Making Student Ratings Meaningful to Faculty: Creation of a New Type of Student Ratings Instrument

Kathleen Dyer, PhD

Professor & Chair, Dept of Child and Family Science

Dermot Donnelly-Hermosillo, PhD

Associate Professor, Dept of Chemistry and Biochemistry

FRESNO STATE

Agenda

Preface

1. Criticism of student ratings
2. Why to keep student ratings
3. Solution: a better instrument
4. Psychometric Properties & Correlates
5. Potential Bias

Conclusion

Preface: Fresno State's story

- Numerical student ratings have been required by our union and by our APM since the 1980s.
- Early 2000s: documentation of wide variety of instruments, never tested. Revised policy in 2011.
- Adopted the IDEA instrument because it was, essentially, the only standardized instrument on the market. It did not align with our policy except for its standardization.
- 2018: Widespread discontent. Senate Task Force.

1. Criticism of Student Ratings

INSIDE HIGHER ED

Even 'Valid' Student Evaluations Are 'Unfair'

New study says student evaluations of teaching are still deeply flawed measures of teaching effectiveness, even when we assume they are unbiased and reliable.

By Colleen Flaherty // February 27, 2020

aaup
AMERICAN ASSOCIATION OF UNIVERSITY PROFESSORS

Student Evaluations of Teaching are Not Valid

It is time to stop using SET scores in personnel decisions.

By John W. Lawrence

THE CHRONICLE OF HIGHER EDUCATION

Students Evaluating Teachers Doesn't Just Hurt Teachers. It Hurts Students.

By Nancy Bunge | NOVEMBER 27, 2018

CALIFORNIA STATE UNIVERSITY FRESNO 19 11

1. Criticism of Student Ratings

"Not Valid"

Studies in Educational Evaluation
Volume 54, September 2017, Pages 22-42

Meta-analysis of faculty's teaching effectiveness: Student evaluation of teaching ratings and student learning are not related

Bob Uttl, Carmela A. White [1], Daniela Wong Gonzalez [2]

Show more

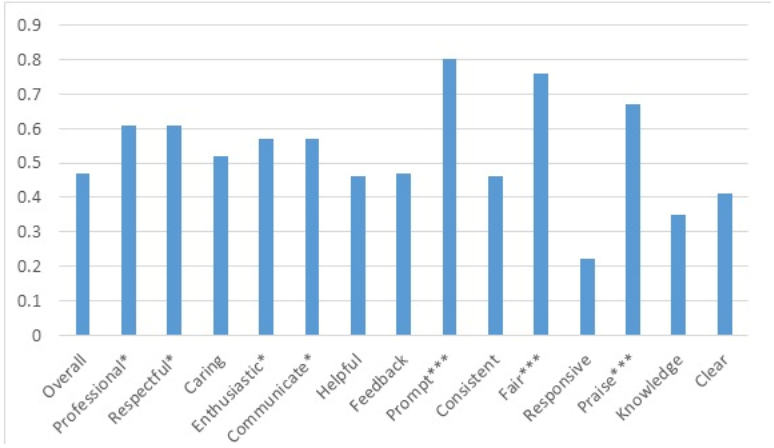
https://doi.org/10.1016/j.stueduc.2016.08.007 Get rights and content

Highlights

- Students do not learn more from professors with higher student evaluation of teaching (SET) ratings.
- Previus meta-analyses of SET/learning correlations in multisection studies are not interprettable.
- Re-analyses of previous meta-analyses of multisection studies indicate that SET ratings explain at most 1% of variability in measures of student learning.
- New meta-analyses of multisection studies show that SET ratings are unrelated to student learning.

"Biased"

Figure 3. Difference in mean ratings and reported instructor gender (male minus female)



https://blogs.lse.ac.uk/impactofsocialsciences/2016/02/04/student-evaluations-of-teaching-gender-bias/

1. Criticism of Student Ratings

- What produces the lack of validity and the bias?
- There are many instruments, and most are not constructed by scientists or tested for reliability/validity. But they all get lumped together.
- Common Problems with existing instruments:
  - Global ratings that allow bias (“how good is this class?”)
  - Judgement of the invisible (“my instructor cares”)
  - Impossible judgements (“my instructor is knowledgeable”)
  - Evaluation of learning, despite a well-documented overconfidence effect that makes these self-assessments worse than useless

2. Why Keep Student Ratings

2. Why Keep Student Ratings

- Student Voice
  - An official mechanism: the University verifies that only enrolled students are able to submit a response, while allowing anonymity.
  - Feedback is seen by the supervisor of the instructor.

- Professional Development
  - Student feedback improves instruction (Cohen, 1980)

3. Solution: a Better Instrument

- These are not problems inherent to the concept of student rating.
- These are measurement problems.

- Our task force, which included experts in survey construction as well as content experts from all colleges on campus, set out to create a new instrument that would not have these measurement problems.
- A new type of instrument – one designed to be meaningful to instructional faculty.

3. Fresno State Student Ratings of Instruction questionnaire

<u>APM Requirements:</u>

1. Assess
    - Instructional Design
    - Instructional Delivery
    - Assessment
    - (Content should be assessed by expert peers rather than students.)
2. Demonstrated reliability and validity
3. Pool of items for dept/instructor choice

3. Fresno State Student Ratings of Instruction questionnaire

- Abandon the problematic strategies already described.
- Instead, ask students to report:
  - directly observable behaviors of faculty (”my teacher did x”)
  - or student self-report of their own understanding (“I understood y”)
  - that are evidence-based (peer-reviewed scientific publication)

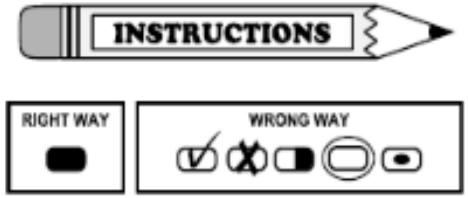
3. Fresno State Student Ratings of Instruction questionnaire

- We reviewed the scientific literature and identified such practices in our three categories.
- Instructional Design: clear objectives, clear expectations, relevant course materials, logical organization
- Instructional Delivery: scaffolding, active learning, connections, welcoming environment
- Assessment: frequent low-stakes assessment, timely grading, clear purpose, constructive feedback

Fresno State Student Ratings of Instruction Questionnaire

Default Items

For each item, please indicate your level of agreement.



	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
Instructional Design					
1. The objectives of this course were clear to me.	A	B	C	D	E
2. The syllabus was an accurate guide to course requirements.	A	B	C	D	E
3. The assigned readings were integral to the course content.	A	B	C	D	E
4. The course content was presented in an organized manner.	A	B	C	D	E
Instructional Delivery					
5. The atmosphere of the class invited students to seek additional help if needed.	A	B	C	D	E
6. My instructor explained new ideas by relating them to familiar concepts.	A	B	C	D	E
7. The class was structured so that students regularly talked with one another about the concepts.	A	B	C	D	E
8. The class connected course content to students' future work.	A	B	C	D	E
Assessment Methods					
9. Feedback was provided to help guide students' progress in this course	A	B	C	D	E
10. Graded assignments were returned to me in a timely fashion.	A	B	C	D	E
11. The class included multiple graded assignments.	A	B	C	D	E
12. The purpose of assignments was clear to me.	A	B	C	D	E

3. Meaningful because…

- Faculty-generated
- Students know exactly what they are being asked.
- A low score on any of these items contains information about exactly how to improve on very tangible pedagogical practices

4. Psychometric Properties

- Spring 2019 semester pilot test of the FSSRI
- 53 instructors, 81 course sections, 2013 student surveys (all colleges on campus, all ranks, some gender and racial diversity)
- Internal Reliability: Cronbach's alpha = .93 - .97
- Convergent Validity
  - Student centered practices subscale of the Post-Secondary Instructional Practices Survey: r=.25, p=.03
  - Student scores and instructor self-report: r=.29, p=.01
  - IDEA instrument: r=.80, p<.0001

4. Correlates

	Correlation with overall FSSRI score r (p-value)
Overall course rating	.76 (p<.0001)
Instructor starts and ends class on time	.35 (p<.0001)
Expected grade	.28 (p<.0001)
Class difficulty	-.22 (p<.0001)
Frequency of instructor absence	-.12 (p<.0001)
Frequency of student absence	-.07 (p=.001)

5. Potential Bias

- Scores NOT correlated with:
  - Class size
  - Upper/lower division
  - Instructor rank (PTF, FT Lecturer, Asst, Assoc, Prof)
  - Instructor gender (female, male)
  - Instructor race (white, non-white)
  - Student level (frosh, soph, jr, sr, grad)
- Scores are consistently lower in classes with quantitative content
  - 3.9 in engineering college, 4.6 in education college
  - By instructor report: 4.4 with no quant content, 4.3 with some, and 4.1 with all

5. Potential Bias

- Gender
  - Well-documented in published literature, effect size is almost half a point on a 5-point scale (MacNell, Driscoll & Hunt, 2015),
  - Not identified here because we did not compare within a single discipline or course and we did not look at the gender match of the instructor/student.
- Race
  - Not so well-documented.
  - Much more complicated because there are more potential groups, and more identified correlates (e.g., language, immigrant status, etc.)
  - Our sample was far too small to look for the nuances.

Conclusion

- The contemporary narrative about student ratings in higher education is appropriately concerned about validity and bias.
- But these are measurement problems, not an existential crisis for student ratings. We need to be careful not to eliminate student voice and lose our access to feedback that can help improve instruction.
- Therefore, a new approach to measurement should be considered. The FSSRI is one such attempt. It needs further analysis, but shows promise.