Howard R. Mzumara, Ph.D.,

Director of Evaluation and Psychometric Services

Office of Institutional Research and Decision Support (IRDS)

Indiana University – Purdue University Indianapolis (IUPUI)

Revisiting Course Evaluations: Strategies to Minimize Gender and Racial Biases in Student Evaluations of Teaching

Presentation given at the Bluenotes GLOBAL 2019 Conference, Chicago, IL

August 6, 2019

Outline

- Objectives of this Session

- Quick Facts and Data about IUPUI

- Gender/Racial Biases in Student Evaluations of Teaching (SET)

- Potential Sources of Biases in SET

- Strategies to Mitigate Biases in SET

- Tips to Optimize the Utility of Course Evaluation Data

  ➢ Using Student Feedback More Effectively for *Formative* and *Summative Evaluation*

Objectives

1. Increase awareness of growing evidence of gender or racial bias in course evaluations.

2. Provide illustrative examples of a growing body of literature examining gender and/or racial biases in course evaluations.

3. Identify potential sources of gender or racial bias in student evaluations of teaching.

4. Discuss strategies to mitigate implicit/explicit bias that may exist in course evaluations.

IUPUI: Quick Facts and Data

- Indiana University-Purdue University Indianapolis; Established in 1969

- ~30,000 students [22,000 undergraduate & 8,000 graduate students from 146 countries and all 50 states in USA; 25% minorities; 56% female, 44% male]

- 17 distinct schools and 2 colleges that confer degrees thru IU & PU
  - 2 Purdue schools (Engineering & Technology, Science)
  - 15 IU schools (Business, Dentistry, Education, Herron Arts & Design, Health & Human Sciences, Informatics & Computing, Law, Liberal Arts, Medicine, Nursing, Philanthropy, Public & Environmental Affairs, Public Health, Social Work, University Graduate School; and
  - 2 Colleges (Honors College, University College)

- Offers more than 350 undergraduate, graduate, and professional programs

- Strong research focus

- Destination campus for health and life sciences

IUPUI – BLUE Users
Fall 2018

Total # of Academic Units Using BLUE Course Evaluations: 15
Total # of Courses marked "Evaluate": ~ 4,547 out of 7,782 courses
(Decentralized process ... with a wide variety of instruments in use)

- Business (Indianapolis)
- Education
- Engineering and Technology
- Health and Human Sciences
- Herron Art and Design
- Informatics and Computing
- Law
- Liberal Arts

- Medicine (Radiology and Imaging Sci.; Pathology & Lab Medicine Departments)
- Philanthropy
- Public Health
- Science
- Social Work
- Public & Environmental Affairs (via Blue Project administered at IUB)
- University College

Brief Overview of Gender/Racial Bias in SET

Key questions: Are student evaluations of teaching (SET) biased against women faculty, and what are the implications of this bias? (Why this matters…)

❖ Recent research studies (e.g., Boring, 2017; MacNell, Driscoll, & Hunt, 2015; Martin, 2016, Mitchell & Martin, 2018; Peterson, Biederman, *et al.*, 2019; Rosen, 2017; etc.) provide evidence of gender bias in course evaluations. Specifically, research suggests that:

   o **women professors obtain lower ratings** than male professors.

   o **women faculty are judged more critically** (or on a **different set of criteria)** than their men counterparts in (at least two ways):

      ➤ qualification/competence and personality

   o "Minority faculty obtain significantly lower ratings than white professors, even after controlling for tenure status and course type" [e.g., see research articles by Chisadza, Nicholls, & Yitbarek, 2019; Merritt, 2008; Reid, 2010; Smith & Hawkins, 2011; etc.]

Implications for Gender/Racial Bias in SET

Consider the implications for using student feedback for **Summative Evaluation** Purposes:

➢ If, as the mounting empirical evidence suggests, course evaluations are biased against women, then the use of evaluations to make personnel decisions (such as hiring, awards, promotion and tenure) might be **discriminatory** or yield **negative outcomes**.

➢ Also, consider potential bias in **race**, **ethnicity**, and **English proficiency of faculty**…

Potential Sources of Bias in SET

❖ Use of global, generic, or ambiguous questionnaire items that are **unspecific** and not linked to any particular instructor behavior or specific course elements, such as:

- **"Overall, how do you rate this instructor?"**

- **"Overall, this is an excellent instructor."**

- **"Overall, this is an outstanding professor."**

- **"What is the overall rating of the instructor's teaching effectiveness?"**

- **"Overall, how do you rate this course?"**

- **"Overall, this is an excellent course."**

- **"What is your overall rating of this course?"**

Experimental Research Study on Mitigating Gender Bias in SET
(Ref: Peterson, Biederman, Andersen, Ditonto, & Roe, 2019)

❖ The article (published in *PLOS ONE*) cites experimental research showing that **gender bias accounts for up to a 0.5-point negative effect for women faculty** (based on a five-point response scale)

❖ Experimental research by Peterson, Biederman, *et al.* (2019) **investigated the impact of limiting the problem of biases in SET by** *"cuing students to be aware of their biases, providing motivation to not rely on them, and suggesting alternatives to their stereotypes."*

❖ Among other questionnaire items, each student involved in the study was asked the following questions about their **instructor**, on a five-point scale:

- **Your overall rating of this instructor is?**

- **What is your overall rating of the instructor's teaching effectiveness?**

- **And your overall rating of this course is?**

Text Added to Course Evaluations*

"Student evaluations of teaching play an important role in the review of faculty. Your opinions influence the review of instructors that takes place every year. <*The University*> recognizes that student evaluations of teaching are often influenced by students' **unconscious** and **unintentional** biases about the race and gender of the instructor. Women and instructors of color are systematically rated lower in their teaching evaluations than white men, even when there are no actual differences in the instruction or in what students have learned.

As you fill out the course evaluation please keep this in mind and make an effort to resist stereotypes about professors. Focus on your opinions about the content of the course (the assignments, the textbook, the in-class material) and not unrelated matters (the instructor's appearance)."

*Text adopted from research paper by Peterson, Biederman, Andersen, Ditonto, & Roe (2019, May 15) (https://doi.org/10.1371/journal.pone.0216241)

Gender Bias Research Study Finding
by Peterson, Biederman, *et al*. (2019)

When students were reminded to be aware of bias, their **evaluations of female instructors** were:

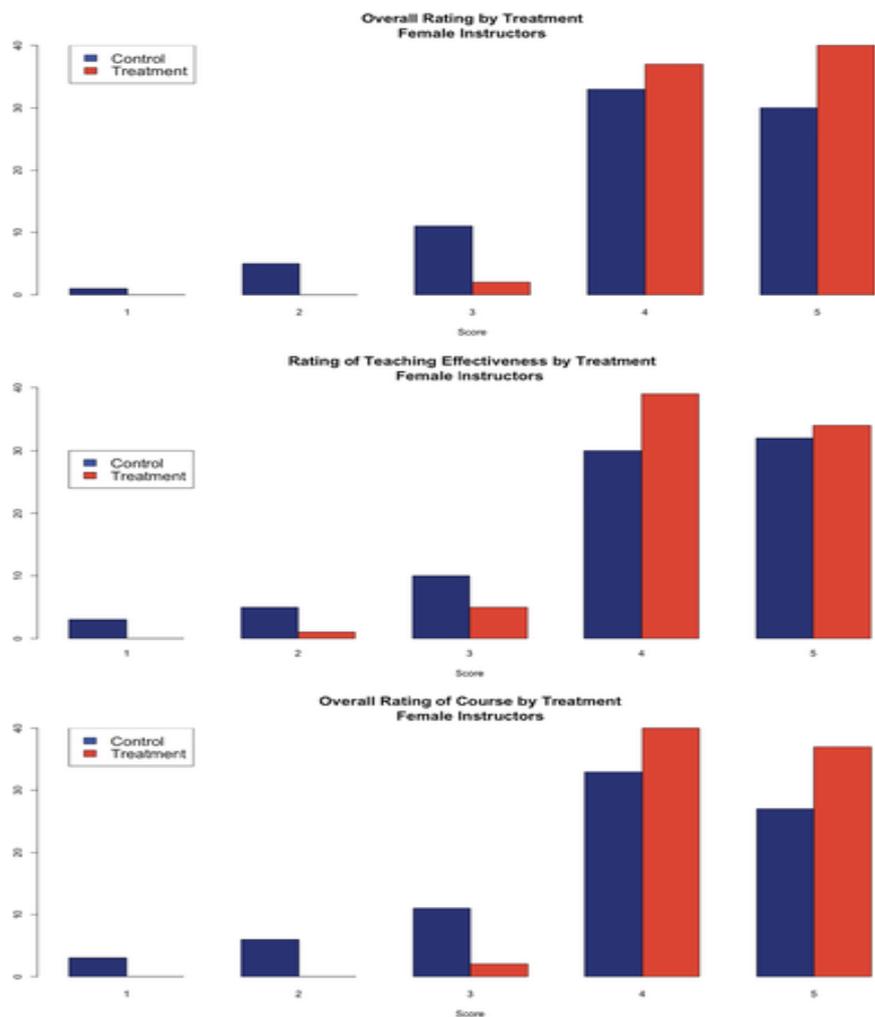**0.41 points higher** as an **overall evaluation**

0.30 points higher for **teaching effectiveness**

0.51 points higher for the **overall course evaluation**

A Note of Caution: The authors are somewhat **uncertain about the broad applicability** of the study results. … Hence, **further research is needed**.
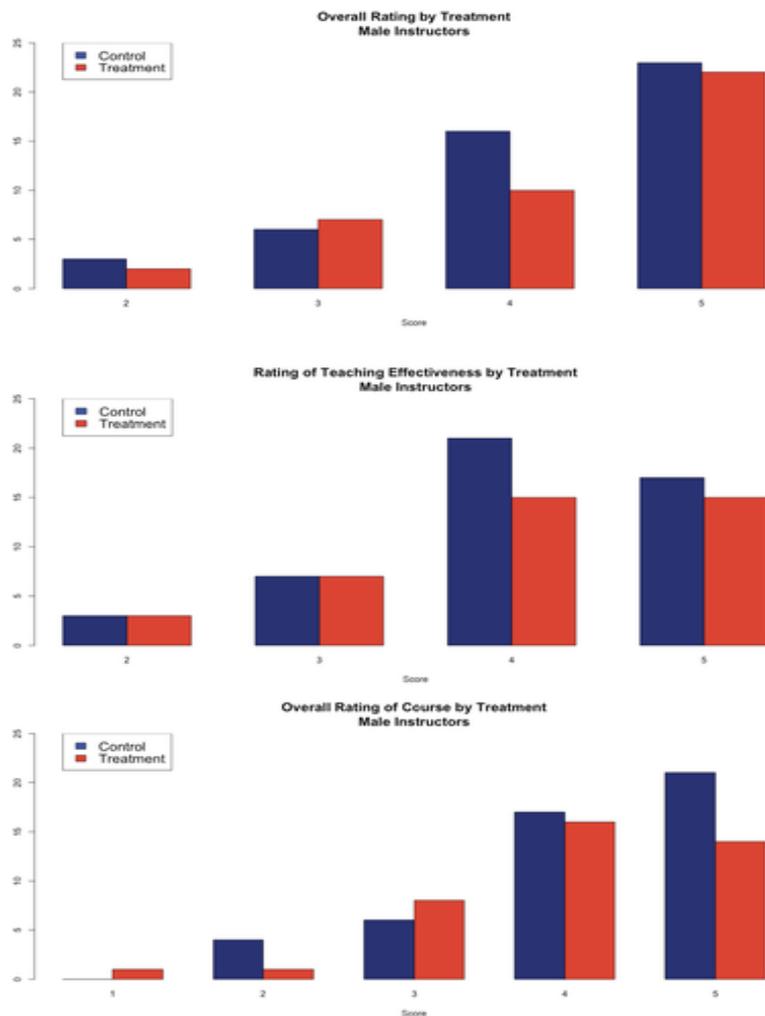
Fig 1. Analysis of student evaluations of teaching by experimental condition (female faculty)*.

Overall Rating by Treatment Female Instructors

Rating of Teaching Effectiveness by Treatment Female Instructors

Overall Rating of Course by Treatment Female Instructors

A. Student's **overall rating** of the **instructor**. Higher values are more positive ratings.

B. Student's rating of the **instructor's teaching effectiveness**. Higher values are more positive ratings.

C. Student's **overall rating** of the **course**. Higher values are more positive ratings.

- For each panel, the **left (blue) bar** are the students in the **control** condition and the **right (red) bar** are students in the **treatment** condition.

Fig 2. Analysis of student evaluations of teaching by experimental condition (male faculty)*.

A. Student's **overall rating** of the **instructor**. Higher values are more positive ratings.

B. Student's **rating** of the **instructor's teaching effectiveness**. Higher values are more positive ratings.

C. Student's **overall rating** of the **course**. Higher values are more positive ratings.

- For each panel, the **left (blue) bar** are the students in the **control** condition and the **right (red) bar** are students in the **treatment** condition.

PLOS | ONE

Potential Strategies to Mitigate Bias
(How to Use Student Evaluations Wisely)

- Design survey instruments that **measure concrete behaviors**.

  - **Specify a time frame**:
    "*This instructor returned graded assignments within two weeks.*"

  - **Prompt more reflective, rational input; Encourage students to reflect on particular interactions with their instructor that inform their assessment**:
    "*This instructor was always in their office during their office hours*".

- **Write better (effective) questions** or questionnaire items.

  - Structure course assessments to help students avoid their own biases (e.g., by being thoughtful about what questions instructors ask students on course assessments, and how those questions are worded.

    - ✓ So watch out for global agreement statements such as
      "Overall, this person is an effective teacher."

Potential Strategies to Mitigate Bias
(How to Use Student Evaluations Wisely) (continued)

- **Use multiple measures** involving **multiple data sources** to **assess teaching more holistically.**

  - ➤ **ensure that student ratings are <u>not</u> the only source of evidence used to assess instructional effectiveness**

- **Consider using alternative methods of evaluating teaching (e.g., solicit feedback from outside observers / peer evaluations; observe faculty members teach; review an instructor's course materials; reflective statement / read faculty self-evaluations; mid-term course evaluations; focus groups / small group feedback**

- **Offer mentoring of faculty to raise awareness** (re: **bias in SET**)

- **Contextualize student ratings data**

- **Allow faculty to have a voice in the course evaluation process**

Suggested Options for Using Course Evaluations Effectively

❖ Use questionnaire items that are more **specific** such as questions about whether a student feels that *learning objectives have been accomplished,* or other specific items about

- o *self-efficacy* (did the student feel capable of succeeding?)

- o teaching methods the student observed, or

- o the student's motivation to take the course.

Suggested Options for Using Course Evaluations Effectively (continued)

❖ Consider using questionnaire items that take some of the onus of student learning off the instructor and make it clear it is a shared responsibility between students and instructors

- For instance, the University of Southern California (USC) revised its course evaluation instruments to improve the specificity of the questions to focus on the following elements:

    o Examine whether the course objective were explained,

    o Assignments reflected the material covered, and

    o Instructor(s) sufficiently explained difficult concepts, methods, and subject matter.

Suggested Options for Using Course Evaluations Effectively (continued)

❖ Consider using an approach that employs questionnaire items that prompt more-thoughtful feedback that evaluates teaching across several categories that might include:

- o Teaching Effectiveness (comprises of multiple measures …; and peer-reviewed measures that might include teaching reflection statements; syllabus/course materials review; review of assignments and grading; classroom or peer observation, etc.

- o Evidence of teaching development

- o Inclusive practices

- o Student engagement (place greater emphasis on student engagement and the shared responsibility between instructor and student)

One Way to Take Out the Sting of Student Feedback
(Ref: *The Chronicle of Higher Education*, March 14, 2019)

❖ Beckie Supiano (2019) describes one professor's strategy for analyzing student course evaluations … that involves a professor employing an assistant (a faculty colleague or consultant, partner/spouse, or any another resource) to read the students' feedback and divide the comments into three categories:

- Positive

- Constructive criticism

- Unconstructive criticism

o Professor's assistant then creates a document containing verbatim quotes of positive comments and a summary of the constructive criticism.
Note: Negative (or angry) comments are left out of the document entirely.

o Screening out angry evaluations (or "skipping to the good comments") produces a much more positive experience for a professor.

✓ Potentially, the strategy could result in a valuable outcome particularly for women faculty and faculty of color (i.e., groups that research suggests bear the brunt of needlessly personal comments from students).

One Way to Take Out the Sting of Student Feedback
(Continued)

❖ A key benefit of the approach (to screen out negative feedback) is how one characterizes the constructive criticism.

➤ Illustrative Example: Consider a student's comment such as: "Interactions with Professor Brown are difficult outside of class. Make office hours more accessible, increase frequency."

➤ One can categorize the student's comment into **actionable advice**, such as: "have more office hours"

✓ Note: The actionable advice "removes the emotional intensity" of the full comment in the example above.

Making Course Evaluations Effective
[Source: Felder (1993)]

- ❖ **Constructing, administering, & interpreting evaluations:**

  - ❖ **Collect overall course-end ratings of instruction**.
    - ➢ <u>Note</u>: Clearly define the numbers on the response scale
      - o Example: "Rate the instruction you received in this course on a scale from 1 to 5, with 5 being the highest response."

  - ❖ **Collect ratings of individual aspects of instruction**.

  - ❖ **Collect evaluations midway through a course** rather than waiting until the end. (purpose: *formative evaluation*)

  - ❖ Use **multiple methods to collect student feedback**
    - ➢ (e.g., from small groups of students; focus groups; interview student reps.)

  - ❖ **Use a variety of sources of feedback**
    - o Have faculty colleagues observe your teaching and provide feedback.
    - o Video record one of your classes and review the recording

  - ❖ **Work with an instructional consultant** to interpret student feedback and plan teaching improvement strategies

Useful Points to Consider …

- Instructors benefit most from *formative evaluation* if they have:

    - helped to shape the questions posed (e.g., via **Question Personalization** process);

    - a good understanding of the feedback provided; and

    - assistance and resources available for making improvements.

Formative Uses of Student Ratings

- **Formative** purposes: Useful feedback for *faculty development* and *enhancement of instruction*

  - ✓ For instance, student feedback can lead faculty to revise teaching methods, refine their courses, change textbooks, revise assignments, or make other changes to provide students with better learning experiences)

  - ✓ Research suggests that students are most qualified sources to report on the extent to which the learning experience was productive, informative, satisfying, or worthwhile.

Formative Uses of Student Ratings (continued)

- Provide useful information on students' perceptions of their engagement, learning outcomes, instructor's behavior, and course activities.

- **Identify teaching strengths and weaknesses**, as perceived by students.

- Play a positive role in **improving the climate of teaching and learning** at a college/university.

- Asking for student feedback regularly sends a clear message that **teaching effectiveness matters**…

Beneficial Ideas on How Instructors Might Improve their Courses

- Reflecting on goals for the course

- Reflecting on teaching methods

- Considering one's strengths & weaknesses as a teacher

- Targeting key areas that need improvement

- Identifying strategies for change
  (e.g., clarifying points or chunking content)

Ref.: Using Student Evaluations to Improve Teaching. *Speaking of Teaching*. (Stanford University Newsletter on Teaching), Fall 1997, *9*(1).

Additional Tips for Using Student Ratings to Enhance Effective Instruction

- Look for trends or patterns in the data

- Focus on key aspects of your course evaluations

- Don't give undue weight to open-ended comments from respondents

- Take into account course characteristics

Summative Uses of Student Ratings

- ***Summative*** (high-stakes) purposes: Student ratings are <u>one source of data</u> about teaching effectiveness for specific uses, such as:

  o Program review, and/or for meeting accreditation data requirements.

  o Assessment of student learning outcomes (assuming the instruments in use are designed appropriately and have good technical properties)

Summative Uses of Student Ratings
(continued)

- Make appropriate decisions about course and program level modifications

- Evaluate strengths and weaknesses of various instructional delivery modes, including the intersection of content and mode of delivery

- Measure program level and general education learning outcomes

- Evaluate and address concerns for preparatory course (e.g., the first course in a required sequence of courses)

- Meet accreditation standards and data requirements

Selected References

- Boring, A .(2017). Gender biases in student evaluations of teaching. *Journal of Public Economics*, *145*, 27–41. https://doi.org/10.1016/j.jpubeco.2016.11.006

- Boring, A., Ottoboni, K., & Stark, P.B. (2016). Student evaluations of teaching (mostly) do not measure teaching effectiveness. *ScienceOpen Research* (https://www.scienceopen.com/document/read?vid=818d8ec0-5908-47d8-86b4-5dc38f04b23e)

- Chisadza, C., Nicholls, N. & Yitbarek, E. (2019, June). Race and gender biases in student evaluations of teachers. *Economics Letters, 179*, 66-71. https://doi.org/10.1016/j.econlet.2019.03.022

- Fan, Y., Shepherd, L.J., Slavich, E., Waters, D., Stone, M., Abel, R., et al. (2019) Gender and cultural bias in student evaluations: Why representation matters. *PLoS ONE 14*(2): e0209749. https://doi.org/10.1371/journal.pone.0209749

- Falkoff, M. (2018, April 25). Why we must stop relying on student ratings of teaching. *The Chronicle of Higher Education*. https://www.chronicle.com/article/Whay-We-Must-Stop_Relying-on/243213

- Flaherty, C. (2019, May 20). Fighting gender bias in student evaluations of teaching, and tenure's effect on instruction. *Inside Higher Ed*

- Lilienfeld, E. (2016, June 10). How student evaluations are skewed against women and minority professors. *Commentary Higher Education*.

Selected References (continued)

- MacNell, L., Driscoll, A., & Hunt, A. N. (2015, August). What's in a name: Exposing gender bias in student ratings of teaching. *Innovative Higher Education*, *40*(4), 291-303. https://doi.org/10.1007/s10755-014-9313-4

- Mengel, F., Sauermann, J., & Zolitz, U. (2019, April). Gender bias in teaching evaluations. *Journal of the European Economic Association, 17*(2), 535-566. https://doi.org/10.1093/jeea/jvx057

- Merritt, Deborah J. (2008) "Bias, the Brain, and Student Evaluations of Teaching," *St. John's Law Review*, *82*(1), Article 6. https://scholarship.law.stjohns.edu/lawreview/vol82/iss1/6

- Mitchell, K. M. W., & Martin, J. (2018). Gender bias in student evaluations. *Political Science & Politics*. *51*(3), 648–652.

- Peterson, D.A.M., Biederman, L.A., Andersen, A., Ditonto, T.M., & Roe, K. (2019, May 15). Mitigating gender bias in student evaluations of teaching. *PLoS ONE 14*(5). https://doi.org/10.1371/journal.pone.0216241

- Reid, L. D. (2010). The role of perceived race and gender in evaluation of college teaching on RateMyProfessors.com. *Journal of Diversity in Higher Education*, *3*(3), 137–152. https://psycnet.apa.org/doi/10.1037/a0019865

- Rosen, A. S. (2018). Correlations, trends, and potential biases among publicly accessible web-based student evaluations of teaching. *Assessment & Evaluation in Higher Education, 43*(1), 31-44.

- Smith, B. P. & Hawkins, B. (2011). Examining student evaluations of black college faculty: Does race matter? *The Journal of Negro Education*, *80*(2), 149-162.

Q & A Session…

Question & Answer

and

Thank You!

Contact Information

❖ **Howard Mzumara, Ph.D.**

Director, Evaluation and Psychometric Services

Office of Institutional Research and Decision Support (IRDS)

Indiana University – Purdue University Indianapolis (IUPUI)

(317) 278-2214 (office phone)

hmzumara@iupui.edu

irds.iupui.edu