



Guidelines for the Use of Results of the Student Instructional Report II (SIR II)

The Student Instructional Report II (SIR II) typically and appropriately is used for instructional improvement; for tenure, promotion, or salary decisions; and by students for course selection. These guidelines provide information to teachers, administrators, and students who use SIR II in any of these ways. Although there may be other uses of SIR II results, these guidelines address the three most frequent ones. Each guideline, unless otherwise indicated, is appropriate for all three uses.

It is important that faculty members and administrators understand clearly how the results of student evaluations will be used, who will have access to any results, and how their use relates to local contractual arrangements or institutional policies.

These guideline recommendations were based on a series of studies with the SIR II and other research with similar instruments. A committee of SIR II users, ETS staff, and researchers met to review and discuss the guidelines. The final list represents the experience and knowledge of this group.

1. Use multiple sources of information.

For whatever purposes the results may be used, it is critical to keep in mind that student instructional ratings represent only one source of information about teaching performance. Other information about teaching, in addition to student opinion, also should be included. In particular, SIR II should not be used as the sole basis for evaluating teaching effectiveness.

2. Use multiple sets of ratings.

A pattern of ratings over time is the best estimate of instructor effectiveness as seen by students. Ratings from only one course or from one term may not fairly represent a teacher's performance (although, for course improvement, ratings from a single course can be useful). For personnel decisions, it is essential to examine rating trends or patterns over time (see additional comments in number 4 regarding possible course bias).

3. Obtain a sufficient number of student raters.

The reliability of the SIR II items depends on having a sufficient number of students responding in order to reduce the effects of a few divergent raters. Currently, reports are not printed for a class with fewer than five students. Reports based on responses from fewer than 10 students are flagged with an asterisk and users are advised to interpret them with caution. When fewer than 10 students respond to any individual item, the same caution applies.

The proportion of a class that rates an instructor also is important. If over one third are absent or choose not to respond, the results may not be representative of a class. On the report itself, item means are not computed if 50 percent or more of the students either omit an item or mark it not applicable.

4. Take into account course characteristics.

A few course characteristics appear to affect ratings and should be taken into account by reference to appropriate comparative data or in other ways. Small classes (that is, under 15) often receive more favorable ratings than larger classes, perhaps deservedly, since they often provide a better learning environment. Courses required by the college that are not part of a student's major or minor field tend to receive somewhat lower ratings than other courses. Ratings also may differ because of the subject field of the course. For each of these characteristics, the differences may not be large, but together they can be significant.

5. Rely more on global ratings than other items for personnel decisions.

Overall ratings of the teacher or the course tend to correlate higher with student learning scores in a course than do other items. Decision makers, therefore, should focus initially on the overall evaluation items. Items that are useful for diagnosing teacher or course strengths and weaknesses are important for improvement purposes and for interpreting the overall ratings in personnel decisions. These items tend to reflect different teaching styles and therefore should not be summed or averaged to provide a total score.

6. Supplement diagnostic information for teaching improvement.

SIR II results help to diagnose teachers' strengths and weaknesses. Although studies have shown that some teachers can improve after receiving results, others may not know how to change. Instructional development services and resources can help teachers who want to do something about these weaknesses. It is appropriate to use SIR II results in instructional counseling and to direct teachers to resources for instructional improvement.

7. Use comparative data.

Since student ratings typically tend to be favorable, comparative data (both national and appropriate local data) provide a context within which teachers and others can interpret individual reports. In making comparisons, it is important to look at the distribution of students' responses in each class as well as at means, and not over interpret small differences. Differences of less than 10 percentile points on any item generally are not critical. Data are presented only at 10 percentile intervals. In most cases differences of at least 20 percentile points are needed to be significant relative to the national comparative data.

Users of SIR II are reminded that the national data are user data and therefore are comparative rather than normative, and the tendency toward high ratings may work to the disadvantage of some instructors. Institutions may wish to supplement the national data with local normative data that are developed over time.

8. Employ standardized procedures for administering the forms in each class.

When the results will be used in personnel decisions, it is critical to employ standardized administrative procedures. Each institution will want to develop its own method. One possibility is to have a student, another faculty member, or

someone other than the teacher involved distribute, collect and place the questionnaires in a sealed envelope. (Mailing the forms to students usually results in a poor response rate.) The teacher should not be present during the process, which probably will take less than 15 minutes of class time. The timing, preferably during the last week or two of class, also should be standard; it probably is best to give results to instructors after grades for the course have been reported.

Additional Suggestions

1. For additional diagnostic information, use the optional items and written comments.

Use of optional items can make the SIR II adaptable to a wider range of courses. Up to 10 additional and locally written items can be added. These might be course specific, provided by the individual teacher or department, or they could be a combination of institutional and course specific items.

2. Teachers should be encouraged to supplement their instructional ratings.

This is especially important in personnel decisions or in student use of results for course selection. Teachers should be encouraged and given the opportunity to describe what they were trying to accomplish in the course and how their methods fit those objectives, or to discuss circumstances they feel may have affected the evaluations.

3. Do not overuse the forms.

If ratings are used in every course every term, students can become bored and respond haphazardly. Faculty members may also resent the lost class time. Strike a balance between the need for external evaluation and the need to experiment freely in instruction.



EDUCATIONAL TESTING SERVICE, ETS, and the modernized ETS logo are registered trademarks of Educational Testing Service. Copyright © 2013 by Educational Testing Service. All rights reserved.

Permission is hereby granted to colleges and universities administering the SIR II to copy and distribute this document for internal use only. No commercial or further distribution is permitted.

283826

