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 there is likely 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/1t38eBhXHpy2Pmelj9F58Uqy5IYCJwCEbBwT4ZRyCLv8/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