

Student Ratings of Instruction

Making Them Work for You

Introduction

- Well over 2,000 studies on SRIs
- Called both SETs and SRIs-many feel evaluation is distinct from rating
- Too much reliance on SRIs in evaluating teaching and lack of guidelines on interpreting results lead to most of abuses and problems with SRIs

Think About It

What are your greatest evaluation challenges in using the current student ratings?

What are some myths?

- SRIs are popularity contests-easy teachers get the highest ratings.
- Students cannot evaluate the effectiveness of their instructors due to a lack of knowledge and experience.
- Students use the evaluations to get even with teachers who criticize them.

What are some myths?

- SRIs can't improve teaching.
- Teaching is a complex activity (or: teaching is an art) and cannot be measured.
- SRIs measure showmanship not good teaching.
- SRIs are based on expected grades.

Think About It

- What research findings do you think will be most valuable in assisting deans and department heads in interpreting student ratings?
- What research findings should tenure committees know about when looking at a candidate's teaching?

What do we know?

- **Validated** SRIs are reliable, and stable.
- **Validated** SRIs are relatively valid against a variety of indicators of effective teaching.
- **Validated** SRIs relatively unaffected by variables hypothesized as potential biases.
- **Most** SRIs are routinely misused and misinterpreted.

What do we know?

- **Validated** SRIs are reliable, and stable.
 - Teachers rated highly in class are rated highly later on (alumni studies); those with poor ratings continue to get poor ratings. Teachers rated highly by one group tend to be rated highly by another, e.g., peers and alumni. (Howard, Conway, & Maxwell, 1985; Marsh, 1982)
- Whether rated within classes, across classes, over time or in other ways, ratings consistent.
(Feldman, 1977; Murray, Rushton, & Paunonen, 1990; Theall & Franklin, 2001)

What do we know?

- **Validated** SRIs are valid against a variety of indicators of effective teaching.
 - Consistently high correlations between students' ratings of "amount learned" as shown on exams and overall ratings of teacher and course (Abrami, d'Apollonia, & Cohen, 1990; d'Apollonia & Abrami, 1997)
 - UConn pilot: Perceived learning: 35% of variance in overall teacher rating

What do we know?

- **Validated** SRIs are valid against a variety of indicators of effective teaching.
 - In multi-section courses, the highest ratings of instructors given by students with best performance on exam.(.40, Cohen 1981; Marsh & Roche 2000)

What do we know?

- **Validated** SRIs are valid against a variety of indicators of effective teaching.
 - **Observer ratings** (Aleamoni & Hexner, 1980)
 - **Alumni ratings** (Aleamoni & Yimer, (1974); Marsh, 1977; Marsh & Overall, 1979; McKeachie, Lin, & Mendelson, 1978)

What do we know?

- Validated SRIs are unaffected by workload/difficulty:
 - Workload/grading leniency (stable over 12 years): Large national studies show more difficult courses given higher ratings than easy courses. (Marsh 2001)
 - UConn pilot report: Perceived difficulty of course accounts for less than 1% of variance

What do we know?

- Validated SRIs are unaffected by expected grade:
 - *Over 500 research articles on this topic!
 - Research with expected grades in multiple section courses: differences in instructor ratings with no difference in expected grades (grading FAIRNESS does influence)

What do we know?

- Validated SRIs unaffected by expected grade:
 - Size of grade-SET relation small (.20 for overall teacher ratings); explained by prior subject interest, better perceived learning outcomes (Marsh & Roche, 2000)
 - UConn: Expected grade 5% of variance
(Remember, how much learned: 35% variance!)

What do we know?

- Validated SRIs are unaffected by other characteristics:

- Rank of instructor: No!
- Class size: No!
- Time of day: No!

(To see full list of studies: Arreola, 2006, pp. 102-103)

What do we know?

- SRIs are unaffected (?) by variables: **gender**.
 - No significant relationship between gender of instructor and overall evaluation (.02 Feldman, 1993; female slightly higher) *some evidence that ratings slightly higher when students same gender as instructor (Ory 2001); some same-gender bias with female students (Basow & Silberg, 1987; Centra & Gaubatz, 2000; Hicks & Santhanam, 2002)

What do we know?

- SRIs are affected by interaction of gender and race:
 - BUT important emerging research: perceptions of warmth and strictness magnified for female faculty of color (Anderson & Smith, 2005)
 - Lower final ratings for female minority but not male (Hamermesh & Parker, 2005)

What do we know?

- SRIs are affected by:
 - Non-native English speaking instructors have significantly lower course evaluations overall, male NNES instructors fared worse than female NNES instructors (Hamermesh & Parker, 2005)
 - Faculty of color received lower evaluations than white peers: Hispanic faculty received lowest, Asian-American better but lower than white faculty (DiPietro & Faye, 2005)

What do we know?

- SRIs are affected by:
 - **Discipline:** Arts and humanities, biological and social sciences, computer science, math, engineering, physical science (Cashin, 1990, 1992; Centra, 1993b; Franklin and Theall, 1992; Hativa & Marincovich, 1995)
 - **Prior interest:** Desire to take the course-both teacher and course ratings increase as prior interest increases (Marsh & Cooper, 1981; Ory, 1980; Perry, Abrami, Levanthal & Check, 1979)
UConn Pilot: 11% of variance in overall teacher rating; 16% overall course

What do we know?

- SRIs are affected by:
 - * **Level of courses**: grad courses and higher level undergrad courses tend to receive higher ratings (Aleamoni, 1981; Braskamp & Ory, 1994; Feldman, 1978)
 - **Required vs. elective**: required courses tend to receive lower evaluations than elective, even in major (Costin, Greenough & Menges, 1971; Feldman, 1978; McKeachie, 1979; Marsh, 1984)

*very small effect

Think About It

- What research findings do you think will be most valuable in assisting deans and department heads in interpreting student ratings?
- What research findings should tenure committees know about when looking at a candidate's teaching?

Current Practices?

- Common issues in construction of SRIs:
 - Inappropriate questions
 - students cannot judge faculty knowledge or currency of materials
 - Only high inference questions
 - Questions not matched to teaching context
 - Inappropriate scales
 - Only high end and low end defined
 - Evaluative rather than diagnostic

Current practices?

- Common Issues in the use of SRIs for evaluation:
 - Inappropriate judgments caused by:
 - Lack of information for faculty or administrators on how to interpret results
 - Comparing apples and oranges
 - Make fine-tuned distinctions (7.4=merit/7.2=no merit)
 - Using written comments
 - Using only student ratings to evaluate teaching

What Is Good Practice?

- Never use SRIs as the only measure of teaching!
 - Ongoing (formative) assessment
 - Student learning/achievement
 - Portfolios: Course or Teaching
 - Peer observations

SRIs are necessary but not sufficient!

What Is Good Practice?

- Student rating form questions:
 - Formative *and* summative questions
 - Should have some common questions and the rest chosen by department and/or instructor
 - Mix of low-inference (specific, unambiguous behaviors that can be isolated and observed) and high-inference questions (overall course and instructor effectiveness)

What Is Good Practice?

- Additional considerations:
 - Comments should be tied to groups of questions-not just at the end: EX. student learning, instructor behaviors, materials and activities
 - Need student demographics: year, prior interest, major, elective/required course, student effort, perceived student learning

What Is Good Practice?

- Interpretation:
 - Faculty and administrators need research in order to correctly weigh and interpret numbers
 - Need to interpret differently ratings from elective vs. required courses
 - Discipline differences-math and natural sciences likely to receive lower ratings than those in other disciplines

Think About It

- What changes would you make to your current data collection/forms and why?
- What would keep the same and why?
- What changes would you make (if any) to your current process for evaluating teaching?

Ready for Change?

- Changing the **forms**:
 - Include all stakeholders in decisions about process and policy
 - Administrators, faculty, students, institutional research
 - Focus groups, town hall meetings, feedback on website, voluntary pilot, feedback on pilot, have students do a “talk-through”

Ready for Change?

- Changing the **practices**:
 - Publicly present clear information about evaluation criteria, process and procedures:
 - Create website with FAQs.
 - Publish protocols for administration and handling of SRIs for students, faculty and administrators.
 - Transparency in how SRIs are used (especially for students), who gets what information and why. (Formative vs. summative)

Ready for Change?

- Improving Teaching:
 - Produce reports that can be understood easily and accurately, i.e., graphic displays using confidence intervals clearly showing individual difference from comparison groups
 - Provide resource information along with reports
 - Use of longitudinal data that values improvement

Ready for Change?

- Improving the whole process!
 - Provide both formative and summative information
 - Keep formative confidential and separate from summative
 - Educate all users to avoid misuse and misinterpretation
 - Adhere to rigorous psychometric and measurement principles and practices (pilot and validation)
 - Regularly evaluate evaluation system
 - Establish system of grievances
- (Theall & Franklin, 2001)

Good (Quick) sources

- Arreola, R. (2007). Developing a Comprehensive Faculty Evaluation System. Bolton, MA. Anker.
- Seldin, P. (1999). Changing Practices in Evaluating Teaching. Bolton, MA. Anker.
- University of Michigan has good summary and bibliography:
(www.crlt.umich.edu/tstrategies/tseot.html)

Prepared by:
Catherine Ross, Director
Teaching & Learning Center
Wake Forest University
rossce@wfu.edu