

**Preface**

In the following article, Diane Enerson provides a thorough overview of using Penn State's Student Rating of Teaching Effectiveness (SRTE). In particular, she makes a case for the importance of choosing questions carefully so that they provide the most useful information possible.

Instructional technology has changed a great deal in the time since Enerson's essay was written (c. 1997), so some departments today might opt to use the more generic question #75—"Rate the effectiveness of the integration of instructional materials (textbooks, handouts, overheads, films, etc.)"—rather than or in addition to question #137—"Rate the instructor's skill in using the blackboard/overhead projector" (mentioned below). Also since the time the article was written, the Schreyer Institute for Teaching Excellence has assumed responsibility for processing SRTE results. Detailed information about SRTE policy can be found at <http://srte.psu.edu/>.

September, 2009

Guide to Using and Analyzing the SRTE

by Diane M. Enerson

The Student Rating of Teaching Effectiveness (SRTE) has become an established part of Penn State's culture and its end-of-the-semester rituals. SRTEs also play a role in virtually every promotion and tenure decision and teaching dossier of Penn State faculty. Despite their omnipresence, however, they are not always well understood. Although some faculty and departments find them useful and feel they can derive meaningful information from them, many do not. Departments and faculty committees often raise questions about them and whether they can be used effectively. The SRTE is obviously only one source of information about teaching effectiveness. Multiple sources of information are generally more desirable.

Nevertheless, student input is clearly crucial to excellence in teaching, and the SRTE is currently the official source of student data about teaching effectiveness.

Accordingly, this guide has been written to address some of the most frequent questions that seem to arise in connection with using the SRTE: how to select items from the B-items list and how to interpret responses. The guidelines that follow are meant to complement, not replace, earlier guidelines published by University Testing Services in consultation with the Faculty Affairs Committee of the Senate that make recommendations about who should be responsible and when the SRTE should be used. The purpose of the present document is to augment those earlier [guidelines](#) by providing a series of questions and context for discussion that departments and faculty committees may find useful when deliberating on existing procedures or establishing new ones. We welcome your [feedback and comments](#).

Background

One question that is often raised about the SRTE is probably also the most basic: Do these data have any merit? Underlying this question is the even more fundamental question of whether student ratings of teaching in the most general sense are useful. Seven decades of research on this question point unequivocally to the answer.



On the whole student ratings can be a useful and reliable source of information about teaching effectiveness and are not easily swayed by trivial or uncontrollable variables.

Student ratings are not simply popularity contests but a fairly stable source of input about course effectiveness. Students, it seems, are quite capable of making reasonable and reliable judgments about the teaching they receive. A recent study at Penn State by Fern K. Willits (1997) found that students agree more than disagree with faculty about characteristics of effective teaching. The results of this study also suggest that students' perceptions of how much they had learned was the single most powerful predictor of their overall level of satisfaction with the course.

Furthermore, Michael Dooris (1997) concludes:

The SRTE appears to be a dependable instrument that probably produces consistent student ratings of individual instructor quality, when enough observations are taken. Expected grades statistically explain only about five percent of the variation in student ratings of instructor quality. The effects of other factors (such as class size and course level) on student ratings of instructor quality are even weaker. The findings on the Penn State SRTE are mostly similar to the results of other research on student ratings.

[Dooris's findings](#) concur with our own experiences and observations. The results of the SRTE are a fairly reliable (i.e., the same group of students would report roughly the same ratings about that teacher at a later time) if somewhat crude index of how a course has gone. In fact, faculty often note with some surprise that their SRTE scores for a particular course are remarkably consistent across different sections of the same course—within a given semester and over time—and are consistent with their own perceptions of how successful that course has been.

The SRTE is not, however, an especially sensitive or powerful instrument. A fairly sizable number of observations are needed for stability. Numerical differences alone cannot—and should not—always be interpreted as meaningful instructional difference. Sudden drops in ratings, for example, are fairly common when teaching a course for the first time or when making innovations in a course. Obviously, we would not want to interpret such declines as evidence of poor teaching. Similarly, the SRTE scores an individual initially receives when teaching an experimental course do not always reflect the positive changes that have been made, even though other sources of student feedback may reflect positively on those changes. And in some instructional situations there may be a powerful course effect that is beyond the immediate control of the individual teaching that course, in which case the SRTE scores for that course may never rise above a certain level irrespective of who teaches it.

Ultimately, then, reflective and thoughtful use of the SRTE requires discussion that goes beyond simply questioning whether students should be asked to rate the teaching they receive and instead focuses on determining which questions students might reasonably be asked and what interpretation will be imposed on the responses they give. Accordingly, the following sections have been prepared as an aid to those faculty, faculty committees, and departments who may have concerns about making the best use of SRTEs.



Reflecting on Which Questions Might Be Asked

The secret to getting the best data from the SRTE is to ask the right questions—questions you want to know the answers to and questions that are reasonable to ask of students. One of the most frequent errors of using and analyzing student data is to assume that students are in a position to answer every conceivable question about an individual's teaching effectiveness. Therefore, the process of reflecting systematically and critically on the SRTE begins with selecting appropriate, relevant, and reasonable questions from the available B items.

The following questions are offered to help structure and give direction to those deliberations.

1. Which dimensions of teaching effectiveness do you wish to measure? All aspects or dimensions of teaching effectiveness represented among the SRTE B-items are not equally applicable in every classroom situation. What characterizes effective teaching in the course (or courses) in question? What are the expected student outcomes? You may find it helpful to begin by reflecting on what you need to know and which questions you will need to have answered in order to make responsible judgments about a course or an individual. For example, if a course is taught predominantly by discussion or small group work, it may not be especially informative to ask students to rate the clarity of the presentations or demonstrations. But it would be useful to have them rate their degree of comfort in asking questions. Conversely, in courses that are taught predominantly as a lecture, asking students to report on their levels of participation may not be as critical as would be the issue of clarity in the presentations and feedback that they received.

2. On which of these dimensions would student input be most appropriate? Although it can legitimately be argued that students are *naive* observers of teaching, they are nonetheless in a unique position to observe the day-to-day impact of teaching on their learning. Who better to report on whether the examples that were used helped to clarify difficult concepts in the course material, whether the feedback they received on exams and assignments was helpful, or whether their interest in the subject matter increased as a result of taking the course? However, they are not able to report on all aspects of a course that an individual (or a department) might like to have answered about the effectiveness of that course or that instructor. And in some cases, better sources of data may simply exist elsewhere.

It is also generally useful to ask whether students would reasonably have access to the judgments you are asking them to make. For instance, students are typically not the appropriate source of information about whether an instructor is knowledgeable about the subject matter, or whether the course adequately represents the full depth of scholarship in a field. These are reasonable questions, but not ones that we should expect students to answer for us. Rather, these are the very questions that ideally would be at the core of substantive and constructive peer review. Students, in fact, will often resent being asked such questions since most will assume that decisions about an individual instructor's subject matter competency are not their responsibility and should have been made long before a teacher enters the classroom.

3. Which B-items could provide you with the data you need? When making a list of relevant items, you may find many more than you expected. Currently, the SRTE B-items exist as an unsorted pool of 177 questions,

many of which are simply linguistic variations of one another. (Note: These variations are not necessarily adjacent to each other.)

Interestingly, but not surprisingly, an analysis of which questions have been used most frequently reveals that some of the questions in this rather large random assortment have been used considerably more frequently than others, and that a few have actually never been used at all.

Historically, departments and faculty committees have actually done quite a good job of selecting the most clearly worded version of those questions. What has been somewhat less successful has been the practice in some units to ask questions that require students to make the same fundamental judgment more than once. Asking students to rate several aspects of "clarity," for example, can cause a certain level of student irritation if they perceive it as redundant because they can't distinguish between the different items. Generally speaking, it is best to avoid such redundancy by selecting the best version of each judgment you want students to make.

5. Which questions, if any, have policy implications you cannot address? Students often express great frustration with and skepticism about the SRTE. Their greatest concern is whether the data they provide via the SRTE will be valued and used. Students want first to be asked reasonable questions and second to be listened to when they provide the answers to those questions. To ask students to respond to a question that you are not prepared to take action on if their response is less than favorable will be generally perceived as insincere, thereby making it harder to get good student cooperation and feedback in the future. In short, it is better not to ask a question than to ask it and then ignore the answer(s) you receive.

Questions That Others Have Found Especially Useful

Below we have listed 36 questions from the SRTE B-item list. These questions are all ones with a good track record for providing information about teaching effectiveness. In assembling this list we have tried to select the least ambiguous versions of questions that are repeated in other places on the SRTE.

Obviously these are not the only good questions that could prove useful. The following list, however, should provide a reasonable place to start. For clarity and ease of cross referencing, the numbering of the items has been retained from the list of SRTE B-items.

1. Rate the clarity of the instructor's presentations.
2. Rate the effectiveness of the examples used to clarify difficult concepts.
3. Rate the clarity of the instructor's explanations.
5. Rate the instructor's skill in relating course material to real life situations.
10. Rate the instructor's skill in handling students' questions and comments.
11. Rate the instructor's pacing of lectures and presentations to allow for note taking.
18. Rate the instructor's skill in using examples and illustrations.



- 24. Rate the instructor's availability during posted hours and appointments.
- 31. Rate the instructor's skill in maintaining a positive atmosphere for learning.
- 37. Rate the effectiveness of the instructor's advice on how to study for the course.
- 45. Rate the instructor's respect for students as individuals.
- 51. Rate the instructor's skill in relating course material to other disciplines.
- 62. Rate the organization of course material.
- 64. Rate the effectiveness of the instructor's explanations of what students were expected to learn.
- 65. Rate the correspondence between announced objectives and what was taught.
- 68. Rate the clarity of the syllabus in stating course objectives, course outline, and criteria for grades.
- 74. Rate the instructor's pacing of the course material.
- 86. Rate the usefulness of papers and term projects in adding value to the course.
- 87. Rate the extent to which interest in the subject matter was generated by this course.
- 89. Rate the effectiveness of the instructor in demonstrating the significance of the subject matter.
- 95. Rate the usefulness of the instructor's comments on papers and exams in indicating errors.
- 98. Rate the clarity of the exam questions.
- 99. Rate the effectiveness of the instructor's suggestions of ways in which students could improve.
- 100. Rate the effectiveness of the instructor's suggestions to students about their progress.
- 101. Rate the promptness with which graded exams, reports, and other materials were returned.
- 105. Rate the appropriateness of the difficulty of exams.
- 114. Rate the instructor's openness to discussion of other viewpoints.
- 137. Rate the instructor's skill in using the blackboard and/or overhead projector.



- 139. Rate the appropriateness of the instructor's demonstrations.
- 141. Rate the instructor's skill in encouraging class participation and discussion.
- 147. Rate the instructor's skill in encouraging students to think.
- 157. Rate the effectiveness of the course in improving problem-solving skills.
- 158. Rate the effectiveness of the course in stimulating further explorations of the subject matter outside of class.
- 164. Rate the effectiveness of the course in stimulating you to work harder.
- 166. Rate the instructor's skill in presenting viewpoints other than his/her own.
- 167. Rate the effectiveness of the instructor in stimulating your thinking.

Reflecting on the Data You Receive

Useful analyses of the data often begin by reflecting on what you might expect and the other kinds of data that you have to give weight to those expectations. As with most other analytic tasks, searching for patterns in the data is an obvious first step. Because the SRTE can be used to make comparisons between individuals, as well as to analyze an individual's performance over time, it is also critical that careful thought be given to which comparisons will be made. For example, although it may be reasonable to make comparison across individuals who are all teaching the same course, what comparisons are reasonable to make between a brand new teacher of that course and one who has been teaching it for 20years?

If you assume that the SRTE assesses teaching effectiveness, and that teaching improves over time, then at what level would you expect those changes to be reflected in the SRTEs? What interpretation might you give and other data might you need if there is a difference? When is it reasonable to make comparisons between instructors? When is it not?

Finally, when reflecting on the data you collect, it is natural to want to bring some meaning to the data. What meaning you ultimately ascribe to these data may hinge on the degree to which they concur with other data you may have about an individual's teaching accomplishment. That is, because no single source of data can be held solely accountable for assessing and improving teaching, it may be useful to ask the following questions. What other data do you have? What other data can you easily get? And what other data do you need? These and other questions related to them, lead naturally to discussions of the [teaching portfolio](#).

**References**

Dooris, M. J. (1997). An analysis of the Penn State Student Rating of Teaching Effectiveness. Report presented to the Penn State University Faculty Senate. Source URL:
http://www.psu.edu/president/pia/planning_research/reports/srte/1997/

Willits, F. K., Moore, B. L., and Enerson, D. M. (1997). Penn State quality of instruction: Surveys of students and teachers at University Park. University Park, PA: Center for Excellence in Teaching and Learning.