The Design and Use of an Effective Student Ratings of Instruction System

June 26, 2016

Ken Ryalls, Ph.D., President
Steve Benton, Ph.D., Senior Research Officer
Elements of an Effective SRI Instrument (Hour 1)

- Focus on what students are competent to rate
- Tailor to course purpose
- Include summary questions
- Control for non-instructional factors
- Provide validity evidence
- Provide reliability evidence
- Provide comparative data
- Standardize administrative procedures
What Are Students Are Competent to Rate?

Students are competent to rate:

- What occurred in class
- Teacher behaviors
- Self-perceptions of learning
- Other self-perceptions
- Overall impressions of course and teacher
• Students are **NOT** competent to rate:
  – Instructor subject-matter knowledge
  – Appropriateness of course content
  – Quality of course goals/objectives
  – Quality of evaluation system
THINK-PAIR-SHARE

• Write down examples of elements in your current SRI instrument students are competent to rate.
• Write down any examples of elements in your system students are NOT competent to rate.
IDEA SRI System

• Progress ratings on 13 learning objectives
• Frequency of 19 teaching methods
• Ratings of amount of coursework, difficulty
• Self-ratings of background preparation, self-efficacy, work habits, motivation
• Overall ratings of teacher and course
Tailor SRI to Course Purpose

• Connect items to course goals/objectives
  – Reflects instructor’s purpose
  – Provides more meaningful feedback
  – Increases validity of ratings
IDEA Objectives Selection Form
(formerly Faculty Information Form - FIF)

- Types of learning must reflect the course purpose
- Rollover
- Autofill
IDEA SRI System

• Evaluation based on student ratings of progress on learning objectives identified by teacher as relevant to course

<table>
<thead>
<tr>
<th>Progress on Relevant Objectives</th>
<th>Overall Ratings</th>
<th>3.1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gaining a basic understanding of the subject (e.g., factual knowledge, methods, principles, generalizations, theories)</td>
<td>Your Average</td>
<td>3.2</td>
</tr>
<tr>
<td></td>
<td>Your Average Comparison</td>
<td>29</td>
</tr>
<tr>
<td></td>
<td>Percent of Students Rating</td>
<td>33%</td>
</tr>
<tr>
<td></td>
<td>1 or 2</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>67%</td>
</tr>
<tr>
<td></td>
<td>4 or 5</td>
<td></td>
</tr>
<tr>
<td>Developing knowledge and understanding of diverse perspectives, global awareness, or other cultures</td>
<td>Your Average</td>
<td>2.9</td>
</tr>
<tr>
<td></td>
<td>Your Average Comparison</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>Percent of Students Rating</td>
<td>22%</td>
</tr>
<tr>
<td></td>
<td>1 or 2</td>
<td>56%</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>22%</td>
</tr>
<tr>
<td></td>
<td>4 or 5</td>
<td></td>
</tr>
<tr>
<td>Learning to apply course material (to improve thinking, problem solving, and decisions)</td>
<td>Your Average</td>
<td>4.1</td>
</tr>
<tr>
<td></td>
<td>Your Average Comparison</td>
<td>49</td>
</tr>
<tr>
<td></td>
<td>Percent of Students Rating</td>
<td>11%</td>
</tr>
<tr>
<td></td>
<td>1 or 2</td>
<td>11%</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>78%</td>
</tr>
<tr>
<td></td>
<td>4 or 5</td>
<td></td>
</tr>
<tr>
<td>Developing specific skills, competencies, and points of view needed by professionals in the field most closely related to this course</td>
<td>Your Average</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Your Average Comparison</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Percent of Students Rating</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1 or 2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4 or 5</td>
<td></td>
</tr>
</tbody>
</table>
Include Summary Questions

• Include questions about:
  – Overall teaching quality
  – Overall quality of course
  – Overall student learning

• Global items more valid than average of dissimilar items

• Self-ratings of learning correlate positively with:
  – Ratings of teaching and the course
Control for Non-instructional Factors

- Control for possible sources of bias
  - Student background preparation
  - Student motivation
  - Student work habits
  - Workload/difficulty
  - Class size
  - Academic discipline
**THINK-PAIR-SHARE**

- Does your instrument include summary items?
  - If so, how are they worded?
- Does it control for non-instructional influences?
  - If so, how is that done?
IDEA SRI System

- Two summary items

<table>
<thead>
<tr>
<th>Overall, I rate this instructor an excellent teacher.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Your Average:</strong></td>
</tr>
<tr>
<td>[43]</td>
</tr>
<tr>
<td><strong>Converted Average Comparison:</strong></td>
</tr>
<tr>
<td>[50]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Overall, I rate this course as excellent.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Your Average:</strong></td>
</tr>
<tr>
<td>[3.3]</td>
</tr>
<tr>
<td><strong>Converted Average Comparison:</strong></td>
</tr>
<tr>
<td>[3.7]</td>
</tr>
</tbody>
</table>

Footer - Presentation Title
• IDEA Controls for
  – Student background preparation
  – Student work habits
  – Student motivation
  – Course difficulty
  – Class size
Summary Evaluation of Teaching Effectiveness

Summary

Your Average: 3.5
Converted Average Comparison: 37

Progress on Relevant Objectives

Your Average: 3.1
Converted Average Comparison: 30

Ratings of Summative Questions

Your Average*: 3.8
*Average of Excellent Teacher and Excellent Course
Converted Average Comparison: 43

Description of Course and Students

Course Description

<table>
<thead>
<tr>
<th>Amount of coursework</th>
<th>Your Raw Average Scores</th>
<th>Compared to Group Averages</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| As a rule, I put forth more effort than other students on academic work. |                         |                           |
|                                                                      | 3.9                      | 59                        |

Student Description

| I really wanted to take this course regardless of who taught it. | Your Raw Average Scores | Compared to Group Averages |
|                                                               |                         |                           |
|                                                                | 3.8                      | 46                        |

<table>
<thead>
<tr>
<th>Difficulty of subject matter</th>
<th>Your Raw Average Scores</th>
<th>Compared to Group Averages</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| When this course began I believed I could master its content. | Your Raw Average Scores | Compared to Group Averages |
|                                                              |                         |                           |
|                                                              | 3.7                      | 43                        |

| My background prepared me well for this course's requirements. | Your Raw Average Scores | Compared to Group Averages |
|                                                             |                         |                           |
|                                                             | 3.6                      | 44                        |
Provide Validity Evidence

• Evidence based on content
• Evidence based on internal structure
• Evidence based on relations to other variables
• Evidence based on consequences of feedback
  – Ultimately, validity is tied to use
Concurrent Validity

- Student ratings of instruction are positively correlated with:
  - Instructor self-ratings
  - Ratings by administrators
  - Ratings by peers
  - Ratings by alumni
  - Ratings by trained observers
  - Review of course materials
Provide Reliability Evidence

• SRI highly consistent:
  – In students’ own ratings
  – Across students in the same class
  – Same instructor across multiple courses

– Should provide:
  • Reliability coefficient
  • Standard error of measurement
Inter-Class Reliability as a Function of Number of Classes Rated

Item 41-Excellence of Instructor

Number of classes vs. Inter-Class Reliability

Number of classes:

- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 10
- 11
- 12
- 13
- 14
- 15

Inter-Class Reliability:

- 0.0
- 0.2
- 0.4
- 0.6
- 0.8
- 1.0
Provide Comparative Data

- Ratings
  - tend to be negatively skewed
  - vary by academic disciplines
- Standard scores provide a comparison
## Summary Evaluation of Teaching Effectiveness

### Overall Ratings

### Progress on Relevant Objectives

<table>
<thead>
<tr>
<th>Objective</th>
<th>Your Average</th>
<th>Your Average Comparison</th>
<th>Percent of Students Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gaining a basic understanding of the subject (e.g., factual knowledge, methods, principles, generalizations, theories)</td>
<td>3.2</td>
<td>29</td>
<td>33% 33% 67% 0% 0%</td>
</tr>
<tr>
<td>Developing knowledge and understanding of diverse perspectives, global awareness, or other cultures</td>
<td>2.9</td>
<td>25</td>
<td>22% 56% 22% 0% 0%</td>
</tr>
<tr>
<td>Learning to apply course material (to improve thinking, problem solving, and decisions)</td>
<td>4.1</td>
<td>49</td>
<td>11% 11% 78% 0% 0%</td>
</tr>
<tr>
<td>Developing specific skills, competencies, and points of view needed by professionals in the field most closely related to this course</td>
<td>4.6</td>
<td>58</td>
<td>0% 11% 89% 0% 0%</td>
</tr>
<tr>
<td>Acquiring skills in working with others as a member of a team</td>
<td>3.1</td>
<td>31</td>
<td>33% 23% 44% 0% 0%</td>
</tr>
</tbody>
</table>
Standardize Administrative Procedures

• Online systems enhance standardization
• Instructor should not be present
• Inform students of standard procedures
• Ensure student confidentiality
Myths and Misconceptions Surrounding SRI (Hour 2)

- SRI and pedagogical study
- Reliability Issues
- Validity Issues
- Myths
- Biases
- Personal factors
The Rise of Pedagogy

• The effectiveness of teaching and learning is recognized as critical to college effectiveness
  – Wiley/Jossey Bass has 175 books dedicated to pedagogy in higher ed over the last 10 years
  – Every major accreditor requires evidence of effective teaching

• Teaching and learning is a legitimate area of research
  – Dedicated Journals (e.g., Journal on Centers for Teaching and Learning)
  – Professional Organizations (e.g. POD)
How Does SRI Fit?

• Can students provide valuable information to guide faculty growth? (Or: Do the students deserve a voice?)
  – The physician analogy
• Observed teaching behaviors
• Perceptions of learning that occurred
• Environmental or student factors
• Not things like qualifications of instructor, importance of curriculum, personal attributes
Myth: Low Reliability

• Is student feedback consistent? A resounding “It depends”
  – Is the SRI tool well designed, or is it a camel?

• Well designed SRI have a great deal of reliability
  – Individual student ratings over time
  – Student ratings as a group
  – Instructors over multiple course

• More ratings lead to more reliability
  – Response rates, or “You have a tail”
  – Holistic assessment
Myth: Low Validity

• Do they measure what we say they measure?
• Well-designed SRI correlate highly with:
  – Actual achievement in the course
  – Instructor self-ratings
  – Other external ratings such as peers, administrators, trained observers
• So what’s the issue?
  – These data are noisy!
  – Exceptions will happen (holistic assessment)
• Let’s explore further with some common myths
Myth: Tough Teachers Get Punished

- Corollary: Easy teachers get better evaluations. Can you ‘buy’ good evaluations with an A?
- Do students care about learning, or grades?
  - Cynical, yes, and insulting to students
  - Warner’s offer of an A at the beginning of his class
- Research evidence:
  - The highest ratings go to instructors whose courses are challenging and who set high goals for students
  - SRI are only weakly related to student grade expectations.
  - Improving teaching is MUCH more impactful
Myth: Only the Angry or the Loving Respond to SRI

- High achieving students are much more likely to respond to an SRI
- Poorly performing students tend not to respond
  - Some logic to this result
- Relationship between overall ratings and response rates tend to be quite low
  - If students were motivated to respond to bad instructors, the response rates should be higher
Myth: “These Kids Today...”

- Is there a generational effect for Millennials, in that they are more punishing if they don’t get their way?
  - This ignores the trend of the Traditional Student, but...
- However, what does the research show?
  - Over last last 10 years, average overall ratings of both the instructor and the course have steadily increased
- Interpretations are numerous, but we argue that the increase in emphasis on teaching over the last 15 years is showing results
Personal Factors

• Are factors unrelated to teaching influencing perception of teaching effectiveness?
  – Gender, race, age, charm, attractiveness, humor

• Is there bias? Of course.
  – Human perception is full of bias
  – Does not negate the ability of students to provide feedback about teaching and learning
  – If students are disqualified due to bias, then any form of human assessment is disqualified

• The question is not “Is there bias?” but instead “How do we use biased data?”
Gender Bias

• Influence of gender is only weakly related to SRI outcomes overall
  – Slight interactive effects tend to favor women over men (female students tend to rate female instructors more highly)

• IDEA data finds remarkably similar scores between men and women (N=30k)

• Research laboratory is different than an actual classroom
  – Influence of time actually tends to negate any initial biases (Contact Theory; Allport)
Racial Bias

• Same caveats apply: Of course there is race bias
• Very few studies of race bias in the classroom as related to SRI
• Those that do exist are inconclusive, and point to the same trend as gender: The effect is not very strong over time
• We will talk about how to use potentially biased data in a bit
• IDEA does not track race information, but would love to have a client partner for research
Other Bias

- Other characteristics do not seem to matter much, such as personality, age, attractiveness, etc.
- Factors that do seem to matter might actually contribute to more effective teaching, such as:
  - Energy/Enthusiasm
  - Positive Self-Esteem
  - Humor
  - High Organization
Bias Checks

• Bias Exists!
• Central tendency measures obscure individual instances
  – It is certainly possible a class/dept/college is a legitimate source of bias
• Check your unit’s data, look for evidence
• Adjust scores accordingly if necessary
• BE CAREFUL – some trends are legitimate, and are truly sources of good data for teaching improvement
So What’s the Issue?

• IF you believe the research, then why is there so much passion and controversy surrounding student ratings/course evaluations?

• We find several factors contributing, which are interrelated:
  – Evaluation vs. development
  – Misinterpretation of messy data
  – Overreliance on SRI data
  – Underutilization of SRI data
Use of SRI Data for Development

• Holistic Analysis
  – Look for patterns over time

• Focus on development when looking at teaching improvement. Evaluation of faculty performance is a different process
  – This may be the single biggest issue surrounding SRI

• Be consistent

• Be communicative about the process

• Be encouraging about innovation and engagement in teaching
Effective Use of SRI for Faculty Evaluation (Hour 3)

• Decide on purpose: summative vs. formative
• Include multiple measures
• Recognize limitations of each measure
• Encourage good response rates
• Train faculty raters and administrators
• Don’t “set it and forget it”
What’s the Purpose?

- SRI over-emphasized for summative, and
  - Under-utilized for formative purposes

- If summative, collect ratings across multiple courses
- If formative, use the ratings to improve
Include Multiple Measures

• Combined data can increase reliability
• Be selective
• Source should have actual knowledge of what is being measured
Measuring Teaching Effectiveness: Include Multiple forms of Assessment

- Instructional Delivery
  - Students
- Instructional Assessment
  - Students, Peers
- Balanced Plan For Summative Evaluation
- Course Management
  - Administration
- Content Expertise
  - Peers
- Instructional Design
  - Students, Peers
- Learning Outcomes
  - Students, Peers
Dimensions and Information Sources for Assessing Teaching Performance

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Instructor</th>
<th>Student</th>
<th>Peers</th>
<th>Administrator</th>
</tr>
</thead>
<tbody>
<tr>
<td>Content</td>
<td>YES</td>
<td>NO</td>
<td>YES</td>
<td>MAYBE</td>
</tr>
<tr>
<td>Instructional design</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>NO</td>
</tr>
<tr>
<td>Instructional delivery</td>
<td>NO</td>
<td>YES</td>
<td>NO</td>
<td>NO</td>
</tr>
<tr>
<td>Assessment</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td>Course management</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>YES</td>
</tr>
</tbody>
</table>
Recognize the Limitations of Each Measure

- Instructor self-ratings lack objectivity
- Peers differ in philosophies, time commitment
- "Friends as peer reviewers" should be avoided
- Is the external recognition truly significant?
- Was professional development legitimate?
- Was contribution truly exemplary?
Train Faculty and Administrators

• Emphasize holistic assessment
• Explain that all ratings have error
• Explain limitations of each measure
• Help interpret SRI class reports
• Show value of SRI aggregate data
• Provide consultation for improvement
Don’t ”Set it and Forget it”

• Practice feedback loop
• Provide incentives for faculty efforts at improvement
• Schedule periodic reviews of evaluation system
WE ARE HERE TO HELP

• One on-one consulting services
• Unlimited training
• Unlimited support
• Documentation and training modules
• Implementation Assistance
Which form/instrument do I choose?

- **Diagnostic Feedback**
  - Feedback on teaching methods and learning outcomes
  - Overall summary measures
  - Focus on improving teaching and learning

- **Learning Outcomes**
  - Feedback on learning outcomes
  - Overall summary measures
  - Commonly used in clinical classes, practicums, and labs etc.

- **Teaching Essentials**
  - Feedback on methods associated with excellent teaching
  - Overall summary measures
  - Useful for adjunct and seasoned faculty,
Questions?