



## **Student Rating of Teaching Effectiveness**

### **Recommendations for Use by Departments and Faculty**

**S. Cross, M. Dooris, S. Weinstein**

**April 2004**

### **Introduction**

This paper is intended to provide recommendations for the interpretation of the SRTE to departments and faculty as well as alternative tools designed to improve teaching.

### **I. History of the Penn State SRTE**

In April 1985, Penn State's University Faculty Senate mandated the use of what has come to be known as the SRTE – the locally developed Student Rating of Teaching Effectiveness. Penn State researchers have examined various aspects of this in-house instrument, but a complete psychometric evaluation of the Penn State SRTE, to our knowledge, has never been conducted. However, the questions raised about the Penn State SRTE tend to be similar to those asked about student ratings in general. There is a large, well-established, and relevant body of research on those questions for instruments similar to the SRTE.

### **Assessment of Similar Instruments**

Research on student ratings goes back at least to the 1920s, and can help address many persistent questions. For example, some faculty and administrators believe that student ratings are mostly popularity contests that have little to do with learning. Such a hypothesis can be readily tested – for example, by analyzing student ratings and grades in large, multiple section courses (such as introductory chemistry or physics) taught by several instructors who use common textbooks and give identical examinations. Literally hundreds of such studies have, in fact, been conducted. A review of this literature, relevant to the Penn State SRTE, is available elsewhere ([www.psu.edu/president/pia/planning\\_research/reports/index.htm](http://www.psu.edu/president/pia/planning_research/reports/index.htm)) and need not be repeated here; highlights are as follows.

1. Professionally developed instruments can be both valid and reliable. Reliability – the extent to which a scale produces the same results on repeated trials – can be readily tested and is quite high for such instruments. It is often in the range of .8 to .9. Validity – the extent to which ratings measure what they are intended to measure – is more difficult to test, but students' ratings have been positively correlated with other indicators of teacher competence, such as peer reviews and measures of student learning.

2. Locally developed student rating metrics are often of questionable reliability and validity, according to the same body of research. Validating such instruments is a significant undertaking.
3. The single most frequently researched issue on student ratings is this: Do grades that students receive in a course correlate with their ratings of the course and/or instructor? According to the literature cited above, various studies have found mildly positive, zero, and even mildly negative correlations. Substantial evidence does show that students are discriminating judges of instructional effectiveness.
4. Upper-division and graduate students tend to rate instructors somewhat more favorably than do lower-division students.
5. The evidence on many other factors commonly suggested as possible influences – gender, class size, faculty rank, and so on – tends to be mixed.

### **Assessment of the Penn State SRTE**

Previous studies have focused on the empirical analysis of means. A recent in-house assessment of the Penn State SRTE is an analysis of fall 2002 data, completed in September 2003. The complete text, tables, and graphs are available online at [http://www.psu.edu/president/pia/planning\\_research/reports/fall\\_02srte.html](http://www.psu.edu/president/pia/planning_research/reports/fall_02srte.html). Again, it is not desirable to repeat that report in this document. In brief, however, the findings are as follows.

To the extent that parallels can be drawn between analyses of the SRTE and those in the research literature, the findings tend to be similar.

On the question of whether instructors who are easier graders get higher student ratings, the data for Penn State's SRTE are consistent with the research literature. Ratings on the SRTEs two global items about quality of the course and quality of the instructor are mildly and positively correlated with grades. That correlation only explains about five to ten percent of the variation in SRTE scores. (It can be argued that such a connection is reasonable, if one believes that more effective teaching, higher grades, and superior student ratings should go hand-in-hand.)

Ratings are also mildly and positively related to course level. On the SRTEs seven-point scale, graduate courses are rated about .3 higher than undergraduate courses, on average. Upper-division undergraduate courses are rated about .1 higher than lower-division undergraduate courses.

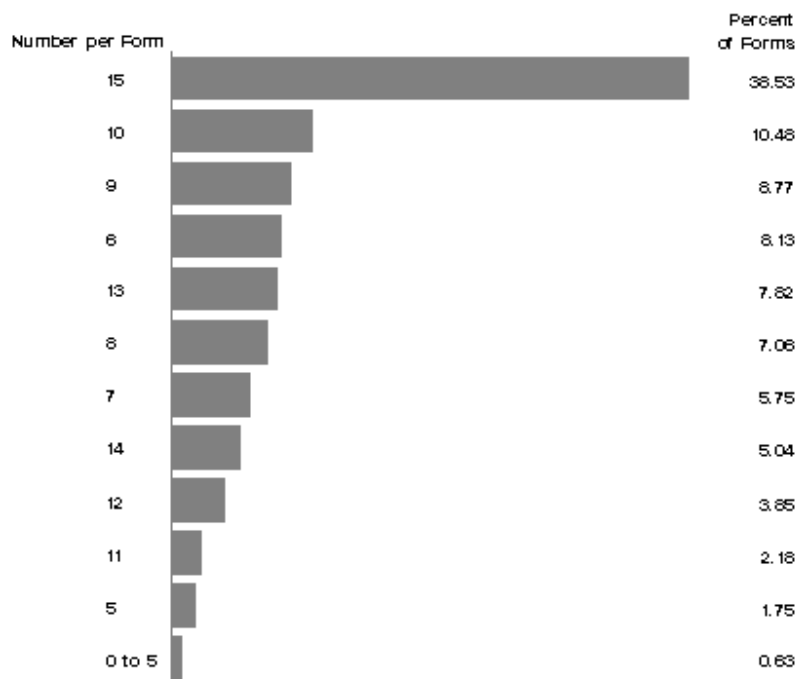
There is some variation on two other dimensions. First, teaching assistants on average score about .3 lower than regular faculty on the SRTEs seven-point scale. Second, there is also some variation across colleges and campuses. However, the average scores by college are all within the range of plus or minus .4.

There is little to no variation among averages when the data are compared in other obvious ways (by rank, gender, appointment type, and so on). For example, for averages at the university level, there is no difference for assistant professors versus associate professors versus professors. There is also essentially no difference in the average scores between standing and fixed-term faculty.

## II. Current Patterns of Use

The SRTE was designed to enable departments to adapt it for their own courses. There are currently 180 unique forms in use with 10% of the departments using more than one form depending on the course. The database of “B” items allows departments to design multiple forms of the SRTE by selecting up to 15 items for each form. The following pareto chart and table show about 40% of the forms ordered take advantage of all 15 questions. One-third of the forms list fewer than 10 questions.

**Number of Questions per Form**  
(ordered by frequency)



**Number of Questions per Form (percentage distribution)**

<b>Number per Form</b>	<b>0 –5</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>9</b>	<b>10</b>	<b>11</b>	<b>12</b>	<b>13</b>	<b>14</b>	<b>15</b>
Percent	2%	8%	6%	7%	9%	10%	2%	4%	8%	5%	39%
Cumulative Percent	2%	10%	16%	23%	32%	42%	44%	48%	56%	61%	100%

Although there are 177 questions in the databank, most forms contain questions from a subset of 32. The most popular questions in current use come from a variety of instructor-related categories. A complete list of questions ranked by frequency of use can be found in Appendix IV.

<b>Top Fifteen Questions</b>
(1) 1 Rate the clarity of the instructor's presentations
(2) 63 Rate the instructor in terms of his/her preparation for class
(3) 49 Rate the adequacy of the instructor's knowledge of the subject matter
(4) 147 Rate the instructor's skill in encouraging students to think
(5) 24 Rate the instructor's availability during posted hours and appointments
(6) 25 Rate the instructor's willingness to help students make progress
(7) 88 Rate the instructor's enthusiasm about the subject matter
(8) 122 Rate the appropriateness of the instructor's encouragement of student participation and discussion
(9) 26 Rate the students' freedom to ask questions and express opinions
(10) 112 Rate the fairness of the overall grading system in the course
(11) 89 Rate the effectiveness of the instructor in demonstrating the significance of the subject matter
(12) 68 Rate the clarity of the syllabus in stating course objectives, course outline, and criteria for grades
(13) 31 Rate the instructor's skill in maintaining a positive atmosphere for learning
(14) 48 Rate the effectiveness of the instructor's responses to student questions
(15) 62 Rate the organization of course material

These questions are similar to those used in other surveys and are usually categorized into a few groups by content experts or statistical analyses. The questions fall into some standard categories, e.g., presentation skills and teacher-student interaction; and some more interesting ones, e.g., getting students to think, enhancing student motivation, and focusing on student improvement.

### **Use by Departments**

#### 1. As evaluation for promotion and tenure

The SRTE was developed for use in determinations about tenure and rank that require student input as evidence of teaching effectiveness. The scores for courses taught over a number of years are included as part of a dossier including peer review and student input on teaching effectiveness.

#### 2. As evaluation for annual performance review

The SRTE scores are included in the evaluation of teaching effectiveness in addition to instructional and advising activities, instructional improvements, teaching awards and other evidence. The review is used not only for personnel decisions such as salary increases, but as an opportunity for self-evaluation, reassessment and to recognize good work (HR40).

### **Use by Faculty**

#### 1. As evaluation.

The global ratings used by departments and promotion and tenure committees can also give instructors an overall picture of how they are doing. In addition, they can look at the ratings data over time to look for progress. Finally, they can use ratings on the B items to help them gauge their effectiveness in certain aspects of teaching, such as preparation for class.

#### 2. To drive improvement

Many departments give out comment sheets along with the SRTEs that ask for general comments such as “what was the best thing about this class?” These comments are the most useful for faculty interested in improvement because students often give concrete suggestions about what works well and what doesn’t.

### **SRTE Reports**

Below is a sample of the way that SRTE data are currently reported to faculty and departments.

## Sample SRTE Report

### STUDENT RATING OF TEACHING EFFECTIVENESS (SRTE)

Instructor: NAME, SOME INSTRUCTOR

Course & Section: PSU 010 001 Spec: Semester/Year: FA/03 Location: ZZ

Enrollment: 11 Number of SRTEs: 10 Percentage of Enrollment: 91% (Enrollment number is acquired before end of semester and may differ from final figure.)

#### University Items:

A 1 Number/percentage of sheets marked as elective: 0/ 0%

A 2 Number/percentage expecting grade of "A": 8/ 89%; "B": 1/ 11%; "C" or lower: 0/ 0%

Reported for each item are the number(n) and percentage(%) of respondents selecting each rating, and the total number(N) and mean for that item.

	1	2	3	4	5	6	7	N	MEAN
A 3 Rate the overall quality of this course	n: 0	0	0	2	0	3	5	10	6.10
	%: 0%	0%	0%	20%	0%	30%	50%		
A 4 Rate the overall quality of the instructor	n: 0	0	0	0	1	2	7	10	6.60
	%: 0%	0%	0%	0%	10%	20%	70%		

#### Department Items:

B 1 Rate the clarity of the instructor's demonstrations and explanations.	n: 0	0	0	0	0	3	7	10	6.70
	%: 0%	0%	0%	0%	0%	30%	70%		
B 2 Rate the instructor's availability for individual help and consultation.	n: 0	0	0	0	2	5	3	10	6.10
	%: 0%	0%	0%	0%	20%	50%	30%		
B 3 Rate the instructor's availability to assist students outside of class.	n: 0	1	0	2	0	3	4	10	5.60
	%: 0%	10%	0%	20%	0%	30%	40%		
B 4 Rate the instructor's interest in whether or not students understood course content.	n: 0	0	0	1	0	1	8	10	6.60
	%: 0%	0%	0%	10%	0%	10%	80%		
B 5 Rate the instructor's skill in using class time effectively.	n: 0	0	0	1	0	3	6	10	6.40
	%: 0%	0%	0%	10%	0%	30%	60%		
B 6 Rate the instructor's skill in converting experimental difficulties into positive learning experiences.	n: 0	0	0	0	1	2	7	10	6.60
	%: 0%	0%	0%	0%	10%	20%	70%		

SRTE

PAGE 2

Instructor: NAME, SOME INSTRUCTOR

Course & Section: PSU 010 001 Spec: Semester/Year: FA/03 Location: ZZ

	1	2	3	4	5	6	7	N	MEAN
B 7 Rate the effectiveness of the instructor's explanations of how student performance would be evaluated.	n: 0	0	0	0	1	3	6	10	6.50
	%: 0%	0%	0%	0%	10%	30%	60%		
B 8 Rate the instructor's skill in encouraging students to think.	n: 0	0	0	1	0	3	6	10	6.40
	%: 0%	0%	0%	10%	0%	30%	60%		

### III. Recommendations for interpreting and displaying SRTE data

1. Instructor ratings over five or more courses can be considered dependable measures.

Measures of reliability in terms of internal consistency are easily calculated for the forms in current use and range between .8 and .9 in a random sampling. Analysis of variance methods have found a measure of reliability to be around .7 when instructor ratings are obtained from 5 to 20 students per course and five courses taught by one instructor; and around .8 when ratings are obtained for ten courses (Deike, 1993). This means that the SRTE is a reliable instrument.

2. Mean ratings should be reported to one decimal point.

Reporting means to more than one-decimal point encourages reviewers to infer more precision than is due.

3. There is a need to develop sound methods to make intra-departmental comparisons.

Although an early trial period of reporting norms ended with negative results, about half of the department/division heads responding to a recent survey use norms or averages. The ratings are intended to be an objective measure with “4” defined as average. Perhaps because the observed mean is well above “4”, relative standards have been defined. These include college, school and department averages, “typical” and “model” scores.

For the ninety-percent of the departments using one form, it makes sense to define relative standards. For the purpose of recognizing teaching excellence, comparisons to a standard or benchmark can be made using critical values of binomial distribution.

4. Comparisons of course and instructor ratings between departments cannot be made between forms.

The flipside of having a flexible department-specific form for evaluating courses within a department is a loss of comparability between departments.

5. Use of graphical methods for display that will allow for easier and more appropriate interpretation of the data.

Below are sample methods for display of SRTE data to help faculty and departments interpret results more appropriately.

Sample Online Report



SRTE Instructor's report

This report contains confidential data and is not for public use

<b>Instructor:</b>	<b>Course &amp; Section:</b>	<b>Special:</b>	<b>Semester/Year:</b>	<b>Location:</b>
SOME INSTRUCTOR NAME	PSU 010 001	A	FA03	ZZ

<b>Enrollment</b>	<b>Number of Forms</b>	<b>Percentage of enrollment</b>	<b>NOTE:</b> Enrollment numbers are acquired before the end of the semester and may differ from final figure.
0019	16	84.2%	

A1	Number/Percentage of sheets marked as elective:	3	18%			
A2	Number/Percentage expecting grade of	A: 9	B: 7	C: 0	D: 0	E: 0

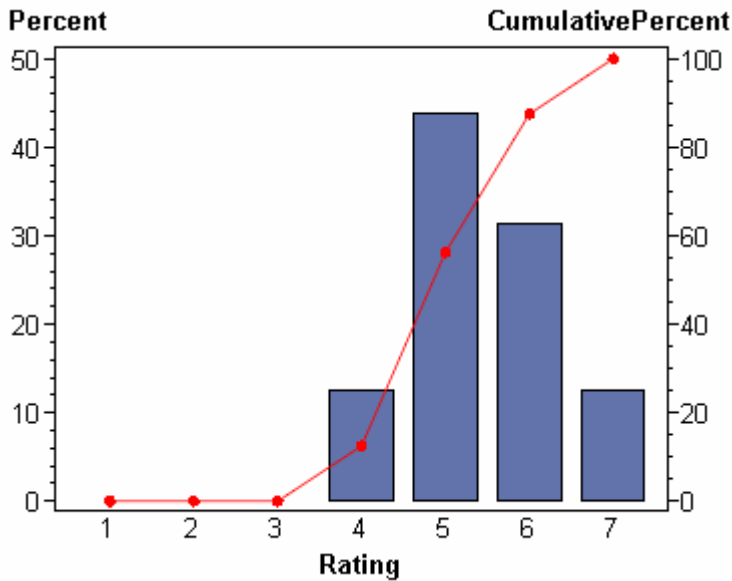
		Reported for each item are the number(n) and percentage(%) of respondents selecting each rating, and the total number(N) and mean for that item.								
		1	2	3	4	5	6	7	N	MEAN
A3	Rate the overall quality of this course.	0	0	0	1	5	8	2	16	5.68
		0%	0%	0%	6%	31%	50%	12%		
A4	Rate the overall quality of the instructor.	0	0	0	0	3	8	5	16	6.12
		0%	0%	0%	0%	19%	50%	31%		
B1	Rate the clarity of the instructor's presentations.	0	0	0	2	7	5	2	16	5.43
		0%	0%	0%	12%	44%	31%	12%		
B2	Rate the instructor's skill in organizing classroom activities.	0	0	0	2	3	8	3	16	5.75
		0%	0%	0%	12%	19%	50%	19%		

## Graphic Interpretations

A chart showing the reported percentage distribution of responses overlaid with a line representing the cumulative distribution gives a complete picture of responses. Examples from several courses follow.

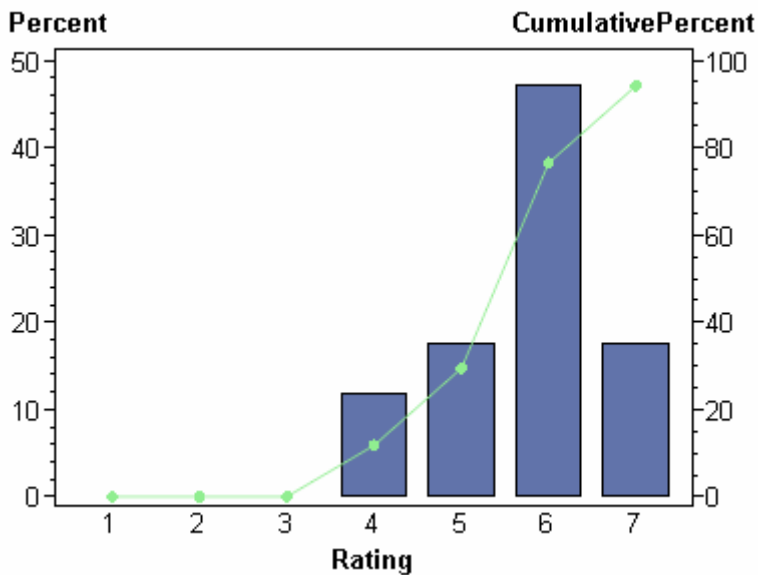
PSU 100 001

Rate the overall quality of the instructor.

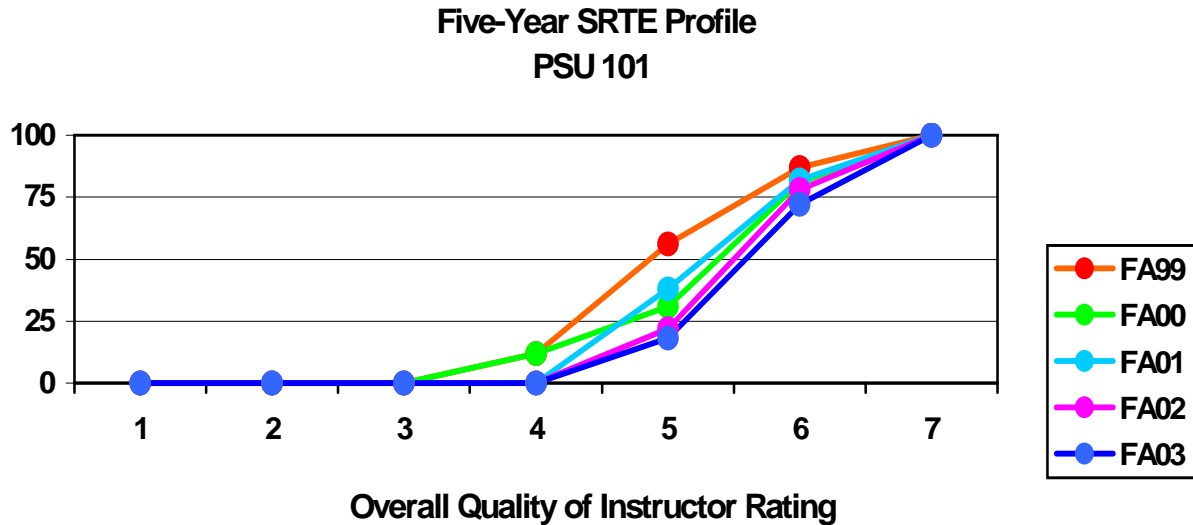


PSU 101 001

Rate the overall quality of the instructor.



The figure below depicts instructor ratings for several semesters. The closeness of the lines provides a measure of reliability. The figure also reveals improvement in ratings over time. A faculty member can easily maintain charts such as this one as part of a teaching portfolio.



6. Faculty could use other methods in addition to the SRTE to help them improve their teaching.

#### IV. Using midsemester feedback for course improvements

An excellent method to help instructors answer the question “How can I improve?” is the use of mid-semester feedback. There are a range of instruments available from SRTE-like forms to instructor-designed surveys.

(a) SEEQ

The Student evaluation of Educational Quality is a well-established, standardized feedback questionnaire written by Herbert W. Marsh (1976). At Penn State, this 40-item form is available as a machine-scannable form or as a web-based application. It has proven to be especially useful for instructors who teach large classes (for whom interpreting hand-scored surveys can be a daunting task) as well as for new teachers (who are often unclear about what questions to ask). While most instructors use it to collect mid-semester feedback, it can also be used at the end of the semester in addition to the SRTE.

(b) SALG

The Student Assessment of Learning Gains instrument is designed for instructors from all disciplines who wish to learn more about how students evaluate various course elements in terms of how much they have gained from them. Feedback from the instrument can guide instructors in modifying their courses to enhance student learning. It may be used at any

point during a course (for formative feedback) as well as at the end.

<http://www.wcer.wisc.edu/salgains/instructor/SALGains.asp>

(c) ANGEL survey tool

The Penn State course management system, ANGEL, includes a feature to create a survey. The tool is very easy to use, flexible and supports several question types, from Likert scales to open-ended questions. Because the structure of the feedback is not fixed, the instructor can construct it to gather information specific to the course.

(d) Classroom Assessment Techniques

At the other end of the spectrum of methods to assess student learning and/or student satisfaction with aspects of the course is to use various methods of classroom assessment throughout the course. Angelo and Cross (1993) have described 50 of these methods, each of which can be used for a specific assessment purpose.

## V. Recommendations for Improvement of the SRTE process

1. From the study of patterns of use, we may be able to identify a standard form for departments wanting to make comparisons.
2. Questions of validity are unanswered. Additional methods of assessment should be supported for some purposes. Ratings used in different ways may be used for self-study and to develop new teaching resources and programs.

The significant investment in time and effort used in the administration of the SRTE forms has contributed to increased pressure to use the SRTEs for purposes other than evaluating teaching effectiveness. The validity question (i.e., is the use of the course and instructor ratings by departments and colleges appropriate?) can be viewed in the context of current practices:

a. Promotion and tenure.

Does the measure of instructor quality and course quality measure teaching effectiveness when viewed in conjunction with other measures of teaching effectiveness or excellence as a collection of evidence collected over 5 or more years? The use of a variety of measures makes it possible to provide both a reliable and valid evaluation package.

b. Rehiring first-time instructors.

Since effective teaching is viewed as a learned-skill, a grace period in the evaluation of new faculty may be considered in conjunction with faculty development activities.

c. Improvement of instruction (individual self-study).

An instructor can chart the reported percentage distribution of responses over several semesters to show improvement. The development of resources (online, tools, diagnoses, and prescriptions) may prove useful.

d. Development of new teaching program and activities.

One study of teaching evaluations found that ratings of teaching effectiveness increased with experience, but only to a point. After about 13-15 years of teaching experience, the ratings decreased (Langbein, 1994). This finding points to the opportunity to improve teaching by increasing support of these faculty members, including recognizing their accomplishments.

e. Prescribing faculty development activities or resources.

Categorized responses to the categories of the “B” items developed by context experts could be used to prescribe faculty development activities or point to the need for teaching resources in, for example, presentation skills. Exploratory factor analyses of several SRTE forms indicate that this method may yield some instructor and course-related factors useful for this purpose.

3. Students may need to learn how to give careful evaluations of teaching.

Instruction in the art of giving careful evaluations early in a student’s Penn State experience may be worth the effort in terms of giving useful feedback to improve teaching. Ideas include developing a “lesson plan” or tutorial with the goal of training students to become careful evaluators. Faculty can include this training as part of a class or on the course ANGEL site. The lesson could become part of orientation for new students or a FYS. Minimally, a scripted statement should be made on the first day of class to students explaining their role in the learning community.

The Manual of the Irvine Division of the Academic Senate, University of California uses a scripted statement to convey the importance of ratings during the administration of the Standardized Teaching Evaluation Form to students (see [http://www.senate.uci.edu/senateweb/9\\_IrvineManual/3ASMAppendices/Appendix11App2.html](http://www.senate.uci.edu/senateweb/9_IrvineManual/3ASMAppendices/Appendix11App2.html)).

4. Consider the addition of some “A” section items to learn more about the context, “B” items that are course-related and “C” items to monitor University-wide initiatives.

Establishing validity of the SRTE is an ongoing process. The SRTE items could be enhanced to include more course-related items (at some point in time these were stripped out of the survey), student characteristics (student effort, concurrent coursework, etc.), and environmental characteristics (classroom facilities, etc.). For example, measures of student effort such as number of hours spent studying outside of class have been shown to be related to ratings of teaching effectiveness (Langbein, 1994). Another variable found to be associated with student ratings is whether it is a non-traditional course or not. Discovering the relationship between

ratings and other variables and controlling for them when making comparisons addresses some questions of bias.

5. Some further studies of the SRTE that may improve its psychometric properties and play a role in improving teaching and learning.

Preliminary analyses of a random sample of courses suggests that the “B” items are highly correlated so that the reliability measure of internal consistency is not lowered much by removing some “B” items. Further analyses may indicate recommendations for the numbers and types of “B” questions to use to measure teaching effectiveness, and provide diagnostic and prescriptive advice with implications for resource development..

6. A study of best practices may help reviewers define weighting schemes to calculate a fair total score and other methods that increase the fairness of using SRTE ratings for evaluating teaching effectiveness as part of a teaching portfolio.

A survey of methods used could yield best practices and new ideas to share as an online resource for defining an evaluation plan.

7. Explore and approve an alternative to the SRTE to measure non-traditional teaching and learning approaches.

An alternative survey that clearly recognizes faculty who are making the classroom enriching and rewarding will provide documentation for promotion and tenure and help show that Penn State endorses this effort. The guidelines for promotion and tenure could be uniformly changed from “SRTE” to “SRTE or other approved instrument”.

## **VI. Additional Resources**

### **A. University of Washington**

Guidelines for the use of student course ratings in personnel decisions.  
<http://www.washington.edu/oea/resources/recommendations.html>

The nine recommendations listed below provide guidelines for the use of student course ratings in personnel decisions.

1. *Student ratings must be used in concert with other data that relate to the quality of a faculty members teaching, rather than as a sole indicator of teaching quality.* Other sources such as peer reviews of classroom sessions, peer reviews of curricular materials, and faculty self-reflection should be assessed in addition to student evaluations to gain a true sense of the teaching skills and performance of a faculty member. Consideration of these other sources of evidence is especially important because student ratings alone do not provide sufficient evidence of the extent of student learning in a course.

2. *Evaluations from more than a single section should be used in making any decision about teaching quality.* Research has shown that ratings from at least five courses are necessary to assure adequate reliability. The validity of the ratings for measuring teaching quality is increased as a greater variety of course formats is represented in the data upon which decisions are based. Trends in ratings across years may also be important in assessing teaching.
3. *Only global ratings of teaching effectiveness, as represented by the first four items on Instructional Assessment forms, should be used in personnel decisions.* Other, more specific items should be used by the faculty member for review of specific skills and areas for improvement.
4. *Small differences in individual evaluations should not be used as a basis for differential decisions.* Because student ratings yield numerical averages, there is a temptation to overestimate the precision of the averages that are presented. Small differences in ratings may not be meaningful. It is better to deal with much broader classifications, such as Excellent vs. Very Good.
5. *Interpretations of student ratings averages should be guided by awareness that, in a university that uses teaching excellence as a hiring criterion, teaching is typically at a high level.* Also, students tend to rate faculty at or near the high end of the scale. Only about 5% of University of Washington faculty receive average ratings below the scale point labeled "Good." It is therefore not appropriate to use the median (or 50th percentile) as a presumed dividing line between strong and weak teachers. More appropriate would be to assume that the great majority of teachers are strong. It is also appropriate, when evaluating average ratings of individual instructors, to consider relevant comparisons (see Recommendation 6) and specific characteristics of courses taught (see Recommendation 7).
6. *Comparative data should be used but with caution.* Colleges and departments should compile information about evaluations for faculty in their courses over time for comparative purposes. However, for comparisons to be useful, the normative group should be based on more than a narrow population of instructors. Smaller departments may not want to rely on departmental norms but use norms calculated for a number of similar departments or for the school or college as a whole. At times, it may be better to compare ratings of similar courses across departments rather than ratings of dissimilar courses within departments.
7. *Course characteristics should be considered when interpreting results.* For example, large lecture courses typically receive lower ratings than smaller courses, new courses being taught for the first time receive lower ratings than well-established courses, introductory courses for non-majors receive lower ratings than higher division courses for majors, and courses in departments that have high student workloads typically receive lower ratings. Adjustments for course type should be made in order to have a fairer sense of the faculty member's teaching skills. One way to adjust for course types is by choosing

similar courses for normative comparisons.

8. *Faculty members should be given an opportunity to respond to evaluation results.* Faculty should have an opportunity to discuss the objectives of the course, how the teaching methods were used to meet that objective, and how circumstances in the course might have affected evaluations. Furthermore, other evaluation information gained from a given course (see Recommendation 1) can aid with the interpretation of ratings results.
9. *Administration of course ratings should be scheduled to maximize the number of respondents.* Generally, evaluations will have greater validity when higher proportions of the enrolled students complete evaluation forms. Ratings may not be an accurate reflection of the entire class when smaller proportions of students respond. This problem can be particularly acute in small classes. Generally, a minimum of a third of enrolled students must be included in the results to have any confidence in the results. As proportions decrease, particularly in small classes, there is greater opportunity for the rating of one or a few students to disproportionately affect the results.

## **B. National Center for Postsecondary Improvement**

Institutional Steps to Improve Teaching Through the Use of Student Evaluations  
[http://www.stanford.edu/group/ncpi/documents/pdfs/4-02\\_enddisconnect.pdf](http://www.stanford.edu/group/ncpi/documents/pdfs/4-02_enddisconnect.pdf), p.4

1. Situate the Evaluation System Firmly With the Academic Context
2. Strive for Quick Processing of Forms
3. Help Faculty to Interpret Their Evaluation Results
4. Create Opportunities for Peer Evaluation
5. Create a Grace Period in the Evaluation of New Faculty
6. Educate Students Regarding their Role in an Evaluation System
7. Stress the Importance of Midterm, as Well as End-of-Term Feedback
8. Create Opportunities for Reflection on One's Teaching Evaluations

## **C. Alternative Faculty Evaluation Forms**

1. Instructional Assessment System – University of Washington
  - Well-established system can be delivered with scan forms or online
  - Eleven machine-readable course evaluation forms (reflecting various course formats) and an open-ended *Comment Sheet*, with course-specific questions.
  - Forms for lecture classes, discussion classes, seminars, skills courses, distance learning classes, etc.
  - Course reports printed on a class-by-class basis.
  - *Comment Sheets* are forwarded to faculty
  - IAS provides annual ratings summaries for each department and college, and lists of faculty who participated in the ratings program.
  - Cost is \$.09 per form and \$.15 per form for processing, \$.67 to \$1.57 per report.
  - See sample in Appendix

2. Instructional Development and Effectiveness Assessment  
Kansas State University  
<http://www.idea.ksu.edu/StudentRatings/index.html>

The IDEA system focuses on student learning rather than the instructor's teaching techniques or characteristics. The IDEA Center tailors each report to fit the instructor's teaching objectives. Teaching effectiveness is determined by student progress on goals chosen by the instructor. Diagnostic assistance can be provided for those with disappointing results. Fairness is improved by taking into account the influence of factors outside the instructor's control. The following are features of the IDEA.

- Ratings can be done on paper or online
- Reports are reader-friendly
- Faculty tailor the report to meet instructor's objectives
- Includes a diagnostic component to help faculty make improvements
- Validity and Reliability for the items is continually assessed.
- Scores are adjusted for extraneous influences, such as student motivation, student work habits, class size, student effort and course difficulty.
- IDEA provides group summary reports that can help identify local norms and also be used for assessment and accreditation purposes.
- Cost is \$.13 per form and \$4.00 per class for processing.
- With over 60,000 classes processed annually, IDEA has a solid basis for providing benchmark data for individual instructors, departments, and institutions.

# Appendix A

## Instructional Assessment System



Fill in bubbles darkly and completely.  
Erase errors cleanly.



Instructor \_\_\_\_\_ Course \_\_\_\_\_ Section \_\_\_\_\_ Date \_\_\_\_\_

Completion of this questionnaire is voluntary. You are free to leave some or all questions unanswered.

	Excel- lent	Very Good	Good	Fair	Poor	Very Poor	
1. The course as a whole was:	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
2. The course content was:	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
3. The instructor's contribution to the course was:	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
4. The instructor's effectiveness in teaching the subject matter was:	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
5. Course organization was:	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
6. Clarity of instructor's voice was:	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
7. Explanations by instructor were:	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
8. Instructor's ability to present alternative explanations when needed was:	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
9. Instructor's use of examples and illustrations was:	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
10. Quality of questions or problems raised by instructor was:	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
11. Student confidence in instructor's knowledge was:	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
12. Instructor's enthusiasm was:	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
13. Encouragement given students to express themselves was:	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
14. Answers to student questions were:	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
15. Availability of extra help when needed was:	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
16. Use of class time was:	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
17. Instructor's interest in whether students learned was:	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
18. Amount you learned in the course was:	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
19. Relevance and usefulness of course content were:	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
20. Evaluative and grading techniques (tests, papers, projects, etc.) were:	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
21. Reasonableness of assigned work was:	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
22. Clarity of student responsibilities and requirements was:	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
<b>Relative to other college courses you have taken:</b>							
23. Do you expect your grade in this course to be:	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
24. The intellectual challenge presented was:	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
25. The amount of effort you put into this course was:	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
26. The amount of effort to succeed in this course was:	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
27. Your involvement in this course (doing assignments, attending classes, etc.) was:	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
28. On average, how many hours per week have you spent on this course, including attending classes, doing readings, reviewing notes, writing papers and any other course related work?	<input type="radio"/> Under 2 <input type="radio"/> 2 - 3 <input type="radio"/> 4 - 5	<input type="radio"/> 6 - 7 <input type="radio"/> 8 - 9 <input type="radio"/> 10 - 11	<input type="radio"/> 12 - 13 <input type="radio"/> 14 - 15 <input type="radio"/> 16 - 17	<input type="radio"/> 18 - 19 <input type="radio"/> 20 - 21 <input type="radio"/> 22 or more			
29. From the total average hours above, how many do you consider were valuable in advancing your education?	<input type="radio"/> Under 2 <input type="radio"/> 2 - 3 <input type="radio"/> 4 - 5	<input type="radio"/> 6 - 7 <input type="radio"/> 8 - 9 <input type="radio"/> 10 - 11	<input type="radio"/> 12 - 13 <input type="radio"/> 14 - 15 <input type="radio"/> 16 - 17	<input type="radio"/> 18 - 19 <input type="radio"/> 20 - 21 <input type="radio"/> 22 or more			
30. What grade do you expect in this course?	<input type="radio"/> A (3.9-4.0) <input type="radio"/> A- (3.5-3.8) <input type="radio"/> B+ (3.2-3.4)	<input type="radio"/> B (2.9-3.1) <input type="radio"/> B- (2.5-2.8) <input type="radio"/> C+ (2.2-2.4)	<input type="radio"/> C (1.9-2.1) <input type="radio"/> C- (1.5-1.8) <input type="radio"/> D+ (1.2-1.4)	<input type="radio"/> D (0.9-1.1) <input type="radio"/> D- (0.7-0.8) <input type="radio"/> E (0.0)	<input type="radio"/> Pass <input type="radio"/> Credit <input type="radio"/> No Credit		
31. In regard to your academic program, is this course best described as:	<input type="radio"/> In your major? <input type="radio"/> In your minor?	<input type="radio"/> A distribution requirement? <input type="radio"/> A program requirement?	<input type="radio"/> An elective? <input type="radio"/> Other?				

Mark Reflex® by NCS MMB9769-2 654321 ED06 Printed in U.S.A.

©1995, University of Washington - Office of Educational Assessment

## References

Administrative Guidelines for HR-23: Promotion and Tenure Procedures and Regulations. <http://guru.psu.edu/policies/OHR/hr23.html>

Angelo, T. A., and Cross, P.K. (1993). Classroom assessment techniques: a handbook for college teachers. San Francisco: Jossey-Bass Publishers.

Cornell University Teaching Evaluation Handbook. (2004).  
<http://www.clt.cornell.edu/resources/teh/intro.html>

Deike, R. C. The dependability of the student rating of teaching effectiveness. Unpublished paper. August 17, 1993.

Dooris, M. J. (1997). An analysis of the Penn State student rating of teaching effectiveness. [http://www.psu.edu/president/pia/planning\\_research/reports/srte/1997/](http://www.psu.edu/president/pia/planning_research/reports/srte/1997/)

Marincovich, M. Ending the disconnect between the student evaluation of teaching and the improvement of teaching: A faculty developer's plea. National Center for Postsecondary Improvement. [http://www.stanford.edu/group/ncpi/documents/pdfs/4-02\\_enddisconnect.pdf](http://www.stanford.edu/group/ncpi/documents/pdfs/4-02_enddisconnect.pdf)

Langbein, L. I. (1994). "The Validity of Student Evaluations of Teaching", PS: Political Science and Politics, 27, 545-553.

Office of the Vice Provost for Academic Affairs. Guidelines for Use and Administration of the Student Rating of Teaching Effectiveness (SRTE) Forms:  
[http://www.psu.edu/dept/vprov/pdfs/srte\\_guidelines.pdf](http://www.psu.edu/dept/vprov/pdfs/srte_guidelines.pdf)

Marsh, H. and Roche, L. (1993). The use of students' evaluations and individually structured intervention to enhance university teaching effectiveness. American Educational Research Journal, 30, 217-251.

Policy HR40 Evaluation of Faculty Performance. 27January2004.  
<http://guru.psu.edu/policies/OHR/HR40.html>

Schreyer Institute for Teaching Excellence. Guide to Using and Analyzing the SRTE.  
[http://www.schreyerinstitute.psu.edu/pdf/Guide\\_to\\_Using\\_and\\_Analyzing\\_the\\_SRTE.pdf](http://www.schreyerinstitute.psu.edu/pdf/Guide_to_Using_and_Analyzing_the_SRTE.pdf)

Statement of Practices for the Evaluation of Teaching Effectiveness for Promotion and Tenure. [http://www.psu.edu/dept/vprov/pdfs/srte\\_statement.pdf](http://www.psu.edu/dept/vprov/pdfs/srte_statement.pdf)