

Student Evaluations of Teaching (Mostly) Do Not Measure Teaching Effectiveness

Kellie Ottoboni

Department of Statistics, UC Berkeley
Berkeley Institute for Data Science

October 5, 2015

Student evaluations of teachers (SET) are used to

- ▶ Quantify teaching effectiveness
- ▶ Compare instructors across courses
- ▶ Make hiring, firing, and promotion decisions

Are SET a valid measure of teaching effectiveness?

No!

- ▶ SET measure “customer satisfaction”
- ▶ Ratings are biased against female instructors
- ▶ Biases are inconsistent across universities and disciplines; impossible to “adjust” SET

- ▶ Census of SET by first-year undergraduates, collected 2008–2013
- ▶ Students sign up for class times, don't know instructors; it's "as if" at random
- ▶ Male instructors are rated higher by male students than by female students
- ▶ SET correlate with grade expectations but not final grades

- ▶ Randomized, controlled experiment of students' perceptions of gender in an online course
- ▶ Female-identified TA was rated lower in all categories (especially “fair” and “prompt”) on average
- ▶ Average “overall” rating in each section

	Female-identified	Male-identified
Male instructor	3.75	4.00
Female instructor	3.625	4.333

Permutation tests

- ▶ Distribution-free tests, always give correct significance levels
- ▶ Compare assumptions:
 - ▶ Two-sample t-test: independent samples from normal distribution
 - ▶ Permutation test: randomization was fair
- ▶ Python package: `permute`
<https://github.com/statlab/permute>

- Anne Boring. Gender biases in student evaluations of teachers.
Document de travail OFCE 13, OFCE, April 2015.
- L. MacNell, A. Driscoll, and A. N. Hunt. Whats in a name:
Exposing gender bias in student ratings of teaching. *Innovative
Higher Education*, pages 1–13, 2014.