

# Why Student “Evaluations” of Faculty Are Unethical

Daryl Close

**Professor of Computer Science and Philosophy**

**Fall 2017 Faculty Forum**

**Heidelberg University**

**Tiffin, Ohio**

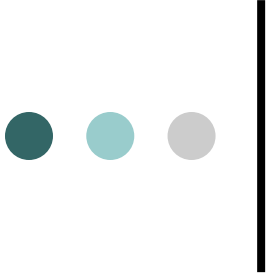
**October 26, 2017**



# Acknowledgments

I want to thank the Heidelberg University Faculty Personnel Committee, Provost Schwartz, President Huntington, and the Board of Trustees for approving my sabbatical leave, with special thanks to James F. Dicke II for his generous support of Heidelberg sabbatical research.

DLC



# Why Are There Scare Quotes in the Title?

- They're not rhetorical
- What students are doing when they fill out a ratings questionnaire is unsettled in the ratings literature



# Why Should We Care What Students are Doing?

- Only if students are competent evaluators or measurers of university teaching does the enormous literature on statistical validity and reliability have any bearing on faculty employment
- Polls of student beliefs or states of mind don't entail statements of fact



# What Are Students Doing?

- Possibility #1: Students are evaluators of the professor's teaching. Issue: are they competent?
- Possibility #2: Students are conducting measurements of objective features of the professor's teaching. Issue: are they accurate?



# What Are Students Doing?

- Possibility #3: Student ratings are tests or experiments regarding student traits. Issue: are they ethically exempt from IRB review?
- Possibility #4: Students are expressing their opinions and beliefs about the professor's teaching. Issue: does belief entail knowledge?



# What Are Students Doing?

- Possibility #5: Students are expressing their (relative) satisfaction with the professor's teaching. Issue: is student satisfaction the goal of university teaching?



# Student Ratings Aren't Evaluations

- Leading ratings researchers agree that student ratings are opinion polls or polls of student satisfaction, not evaluations of teaching (Theall & Franklin 1990; Linse 2017)



# Something of a Scandal

- The use of student ratings in faculty employment decisions is studded with ethical problems
- There are only a few passing references to those problems in more than 90 years of ratings research



# Paper Overview

- **Goal:** identify problems that are fatal to the entire enterprise of the administrative use of student ratings in employment decisions
- **Overall conclusion:** such use of student ratings should be immediately and universally suspended



# What My Research Does Not Include

- Instructor's solicitation of student "feedback" about the course (tests, texts, grading, homework, course organization, content delivery, etc.)
- Professors' ethical obligations to students
- Ethical methods of peer evaluation of teaching and research
- Philosophical problems in test theory



# Criterion 1 for Ethical Faculty Evaluation

- **Principle of Nonmaleficence;** do no harm (primum non nocere)
- Exclude/limit possibility of false negative evaluations
- Evaluator competence
- No coercion



# Criterion 2 for Ethical Faculty Evaluation

- **Principle of Beneficence**; do good
- For faculty, ethical evaluation focuses on the development of both disciplinary knowledge and clinical (pedagogical) skills
- For students, ethical evaluation promotes student learning



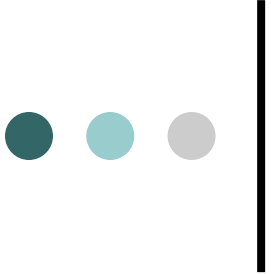
# Criterion 3 for Ethical Faculty Evaluation

- **Principle of Respect for Autonomy**
- Traditional professions (medicine, law, university teaching) are based on professional autonomy
- Evaluation of a professional's disciplinary knowledge and/or clinical skills by a nonprofessional is prima facie unethical
- Academic freedom is a species of professional autonomy



# Criterion 4 for Ethical Faculty Evaluation

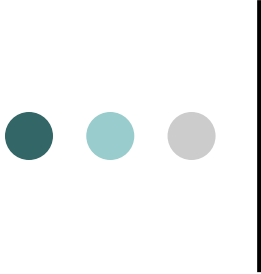
- **Principle of Justice**; fairness, due care, honesty
- Treat those that are alike alike, treat those that are unlike differently (Aristotle; AAUP *Statement on Teaching Evaluation*)



# Principle of Justice

## Part 2

- Due care: establish a national consensus concept of teaching excellence
- Truthfulness: no undisclosed “black ball” criteria for adverse employment decisions



# Example of Unjust Evaluation (Common)

- “[The ethical problem of the] use of unspecified relative weights for the teaching, research, and service categories, with the result that department chairs, committees, and deans can do whatever they like and fabricate some rationalization for it. This is known as inequitable treatment” (Scriven 1982)



# Student Ratings Ethical Problem #1

1. Weak and confused conceptual foundations--Concepts of teaching effectiveness, student evaluations of teaching effectiveness, psychological measurement, construct validity, and evaluation itself are at least unsettled, and in some cases, conceptually inchoate.
- Violates principle of justice (due care, truthfulness)



# What Is Teaching “Effectiveness?”

- Contrary to Scriven’s theory of evaluation (1967), equating teaching behaviors or styles with effectiveness has become nearly universal in SET research (Buck 1998)

BUT . . .



# A Genus, Not a Species

- The expression “student ratings” doesn’t denote a specific thing. There is no uniform “they,” as ratings researchers imagine, but only many distinct student ratings forms—perhaps hundreds and hundreds of them—based on different definitions of teaching effectiveness.



# The Dirty Little Secret

- *There is no consensus definition of teaching effectiveness, the very trait, or collection of traits, that student rating items are alleged to evaluate or measure*
- This ethical problem alone constitutes a sufficient condition for immediately ending student ratings. We literally do not know what we're talking about.



# Student Ratings

## Ethical Problem #2

2. Coercion--Using administrative (and faculty) authority to induce students to participate in a process that may lead to harm is morally wrong. Medical researchers, psychologists, and other human subject researchers are well-acquainted with this fundamental ethical principle.
- Violates principles of nonmaleficence, autonomy, and justice



# Student Ratings

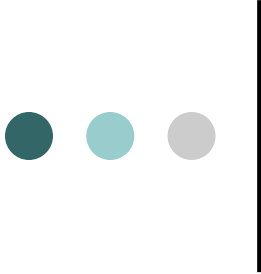
## Ethical Problem #3

3. Unqualified Evaluators--Faculty evaluation can only be ethically undertaken by qualified evaluators, so we should expect to find strong and persuasive arguments in the literature regarding student competence in evaluating university teaching. The argumentation is thin to nonexistent.
- Violates professional autonomy and justice (esp. due care)



# The “Long-exposure” Premise of Competency

- “No one else is as qualified to report on what transpired during the term simply because no one else is present for as much of the term” (Theall & Franklin 2001)
- Asserted repeatedly for decades as if self-evident (Guthrie 1927; see also Aleamoni 1974; Detchen 1940; Kulik & McKeachie 1975; Remmers 1958; Theall & Franklin 2001)



# Competent Evaluation Not a Function of Time

- Long exposure is not sufficient for competent evaluation
- Example: 4-H Mary spends a lot more time with her show chicken than the poultry judges at the county fair, but that fact doesn't make her a competent judge of show chickens
- Long exposure is not necessary (expert judges work quickly)



# Student Ratings

## Ethical Problem #4

4. Interference with Academic Freedom—  
Presenting faculty with a no-win dilemma:  
“You must choose between (a) possible nonrenewal, denial of tenure, denial of promotion, and (b) making pedagogical and behavioral changes that will optimize the student ratings on which those employment decisions are based.” “And, no, student learning doesn’t matter.”
  - Violates professional autonomy, due care



# Student Ratings

## Ethical Problem #5

5. Involuntary Social Experimentation--“Let’s make adverse faculty employment decisions based on student ratings and see if student learning improves.”
  - Violates principles of professional autonomy, nonmaleficence, and justice



# Student Ratings Ethical Problem #6

6. No Evidence that Use of Student Ratings Improves Learning--If teaching “effectiveness”—whatever that is—is relevant to employment, then improving teaching effectiveness as determined by student ratings should somehow connect with improvement in student learning. There is no empirical evidence of any improvement (Armstrong 1998).
- Violates the principle of beneficence, due care



# Opinion/Satisfaction Polls

- Evidence-based conclusions about the clinical skills of a professional cannot be drawn from the opinions or beliefs of the nonprofessional
- More generally, facts cannot be deduced from opinions or beliefs



# Satisfying the Customer

- On the consumerist model of college education, faculty employment decisions could be based on student satisfaction rather than evaluation by professional peers
- Would require full disclosure in hiring, tenure, promotion, and merit pay



# Consumerism and “Accountability”

- A bill, introduced in the Iowa legislature by State Sen. Mark Chelgren in 2015 would have required the state’s public universities to rate professors’ performance based solely on students’ evaluations of their teaching effectiveness. Professors whose evaluation scores didn’t reach a minimum threshold would be automatically fired by the university.



# All In On Satisfaction

- Is that the kind of college you would want to devote your career to?
- What quality of faculty would choose student satisfaction over student learning when faculty employment decisions are driven by satisfaction rather than learning?
- Can that kind of college even survive in today's environment?



# University Teaching Is a Profession

Yet we administer student ratings, semester in, and semester out, year in and year out, deny and award tenure, deny and award promotion, and make other life-changing faculty personnel decisions unethically based on those ratings. Why is that?



Daryl Close  
Heidelberg University  
Tiffin, Ohio

Revised 25-Oct-2017

**Copyright © 2017 Daryl Close** This work is licensed under the Creative Commons Attribution-Noncommercial-No Derivative Works 3.0 United States License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/3.0/us/> or send a letter to Creative Commons, 171 Second Street, Suite 300, San Francisco, California, 94105, USA.



# Appendix A

## What About All of Those Ratings Studies?

- Generic, meta-analytic assertions of ratings validity and bias are scientifically specious precisely because there is no common, cross-study construct that is being tested.



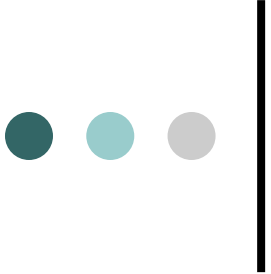
# Looking for Morality in All the Wrong Places

- Statistical validity and reliability are not sufficient for the moral acceptability of a faculty evaluation process
- “If ratings are valid, then they’re ethical” is flatly false



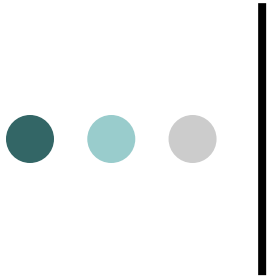
# Definitely Not a Sufficient Condition

- Example: The grotesque ethical faults of the clinical study, “Tuskegee Study of Untreated Syphilis in the Negro Male” are hardly a consequence of statistical weaknesses. Statistical validity (or reliability) is simply not sufficient for moral acceptability.



# Probably Not a Necessary Condition

- Example: Blinding is prima facie morally wrong in certain types of human subject research (Büller, et al., 2008), yet is typically regarded as essential to validity



# References

- Available on request