journal of Applied Psychology* 94, no. 6 (2009): 1591–1599.

[13] While an additional publication is correlated with a roughly 4 percentage point increase in tenure probability for both men and women, women are consistently 17 percentage points less likely to receive tenure than men conditional on having written the same number of papers of similar quality. See Sarsons, Heather, "Recognition for Group Work," *American Economic Review* 107, no. 5 (2017): 141–145.

c. Take the Implicit Association Test, which measures attitudes and beliefs that people may be unaware they hold. There are several different tests to choose from. (https://implicit.harvard.edu/implicit/education.html)

Second, think about past actions and *how your biases may have hindered opportunity for others,* even if unintentionally*.* Acknowledging the impact of your biases on others can motivate you to be more self-aware and to control your biases in the future.

Third, *check in with yourself* while conducting an observation. Every so often ask, "Would I interpret this the same way if the professor represented another gender identity, race, age, language group, etc.?"

Note that *bias shows up not only in negative observations, but also positive ones*. In writing summative observation letters, avoid references to how hard a colleague works, inclusion of irrelevant personal information, and ambiguous praise.*[14]* For example, someone might observe an impressive female faculty member and write "Her effort and dedication is all the more impressive when you consider she has kids at home!" This may seem like praise, but ask yourself if you would write the same thing if the colleague was a male professor? Similarly, in observing faculty whose primary language is not the language of instruction, one might make a comment like "the professor goes to great lengths to be understood." This is meant to be positive, but has the impact of implying that the professor is difficult to understand. A more useful approach would be to acknowledge openly that this faculty's primary language is not English, but offer specific details on how the faculty member has developed strategies to ensure students understand what is being taught despite their accent (e.g. noting the complementary use of white boards, slides, handouts, etc.) or if appropriate note that this is an area which may need some improvement.

Lastly, **be specific, and be consistent in standards you apply** across the peers you observe. Minimizing the impact of implicit bias is one reason to use a standardised, criteria-based approach in all observations. *The Criteria Organiser available at the end of this document can help you evaluate specifically identified criteria and curb implicit bias.*

---

[14] "A study of over 300 recommendation letters for medical faculty hired by a large U.S. medical school found that letters for female applicants differed systematically from those for males. Letters written for women were shorter, provided "minimal assurance" rather than solid recommendation, raised more doubts, portrayed women as students and teachers while portraying men as researchers and professionals, and more frequently mentioned women's personal lives." Quoted from Rutgers School of Arts and Sciences, "Implicit Bias."