

Further research on gender bias in course evaluation data

Emily Greenleaf, PhD (emily.greenleaf@utoronto.ca) & Pamela Gravestock, PhD (p.gravestock@utoronto.ca) – Revised April 2017

In the nearly 10 years since we completed [our survey of course evaluation research](#) and practices, we have continued to work closely with course evaluations and the assessment of teaching. We've also continued to review new research in this area. We'd like to share our updated thinking on why it has been difficult to demonstrate consistent statistical evidence of gender bias (or other forms of bias) in studies of course evaluations, even though there are many examples of cases where student responses are influenced by gender.

Limited information from statistical analyses

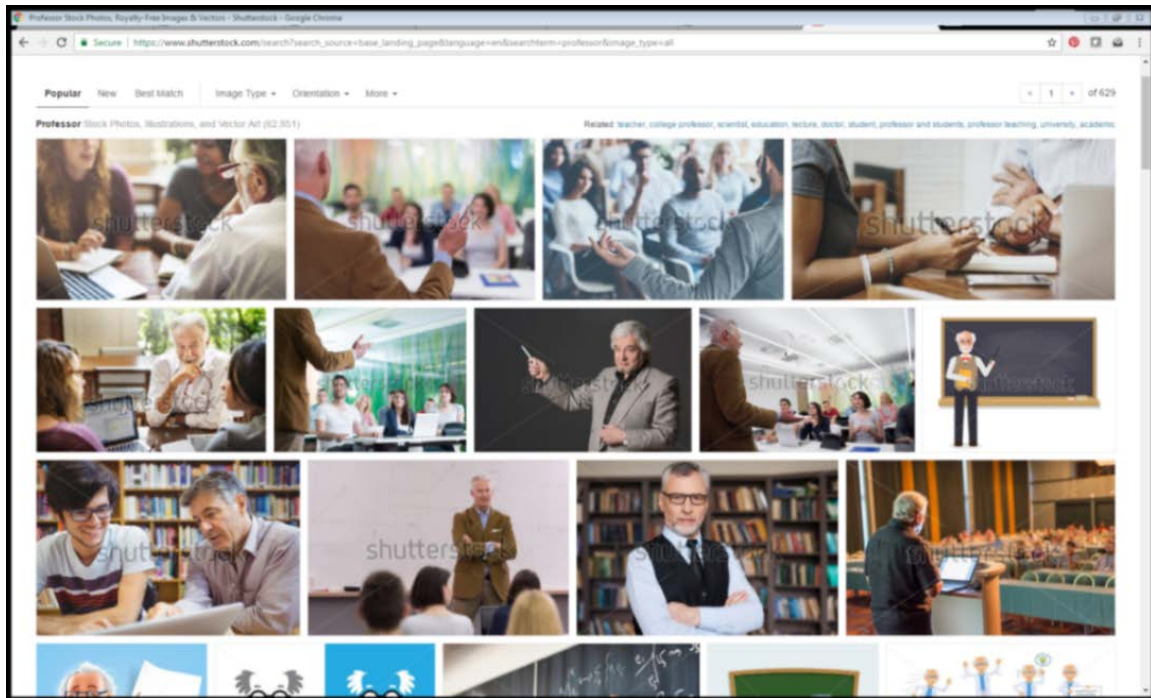
In 2008, we noted that, for quantitative evaluation questions, “studies relating to gender have produced inconclusive results, but most have shown that this variable has little or no impact on evaluations.” This finding has been reiterated in several more recent overviews of course evaluation research ([Benton & Cashin, 2014](#); [Hativa, 2014](#), [DeFrain, 2016](#)). However, while overall statistical analysis only sometimes demonstrates gender bias, evidence of gender bias and other forms of bias related to instructor characteristics has been found in multiple studies ([Arbuckle, J. & Williams, 2003](#); [Boring & Stark, 2016](#); [Boring, 2017](#); [Centra & Gaubatz, 2000](#); [MacNell, Driscoll, & Hunt, 2015](#); [Miller & Chamberlin, 2000](#); [Sprague & Massoni, 2005](#)). We'd like to discuss this ambiguity, and suggest that a lack of consistent gender bias in statistical analysis does not preclude the presence of gender bias in student responses, and should not be taken to indicate that all questions or instruments¹ are free of bias.

We can draw some broad conclusions about bias based on gender and other instructor characteristics from a collective analysis of studies of this particular issue (a number of which are cited above) alongside our more general and ongoing course evaluations research.

Some questions clearly elicit bias

Some instruments inappropriately ask students to provide feedback in areas they can't accurately assess. This might include, for example, questions about instructor knowledge of the subject area or about the quality of course content; students, by definition non-experts, can't make this assessment. Because students do not have sufficient information to respond accurately to such questions, they make a guess, and that guess is shaped by their existing assumptions about what kind of instructor *is*, for example, knowledgeable about the subject area. Such activation of implicit biases will favour instructors who best reflect common conceptions of “a professor.” For an example of what “a professor” might look like, take a look at what a search for “professor” in a stock photo database returns:

¹ The conclusions of studies of a single instrument at a single institution can point to important considerations for instrument development, use and interpretation, but can rarely be extrapolated to draw conclusions about course evaluations in general. Course evaluations – or any instruments designed to collect student feedback on courses, their learning experience, or their instructors – represent a constellation of unique examples, each with their own potentials for bias, misuse, and misinterpretation as well as the potential to gather useful feedback from students.



Such questions, therefore, can lead to clear and direct evidence of gender biases (see, e.g., the results in the research from Boring, 2017); many of the studies that demonstrate strong evidence of gender bias in evaluation scores are assessing instruments that incorporate this type of question.

More complexity: Questions about instructor characteristics and behaviours

Similarly, questions that ask students to rate particular instructor characteristics or behaviours (e.g. “The instructor was enthusiastic,” “The instructor explained concepts clearly”) activate comparisons to the ways in which students expect their instructors to express those characteristics or carry out those behaviours (Miller & Chamberlin, 2000). These expectations will be informed both by the expected characteristics of “a professor,” and the expected modes of teaching that the student associates with other groups the instructor might belong to – age, gender, race, culture, etc.

Given the complex and multifaceted associations at play, questions that ask students to rate particular instructor characteristics or behaviours may be responsible for the sometimes ambiguous or contradictory evidence about bias.

For example, a young woman instructor might conform to gender expectations (e.g. being very nurturing) but not to a student’s mental image of “a professor”; in response to an item that asks, for example, whether “My instructor cares about my learning,” that instructor may be rewarded for conforming to some norms but penalized for deviating from others. These expectations also vary by student, and a number of studies suggest an interaction between a student’s own gender or other

identities and instructor gender or other identities (Boring, 2017; Centra & Gaubatz, 2000, [Das & Das, 2001](#)) – again creating complexities that might interact to obscure a range of influences and biases.²

Conclusions and recommendations

Our overall conclusion from this research is that biases are likely to be activated by questions that ask students to provide feedback on information they can't accurately assess, and questions that ask students to rate particular instructor characteristics or behaviours. In some cases, these biases may pull ratings in multiple directions for any given pairing of instructor and student and therefore may not always be immediately evident in numerical results. Given the demonstrated or theoretical potential for bias, such questions should not be included on evaluation instruments.

With that understanding in mind, we have advocated that questions on course evaluations should be phrased and understood as an assessment of how students experience their learning (as opposed to a direct assessment of an instructor's teaching). As an example, instead of asking students to rate the degree to which "The instructor stimulates my interest in this subject," students might be asked to reflect on the degree to which "I found this course intellectually stimulating."

There are several important reasons to focus evaluations on the student learning experience,³ and one is that it may help to avoid some of the most direct and obvious sources of bias. Learning experience questions ask the student to reflect and report on their *own* experience, shifting the focus of assessment away from instructor behaviours or characteristics. Additionally, the wording of the questions reflects (and, ideally, communicates - to student responders as well as to administrators) that the data collected do not reflect an objective measure of an instructor's performance but rather a subjective, relational assessment of how an instructor has shaped student learning within the context of a course.⁴ In recent years, our conclusions have been reiterated by other researchers who similarly argue that evaluations are best suited to measure a student's experience of their own learning (see, for example, [Stark & Frieshtat, 2014](#)).

² Not to mention the range of influences that might further introduce complexity into an attempt to compare evaluation scores across courses or units, including differences in class size, level, instructional approach, or discipline – which themselves might be gendered because of differences in teaching assignments, gender balance in the discipline, etc.

³ By focusing on outcome (i.e. how the student experiences the teaching) rather than input (i.e. how the instructor performs the teaching), a student-focused approach also allows instructors to use a range of teaching approaches to foster student learning. Given that no universal set of "effective" teaching approaches exists, such an approach allows for flexibility in the teaching approaches understood as effective or desired within a particular disciplinary context.

⁴ Indeed, we refer to these instruments as "course evaluations" because this terminology reflects a foundational conclusion about the use of student feedback about courses and instructors: For students, the meaningful unit for feedback is the course. Students may not make a meaningful distinction – assuming there is one to be made – between an instructor's approach to teaching or classroom instruction and the broader learning context/experience of the course.

Given that evaluations provide this important but limited insight into how students experience an instructor's teaching, course evaluations should only be used as one part of a broader assessment framework that provides additional lenses on an instructor's teaching activities.

Need for analysis in the UofT context

Our assumption is that the institutional items on the UofT instrument, with their focus on students' experiences of their own learning, mitigate the most direct sources of bias. UofT is almost unique in having a course evaluation framework that focuses on the student learning experience, although similar frameworks are now being adopted at a number of other institutions. Because we have such an instrument, and because it has been in place for a few years, the institution now has the opportunity to assess the strength of this approach in practice within our institutional context. This will provide UofT instructors and administrators with the most meaningful data about how gender or other factors affect evaluation scores and comments.

Finally, although our expectation is that a learning experience-focused instrument represents an improvement on traditional instruments, all evaluation instruments will be shaped by broader social norms and constructs. Students' expectations for and assumptions about instructors will also affect how they experience their learning, and institutions must be attentive to the influence of these expectations and assumptions on quantitative and qualitative evaluation data. This again emphasizes the importance of interpreting students' evaluation responses within the context of a multi-faceted framework for the assessment of teaching, and with an understanding that what course evaluations offer is insight into a student's experience of their own learning.