

# Student Evaluations of Teaching (SETs)

Review on the use of SETs, with  
recommendations for assessment of  
teaching faculty for employment  
decisions

# SET questions revised in 2017

- Previous questions had been used since 2011
- Process included an in-depth review of existing research on SETs
- Committee found that SETs are of limited value for *summative* evaluations
  - Summative = absolute evaluation of teaching effectiveness
- Set out to develop questions of greatest value for formative evaluations
  - Formative = with a focus on improving effectiveness
- Questions based on the 7 Principles for Good Practice in Undergraduate Education (Chickering and Gamson 1987)

# NEW SET questions

- I was encouraged to interact with the instructor regarding course content
- I was provided opportunities to cooperate with other classmates about course material
- I was informed of the instructor's high expectations for my work in this course.
- I was provided with an evaluation of my academic progress at regular intervals during the semester.
- I was provided with ample opportunities to apply my learning in this course.
- I was prompted to think critically about the course material.
- I was provided an environment that supported my learning.

# Committee Report on Use of SETs

- Review of existing literature on utility of SETs
- Review expert advice on how SETs should be used for evaluations of teaching effectiveness
- Recommendations on the how teachers should be evaluated at Auburn University

# Existing literature on SETs

- Students are not qualified to evaluate faculty for teaching effectiveness
- Evidence for the validity of SETs for measuring individual teaching effectiveness is weak (at best)
- Many sources of bias in SET scores exist

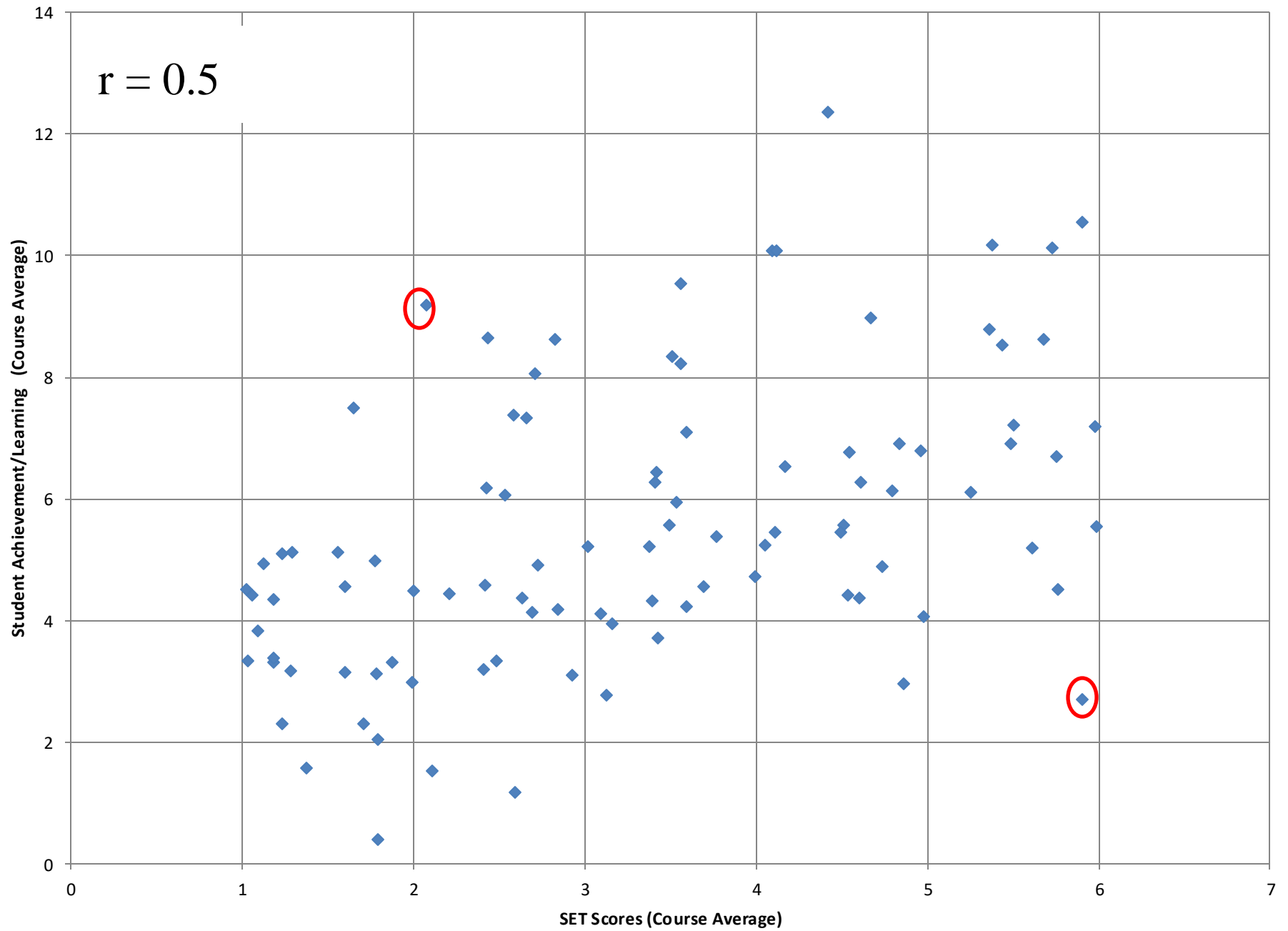
# Students not qualified to evaluate...

- Effectiveness of a teacher
- Numerous skills that define teaching effectiveness
  - Knowledge, methods, course design, use of technology, course materials, grading, etc.

SETs are used to gather the collective views of a group of students about their experience

# SETs as measures of effectiveness

- Numerous meta-analyses of the relationship between SET scores and student learning
  - Cohen (1981, 1982, 1983); Feldman (1987); McCallum (1984); Clayson (2009).
- Correlation coefficient ( $r$ ) from 0.13 to 0.44



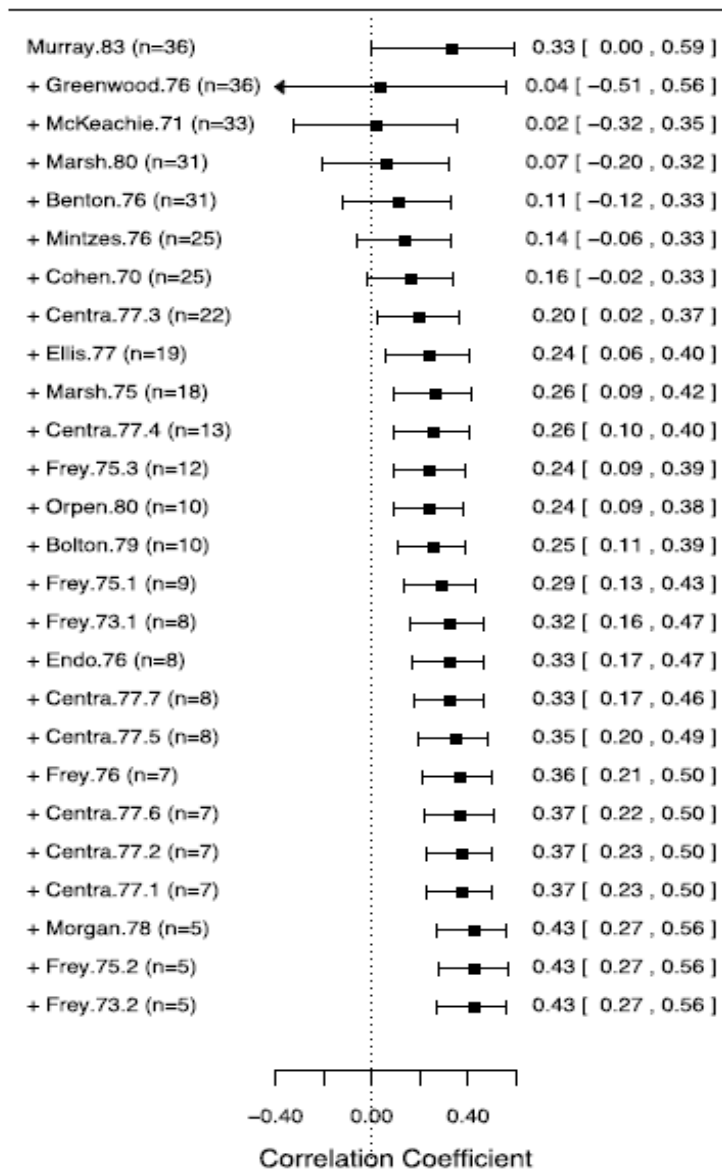


# SETs as measures of effectiveness

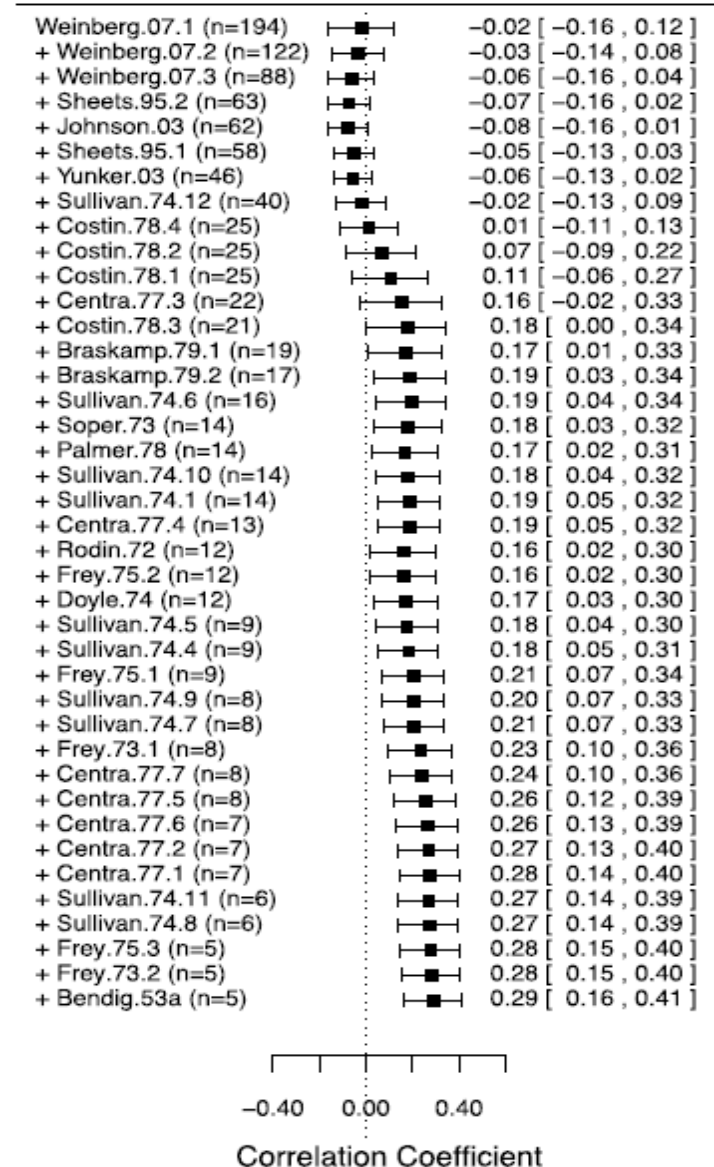
- Numerous meta-analyses of the relationship between SET scores and student learning
  - Cohen (1981, 1982, 1983); Feldman (1987); McCallum (1984); Clayson (2009).
- Correlation coefficient ( $r$ ) from 0.13 to 0.44
- If  $r = 0.44$ , 80% of variation in SET due to something else
- Valid instruments for measuring students learning *in aggregate*.

# Small sample bias in previous studies

- Uttil et al. 2017
- Previous studies did not consider effect of small sample bias
  - Studies with small sample sizes need stronger effects to get significance p-values
  - Studies without significant p-values don't get published



Feldman 1989



Clayson 2009

# Small sample bias in previous studies

- Uttl et al. 2017
- Previous studies did not consider effect of small sample bias
- Re-analysis of previous data
  - Adjusted  $r$  from 0.05 to 0.27
- Analysis of new data
  - Adjusted  $r$  from -0.02 to -0.04

# Research on Bias in SETs

## Numerous potential sources

- Faculty rank
- Faculty gender
- Student gender
- Faculty expressiveness
- Student motivation
- Required course or not
- Expected grade
- Level of course
- Class size
- Academic discipline
- Student workload

# Global questions

- “The instructor’s overall effectiveness was....”
- Particularly inappropriate for evaluating instructors
  - Low reliability
  - Not fair to evaluate based on a single score
    - Use all scores
  - Violates standards related to personnel evaluation

# Implications

- SETs are not valid instruments for measuring effectiveness of individual teachers.
- SETs should not be used in a summative manner for evaluating teachers.

# Recommendations

- SETs should not be used in summative manner
- SETs could be used in a formative manner
  - Regardless of absolute scores, do scores improve over time?
  - Indicative of a teacher that takes student learning seriously



# Recommendations

- Number of evaluations if used
  - Minimum 6 to 8 courses total
  - Courses with fewer than 10 raters excluded
  - For summative employment decisions:
    - 70% response rate minimum
  - For formative considerations:
    - 30% response rate acceptable
- No Global Questions!
  - Use all scores or means from scores

# Recommendations

- No comparisons with other faculty
  - Prone to bias due to gender, race, other characteristics

# SETs MUST be used with other metrics

- Peer review of course material
- Peer review of course instruction
- Review of metrics of teaching effort by supervisor
  - (course policies followed; grades submitted on time)
- Review by expert outside sources (Biggio)
- Exit and alumni ratings

# Other metrics of teaching performance

- Employer ratings
- Teaching scholarship
- Teaching awards
- Learning outcome measures
- Teaching portfolios
- Self-evaluation of teaching

Generation of quantitative rubrics for any