Task Force to Revise Student Rating of Instruction: Preliminary Report to Senate
March 2016

Task Force Members:
Mercedes Bentley (student member)
Christine Boone (Music)
Evelyn Chiang (Psychology; co-chair)
Ann Dunn (Humanities)
Jinhua Li (International Studies)
Karin Peterson (Sociology & Anthropology; co-chair)
Mike Ruiz (Physics)
Samer Traboulsi (History)
Melissa Himelein (Center for Teaching & Learning, ex-officio)
Jeff Konz (Dean of Social Sciences, ex-officio)

Charge:
In the fall of 2015, UNC Asheville’s Faculty Senate created a Task Force to Revise Student Rating of Instruction, charging us to “study student rating of instruction and make recommendations for changes to the Student Rating of Instruction (SRI) process and instrument.” Our task was outlined as follows:

1. review the state of the science on SRIs;
2. examine SRIs used by institutions with similar missions and other academic leadership institutions;
3. submit a preliminary report to the Faculty Senate on findings regarding tradeoffs, strengths, and shortcomings of various approaches to SRIs; and further to make recommendations concerning the UNC Asheville SRI content and implementation.

Process:
In the fall of 2015, this task force convened and elected two co-chairs: Karin Peterson and Evelyn Chiang. Each member of the task force then read key background articles on SRIs. These articles were identified by Melissa Himelein, who has reviewed extensive work in the area. The task force then identified three areas of focus for further reading by subcommittees. The three areas of focus for our readings included 1) administration of student evaluations; 2) best practices in the use and interpretation of evaluations; and 3) current strategies in the design of evaluation instruments. Subcommittees reviewed literature in each area of focus and generated recommendations that were then discussed by the task force as a whole. The present document contains a brief summary of issues highlighted in the current literature, as well as a number of recommendations regarding the design of SRIs, the administration of SRIs, the use and interpretation of SRIs, and next steps.

Please note: Henceforth, we will use the term SFI (Student Feedback on Instruction) rather than SRI (Student Ratings of Instruction) in this report. We propose that this new terminology be adopted for reasons detailed in the first recommendation related to SFI items (page 2).
Summary of Issues and Concerns:
In our review of the research and national discussions, the following concerns about SRIs emerged; many of these concerns have been echoed by colleagues on our campus, either formally or informally:

- Issues related to the wording and content of the quantitative and qualitative questions asked to evaluate teaching performance
- Concerns that SFIs typically under-recognize the variation in styles of teaching and in desired learning outcomes
- Issues related to the administration of SFIs (online vs. paper methods of evaluation, time period in which SFI data are collected, achieving response rates that are sufficiently high to produce reliability)
- Issues related to the usefulness or meaningfulness of SFI numeric scores
- Issues related to the use of SFIs in evaluating faculty teaching, and proposals for multi-dimensional approaches to evaluating teaching
- Issues of gender and racial/ethnic bias in SFI student responses
- Issues related to large introductory required ILS courses. Some research also indicates a bias towards some disciplines or subject matter

RECOMMENDATIONS

As a result of our readings and discussions, we make recommendations pertaining to each of the following issues: the content of SFI items, the administration of SFIs, the use and interpretation of SFIs, and steps to follow in moving forward.

A. Recommendations Pertaining to SFI Items:

Trends in SFI development nationally are suggesting that more attention be given to aligning SFI measures with desired student learning outcomes. In addition, variations across disciplines and instructors in their styles of teaching create challenges in using just one set of SFI items. We make the following recommendations with these issues in mind.

1. The university should adopt consistent terminology in reference to the SFI on all SFI forms and in all communication to students, faculty, administrators, and staff. We recommend adoption of the terminology, “Student Feedback on Instruction” (SFI). In addition, we recommend that questions on SFI forms clearly specify when the student is providing feedback on the Course or on the Instructor.
   - The variety of terms used to address SFIs currently (e.g., student ratings of instruction, student evaluations, course evaluations) is confusing.
   - Research suggests that the language of feedback removes the sense of the student serving as “expert evaluator,” and also removes “ratings” language, which is too commonly associated with consumer surveys. Changing the
terminology further emphasizes the formative purpose of the information students supply.

2. UNC Asheville’s SFI should include multiple question types. While universal questions might be required of all SFIs, departments/majors and faculty should be able to tailor their SFI survey to the course they teach. SFIs should be made up of:
   • Universal questions (a common core of items used on all SFIs across campus, in order to allow for comparisons across divisions and departments)
   • Department- or program-specific questions (in order to reflect department/program student learning outcomes)
   • Course-specific questions that address specific learning outcomes and teaching strategies identified by the course instructor (to be selected from a limited bank of possible items)

3. When possible, survey questions and response scales should be descriptive (hardly ever, occasionally, almost always) rather than evaluative (excellent, fair, poor).

4. Surveys should measure behaviors of faculty or experiences of students that were helpful/unhelpful to their learning.
   • Items are ideally descriptive of behaviors of faculty or experiences of students that were helpful/unhelpful to their learning, as opposed to global strengths and weaknesses; the latter items do not necessarily relate to the ultimate goal of student learning.
   • Items should be criterion-based (e.g., the course was well-organized) rather than norm-based (e.g., in comparison with other courses, this class was well-organized). Surveys should not ask students to compare an instructor to other faculty or a course to other courses, as learning outcomes and teaching strategies vary by course and instructor.

B. Recommendations Pertaining to the Administration of SFIs:

1. The literature suggests that SFIs should be administered during a narrowly defined time range, in order to address issues of consistency across classes. We recommend the last five days of the semester.

2. SFIs should not be administered in courses with fewer than 6 students. This ensures student anonymity.

3. SFIs should continue to be administered online, as the research indicates that online administration is comparable to paper administration in terms of reliability.
4. In order to communicate to students the purpose and importance of student feedback, the University should create a standardized statement to be included on the SFI instrument, as well as in other strategic places that make the value of student feedback more visible. The statement should include information about confidentiality.

5. Faculty should be encouraged to engage in a university-wide established procedure for administering in-class online surveys in order to improve response rates and therefore reliability of results. This procedure should include:
   - The instructor should announce the day when class time will be provided for SFIs, and, if no computer lab is available, ask students to bring a computer, tablet or phone that will allow them to complete the forms during class.
   - The instructor should assign a student to monitor the feedback session.
   - The student monitor should read (out loud to the class) the standardized statement on the purpose and importance of student feedback (standardized by the University; see recommendation 4 above).
   - The instructor should be absent from the classroom while the survey is administered.

6. The University should engage in conversations with students about the importance of their feedback in SRIs – perhaps in the first semester of students’ careers at UNC Asheville. The standardized statement should be shared in orientations and included in the student handbook, or other places where students are likely to seek out such information.

**C. Recommendations Pertaining to SFI Interpretation and Use**

1. While SFIs will always be incorporated into summative merit evaluations of faculty members by Deans, Department Chairs and Program Directors, faculty members should be encouraged to use SFIs as a formative tool for improving teaching. SFIs are one source of feedback for informing faculty of their teaching effectiveness, providing faculty opportunities to address areas of concern. As such, how faculty members respond to SFI scores and comments are important formative aspects of their teaching.

2. The Annual Faculty Record form should include a space for faculty to reflect on their teaching, including addressing how they have used SFIs for their own formative development.

3. Administrators, Department Chairs, and Evaluation Committees (such as PTR and Tenure) should be trained to understand the purpose of SFIs in a larger context, as SFIs relate to other measures of teaching effectiveness (e.g., peer reviews, portfolios, faculty record self-evaluation narratives, chairs’ letters).
4. Administrators, Department Chairs, and Evaluation Committees should receive required training on interpreting both the quantitative and qualitative components of SRIs. This includes understanding confidence intervals and tests of significance, and distinguishing between meaningful and less meaningful qualitative comments.

5. Administrators, Department Chairs, and Evaluation Committees should receive required training on evaluating teaching and reading SFIs in context; this includes contexts shaped by class size, whether a course is a requirement or elective, discipline, course level, and other factors, as well as in the context of persistent patterns of gender and racial bias.
   • While some research concludes there is no definitive pattern of gender or racial bias in student evaluations, other studies do suggest patterns. In addition, there is national concern about the issues that women and racial minorities confront in evaluation processes.
   • All evaluators should also be aware of the current status in the research regarding bias towards some disciplines and subject matter.

6. The university should begin producing and sharing data on gender and race bias in SFIs (Hamilton College does this). This would help all evaluators and faculty understand what the realities look like at our institution and help us formulate strategies for addressing these concerns.

7. Academic Affairs should continue to work on strategies that address gender and racial bias as it relates to teaching (in the evaluation realm, but not necessarily limited to the evaluation realm). The research in this arena produces mixed findings; however, concerns about potential bias are reason enough for the institution to address the issues.

8. Great care should be taken in giving credence to SFI results where the response rate is lower than 66%, a level that research suggests is not reliable. Item means or medians in SFIs for courses with fewer than 8 students should be interpreted with great caution.

9. Item means should always be accompanied by comparison means (such as the home department and division), confidence intervals, and tests of significance. By correlating students’ mean ratings with the class average, we may identify comments that correlate highly as “reliable and representative” and discard comments with low correlations as outliers.

10. SFIs should be read in the context of trends and patterns of the faculty member’s performance across courses and across time.

11. The institution should further explore alternative means for evaluating faculty with a large number of low-enrolled courses.
D. Recommendations Pertaining to Moving Forward:

1. Appoint our committee or a subset of our committee to design and pilot a replacement SFI instrument, and accompanying directives regarding administration and use, which meets the recommendations above. We suggest Summer-Fall 2016, as a time frame for the development of the new instrument. A draft SFI could be piloted by faculty volunteers in Spring 2017.

2. Engage the faculty in a dialogue about the use of SFIs and the proper role they can play in informing us about the experiences of students in the classrooms and in giving us room to formatively address teaching issues, as well as ways that SFIs should not be used. We are sympathetic to complaints that faculty have about SFIs not being valid or useful; nonetheless, we believe that students need to be part of the culture of feedback on teaching at this institution. Addressing how SFIs are to be used, not to be used, improving the SFI instrument itself, and creating a more open culture of feedback as a formative tool should address some of these issues. Creating a means of researching trends in biases will also promote a culture of reasonable use of SFIs.

3. Identify a group to work on refining other means of evaluating and improving teaching through reflection and other means of feedback (e.g., more emphasis on faculty self-evaluation in the annual faculty record, portfolios, peer review, other forms of improving teaching).

4. Identify a working group to address persistent forms of gender/racial/ethnic bias, as well as other forms that are identified by 1) gathering data; 2) creating programming that enhances a university culture that addresses student bias.
References


An investigation of the nature of written comments across 198 separate courses. Suggests comments have value, generally focusing on three dimensions: ideas related to interest generated by course and/or instructor, contribution of course to overall learning, and clarity of teaching.


Blog post from the Rice University Center for Teaching Excellence (CTE), providing overview and synthesis of literature on student and course evaluations. Provides six evidence-based conclusions, including summaries of biasing and non-biasing variables and the relationship between student evaluations and student learning.


Narrated presentation slides and bibliography compiled by Rice University's faculty development center, for use in Rice's recent examination of campus SRIs. Compilation of literature on reliability and validity of SRIs; biases in SRIs; research-validated structures, administration, and analysis of SRIs; and appropriate use.


IDEA (Individual Development and Education Assessment) is a nonprofit organization focused on improving learning in education; one activity of the organization is producing and publishing scholarly reviews of teaching and learning issues. This review examines findings of more than 100 studies of SRIs and summarizes evidence on student and instructor variables related to ratings. The authors also discuss online vs. paper administration and overall usefulness of SRIs.


Investigated the impact of small differences in mean ratings on course evaluations, finding that differences as insignificant as 0.15 (on a 5-point scale) were perceived by faculty members to be important.


A brief review of recent literature with several recommendations: Student ratings are best considered an opportunity for "feedback" rather than evaluation; anonymity should only be granted to comments as long as grades are pending; colleges should withhold grade feedback until ratings are completed (or a formal abstention is requested); margins of error should be considered in reviewing SRIs; multiple sources of feedback on faculty teaching should be considered in assessing faculty.
Summarizes an instrument developed by the author, a professor of economics, assessing proficiencies expected of department majors, e.g., applying existing knowledge, creating new knowledge. Form asked students novel questions, e.g., whether course delivered what promised, effectiveness of instructional materials and pedagogy in developing proficiencies, effectiveness of learning activities in developing proficiencies.

Hativa, N. (2014). Most important behaviors of effective teaching as a basis for designing SRI instruments. Synopsis by Rick Reis of Chapter 1 (Theoretical basis for SRI) of *Student ratings of instruction: A practical approach to designing, operating, and reporting (2nd ed.)*.
Argues that certain items need to appear in SRIs because they contribute to teaching effectiveness: clear, organized, and engaging teaching; classroom interactions with students; and rapport with students.

Summarizes the process followed by a task force to revise SRIs at Thompson Rivers University. Final instrument includes assessment of six teaching dimensions shown to be correlated with student achievement, written in "learner-centered" language, and focusing only on experiences students are “qualified” to evaluate; does not incorporate global statements of assessment (believed to be unhelpful and potentially misleading).

Study suggesting that SRIs using the IDEA Center form do not differ according to type of administration (i.e., paper vs. online). Response rates were higher with paper, but rates were uncorrelated with item means.

Brief presentation on the IDEA Center’s system of SRIs, which emphasizes student ratings of learning (i.e., “How much progress did you make on each of the following student learning outcomes?”) as opposed to evaluation of instructors. Information about the validity of this approach as well as a sample form can be found on the ideaedu.org website.

Another literature review of the validity of SRIs, providing both overlap and update of Benton and Cashin (2012); focus is on studies only since 2000. Dense article that provides somewhat less definitive conclusions than Benton and Cashin.

Argument that from a statistical standpoint, SRIs are problematic. Response rates and response variability are important variables to consider, comparisons of mean scores are inappropriate, and comparisons across types, levels, and sizes of course are unfair. Offers recommendations based on substantive argument and review of the literature.


Review of shortcomings of traditional SRIs. Argues for use of the “Teaching Practices Inventory” to evaluate teaching, a self-report instrument to be completed by faculty.


Brief synopsis of evidence-based recommendations regarding what factors to evaluate, when to evaluate, where to evaluate, and how to use SRIs.


Based on review of SRI literature, suggests a method for evaluating written comments on SRIs: Quantitative ratings would first be compared with class averages to separate out personally biased or outlier comments, with review then focusing on majority opinion.