Using Student Ratings of Instruction in Faculty Evaluation

Although the IDEA Student Ratings of Instruction instrument (SRI) is designed to be primarily a formative feedback tool—that is, it provides feedback to help instructors improve their instruction—we recognize the need for evaluating faculty teaching performance for purposes of tenure and promotion and related personnel decisions. The use of student feedback as the sole measure for evaluating faculty is not an appropriate use of student ratings and often leads to faculty resentment toward the process as well as reluctance to use student feedback to reflectively improve instruction. Therefore, the approach suggested here uses the SRI along with other perspectives to create a comprehensive and balanced system for evaluating faculty. This approach uses student feedback based on SRI data, and evaluation of three instructional criteria that includes instructor self-evaluation. This model is provided as an example of how an institution might use SRI data in a comprehensive system and not as a specific, recommended process. It is hoped that institutions will use this model to develop one that fits specific institutional needs while embracing the balanced approach presented here. Whatever direction you take, keep in mind that faculty evaluation should be as transparent as possible—beginning with faculty involvement in the creation of the evaluation process. Once procedures are in place, make sure faculty are well aware of how they are being evaluated, so they can work toward being as effective as possible—as described in the model—as well as not being caught off guard by any of the criteria involved.

Balanced Evaluation of Teaching (BET) Model

Student Feedback

SRI Summary Score

How does the adjusted overall score (50% Progress of Relevant Objectives and 50% overall ratings) converted average for the chosen comparison group compare to others via the **Converted Average Comparison**?

Assessment of Instructional Effectiveness

Self-evaluation | Peer evaluation | Review of evidence

to assess

Instructional effort

Is there evidence that this instructor is actively working to improve instruction and student learning through reflective practice?

Student Engagement

Is there evidence that this instructor effectively engages students in and out of class?

Design for Learning

Is there evidence that assessments, assignments, learning activities and other learning materials are effective and appropriate, and that courses are purposely designed?

Student Feedback

Student feedback is codified by use of the instructor's Summary Score on the IDEA SRI (a combination of student ratings of Progress on Relevant Objectives and the two Overall ratings) over the period of time being evaluated. This score provides a measure of students' ratings of their progress on relevant objectives along with their more subjective interpretation of the effectiveness of the course and instructor, and therefore, serves as a summary measure of student feedback. Then the Converted Average Comparison is used to compare the score to others in the comparison group. This allows you to determine if an individual faculty member's score is much lower, lower, similar, higher, or much higher than others in the comparison group (see Calculating an SRI score across courses and time below for more details), and this categorization provides the primary basis of student feedback for evaluation. So in this model, student feedback provides an important starting point for evaluating faculty, but as described next in the Assessment of Instructional Effectiveness details, it is just the start. An examination of other evidence is used to create a balanced and comprehensive analysis of teaching quality. However, "student voice" is an essential part of getting a clear picture of the quality of an individual's instruction. See Challenging Misconceptions About Student Ratings of Instruction for more about the importance and validity of including student ratings.

Assessment of Instructional Effectiveness

The three criteria for instructional effectiveness—Instructional Effort, Student Engagement, and Design for Learning—are assessed several ways depending on the policies and practices of the academic unit. These include a self-evaluation, optional peer observation, and a review of evidence. This information is used collectively to make evaluations of the instructional effectiveness of the instructor.

Methods of Evaluation

Self-evaluation

As part of this process, faculty complete a self -evaluation that addresses the three criteria (Instructional Effort, Student Engagement, and Design for Learning) as well as analysis and responses, as appropriate, to student feedback. This may include explanations that some courses in the review period were taught for the first time or under extenuating circumstances that may have led to lower than desired SRI scores for instance. It also includes a thoughtful response to student ratings in specific courses and how the faculty member responded to the feedback. Faculty also provide evidence for changes to course materials, assignments, class activities and such in response to student feedback, student performance on assessments, suggestions from mentors or others and attempts to apply best practices in instruction. Effective self-evaluations include a reflective summary of overall improvements and lessons learned during the review period.

Peer Evaluation

This optional method of evaluation refers specifically to **observation of classroom instruction** conducted by fellow faculty and/or chairs and deans.

Peer observation of instruction should be designed to assess instructor behaviors that support student learning including assessments of the three criteria of instructional effort, student engagement, and to the degree possible, course design elements such as class learning activities. Numerous rubrics and guidelines have been developed for conducting peer evaluations and are available online. They typically include observations of faculty preparedness and organization, engagement of students, effective use of learning strategies, and effective communication techniques among others and often encourage a pre-class review of class goals and conversation with the instructor. As with any data collection, the more the better. So for evaluation purposes, more than one observation—preferably from more than one observer—will yield the most useful information.

Review of Evidence

The chair, dean, and/or P&T committee reviews the self-evaluation, summaries of peer evaluations, and course materials provided (and may request copies of other materials such as exams and syllabi) to determine how well the instructor is meeting each of the three criteria. SRI data may also be reviewed to provide evidence for some criteria. Student comments, for example, may provide evidence for student engagement and trends in SRI scores over time may provide evidence for improvement. Institutions may develop their own rubrics to use in evaluating the criteria.

Evaluation Criteria

Instructional Effort

Faculty are assessed for their instructional effort including improvement over time in SRI scores, responding to feedback from students and peers by adjusting instruction, evidence of modifying instruction to improve student learning outcomes, evidence of trying new things, and whatever other evidence an institution believes is important to assess. In essence, this is a determination of whether nor not an instructor is engaging in reflective practice—the thoughtful consideration of evidence to improve student learning. The combination of these evidences are synthesized into a single assessment of instructional effort. Instructors provide a self-assessment as part of this criterion that is used in the evaluation. The chair or evaluation committee then reviews the available evidence to determine the appropriate placement of the faculty based on the evidence.

Student Engagement

Student engagement affects learning, retention, and student commitment to programs, and this criterion is an attempt to measure the quality of an instructor's contribution to engaging students in learning. Student engagement is a term often used to cover a wide array of concepts but is used here to address those instructor efforts which engage students with course content, and hence provide motivation for learning, as well as interpersonal connections between student and faculty and the larger learning community. Content-related engagement includes effective use of teaching strategies that motivate and draw students into learning including the instructor's own interaction with students about course content that motivates and challenges them. It also includes the direct interactions that an instructor has with students that promote learning and student success. Is the instructor approachable and available to students during and outside of class, for instance?

Design for Learning

This criterion is an attempt to assess course design elements within control of the instructor including assessments, assignments, syllabi, and other elements important to a campus. Are courses purposefully designed to achieve the outcomes most desired in a course? Some institutions might look for evidence that courses are backward designed as a part of this assessment. For online courses, this assessment might include evaluations of course organization, designing for instructor presence and more.

Other measures considered in evaluation criteria could include advising ratings; departmental or institutional learning outcome assessments; or other criteria of importance to the unit. For the criteria of interest, campuses could create rubrics, or borrow existing ones, to further operationalize and standardize the assessment of each.

Putting it all together

Two variations of using these data for evaluation are provided here. The Four Category version allows review committees or chairs/deans to place faculty into one of four categories for more discriminate evaluations for tenure and promotion and/or merit pay needs or for institutions that have other incentives in place to encourage improvement of instruction:

Four Categories:	Category 1 *Improvement needed*	Category 2 *Improving*	Category 3 *Performing*	Category 4 *Exemplary*

The second Two Category variation is a simplified tool to determine if a faculty member is

Two Categories:	*Performing as Expected*	*Needs improvement*

This dichotomous approach may fit the needs of some chairs or deans who are making annual reviews, for instance, and just want to ensure a faculty member is performing and will be retained.

A balanced approach. The matrix is designed so that student feedback via the SRI is taken into consideration, but other sources of evidence are considered so that a balanced view of faculty instructional effectiveness is obtained. SRI scores could be fairly low, placing an instructor in the lower comparison group, but the instructor is placed in a higher evaluation category based on other criteria. This might include, for example, the understanding that an instructor tried flipping their classrooms for the first time with mixed results. In their evaluation, they should be rewarded for trying even though SRI scores may not have shown a positive reaction from students. It is important to keep this example in mind when developing your own system, as a system that is too rigid in its approach may limit or completely stifle innovation in the classroom. In all cases, an instructor with very low SRI scores would need strong evidence of effort and improvement to receive a more favorable evaluation than one with high SRI scores. The result, then, is a fair and comprehensive evaluation that considers multiple data points. Student feedback is part of the formula but only one part of a multifaceted evaluation.

Four Category Evaluation Tool

Assessment of Instructional Effectiveness

Student Feedback	Instructional Effort | Student Engagement | Design for Learning	Evaluation Category
SRI score is **lower or much lower** than others in the same comparison group.	There is **little to no evidence** that this instructor is actively working to improve courses through implementing or improving SRI teaching methods, engaging in course redesign, consulting with instructional designers, attending pedagogy workshops, etc. and little evidence of effective engagement of students.	**Category 1** Improvement Needed
	There **is evidence** that this instructor is actively working to improve courses through implementing or improving SRI teaching methods, engaging in course redesign, consulting with instructional designers, attending pedagogy workshops, etc. and working to improve engagement of students.	**Category 2** Improving
	There is **evidence of exemplary effort** to improve courses through implementing or improving SRI teaching methods, engaging in course redesign, consulting with instructional designers, attending pedagogy workshops, etc. and working to improve engagement of students **resulting in noted progress in course outcomes and/or SRI scores over time** (though still in the lower range).	**Category 3** Performing
SRI score is **similar** to others in the same comparison group.	There is **little to no evidence** that this instructor is actively working to improve courses through implementing or improving SRI teaching methods, engaging in course redesign, consulting with instructional designers, attending pedagogy workshops, etc. and little evidence of effective engagement of students.	**Category 2** Improving
	There **is evidence** that this instructor is actively working to improve courses through implementing or improving SRI teaching methods, engaging in course redesign, consulting with instructional designers, attending pedagogy workshops, etc. and working to improve engagement of students.	**Category 3** Performing
	There is evidence of **exemplary effort** to improve courses through implementing or improving SRI teaching methods, engaging in course redesign, consulting with instructional designers, attending pedagogy workshops, etc. resulting in **noted progress in course outcomes and/or recognition by external groups of pedagogical innovation**.	**Category 4** Exemplary
SRI score is **higher or much higher** than others in the same comparison group.	There is **little to no evidence** that this instructor is actively working to improve courses through implementing or improving SRI teaching methods, engaging in course redesign, consulting with instructional designers, attending pedagogy workshops, etc. and little evidence of effective engagement of students.	**Category 3** Performing
	There **is evidence** that this instructor is actively working to improve courses through implementing or improving SRI teaching methods, engaging in course redesign, consulting with instructional designers, attending pedagogy workshops, etc. and working to improve engagement of students.	**Category 4** Exemplary

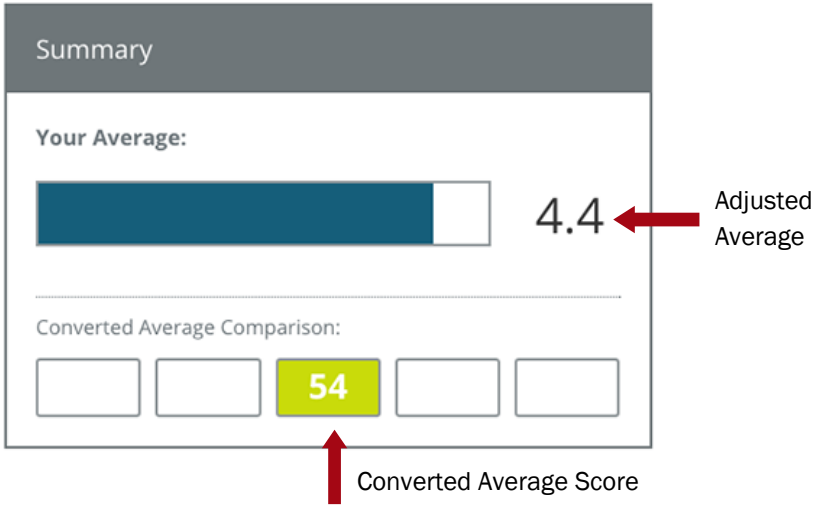
Two Category Evaluation Tool

Assessment of Instructional Effectiveness

Student Feedback	Instructional Effort | Student Engagement | Design for Learning	Evaluation Category
SRI score is **lower or much lower** than others in the same comparison group.	There is **little to no evidence** that this instructor is actively working to improve courses through implementing or improving SRI teaching methods, engaging in course redesign, consulting with instructional designers, attending pedagogy workshops, etc. and little evidence of effective engagement of students **or such efforts need improvement**.	Needs Improvement
	There **is strong evidence** that this instructor is actively working to improve courses through implementing or improving SRI teaching methods, engaging in course redesign, consulting with instructional designers, attending pedagogy workshops, etc. and working to improve engagement of students such that **improvement is evident over time**.	Performing as Expected
SRI score is **similar** to others in the same comparison group.	There is **little to no evidence** that this instructor is actively working to improve courses through implementing or improving SRI teaching methods, engaging in course redesign, consulting with instructional designers, attending pedagogy workshops, etc. and little evidence of effective engagement of students.	Needs Improvement
	There **is sufficient evidence** that this instructor is actively working to improve courses through implementing or improving SRI teaching methods, engaging in course redesign, consulting with instructional designers, attending pedagogy workshops, etc. and working to improve engagement of students.	Performing as Expected
SRI score is **higher or much higher** than others in the same comparison group.	There is **little to no evidence** that this instructor is actively working to improve courses through implementing or improving SRI teaching methods, engaging in course redesign, consulting with instructional designers, attending pedagogy workshops, etc. and little evidence of effective engagement of students.	Needs Improvement
	There **is evidence** that this instructor is actively working to improve courses through implementing or improving SRI teaching methods, engaging in course redesign, consulting with instructional designers, attending pedagogy workshops, etc. and working to improve engagement of students.	Performing as Expected

SRI Score: calculating an SRI score across time and courses

To determine which SRI category a faculty member is in for the selected period of time, gather the Converted Average Comparison Summary score for each course in the selected time period. For someone teaching four courses two semesters a year, this would be 24 scores for a typical three-year period. Then calculate the **median score** for this period of time (this simple spreadsheet is available for your use in calculating the median score). The median is the preferred measure since it is the midpoint of scores ranked from highest to lowest. Therefore, it is less likely to be affected by extreme scores. The mean, on the other hand, could be pulled in either direction by extreme scores. Since all instructors have the occasional course where things go awry, often in circumstances beyond their control, the median converted adjusted score is the best measure to summarize the converted average over a period of time.



This median score then, their representative Converted Average Comparison Summary score for the period of time being evaluated, is used to determine which of the five categories in which to place an instructor in the first part of the evaluation model: much lower, lower, similar, higher, or much higher than others in the comparison group. This chart shows you which median score range is associated with which category.

Converted Score	Percent Distribution	Label
37 or lower	Much lower 10%	Much Lower
38 to 44	Lower 20%	Lower
45 to 55	Similar 40%	Similar
56 to 62	Higher 20%	Higher
63 or above	Much higher 10%	Much Higher

Other considerations. There should be a minimum of six courses in the review period with the goal of having as many courses as possible (see Technical Report 18). IDEA recommends the use of **adjusted scores** over raw scores because they take into account and adjust scores based on factors beyond the control of instructors. See Adjusted Scores at a Glance for more about adjusted scores.

Weighting courses. There may be times when a review of an instructor's work in a particular course may be more important than other courses. For instance, if you have an important introductory course in your department that is vital for recruiting students to your major or ensuring large enrollments in the course is important to the department bottom line, then teaching this large course is a very important part of an instructor's load compared to the other, smaller, upper level courses they teach. In such cases, you might consider ways of weighting some courses over others in your evaluations.