INTERPRETATION GUIDE FOR STUDENT OPINION OF TEACHING EFFECTIVENESS (SOTE) RESULTS

Prepared by

STUDENT EVALUATION REVIEW BOARD

October 2004

NOTE: This Guide applies to the newly revised SOTE rating form adopted in Fall 2003 (Appendix A). The SOLATE rating form (Appendix B) is currently undergoing modifications. Continue to use the previous edition of the Interpretation Guide (attached) when reviewing SOLATE reports.

Following several years of development by the Student Evaluation Review Board (SERB), a new SOTE rating form was adopted for implementation beginning in the Fall 2003 semester (see F02-2). In addition to developing tools for rating teaching effectiveness, SERB is also charged with "developing and making available to the University community, information and guidelines for the effective interpretation of the rating instruments." As per this charge, the interpretation guide presented here provides a description of the new form, explanations for the statistics included in the SOTE report, and factors that influence SOTE ratings. The interpretation guide refers to and explains analyses of SOTE results generated during the Fall 2003 semester when the new form was administered. Based on administration to 2827 classes and returns of 66443 SOTE forms during the Fall 2003 semester, a new set of norms were generated for use in evaluating teaching effectiveness. Additionally, data on a variety of demographic and other variables were collected for use in helping to identify meaningful patterns in SOTE responses. Drawing upon previous guidelines for the interpretation of SOTES, and incorporating the changes dictated by the current SOTE, this interpretation guide should be used to evaluate the both the "statistical" opinion of teaching effectiveness data provided by students, and the written subjective comments of students in order to reach a qualitative judgment about effectiveness in teaching assignment.

Differences between the current SOTE and the previous SOTE

The following are the most important differences between the previous version of the SOTE and the current version. Each of these has implications for interpreting the SOTE, and these implications are noted.

Format

Unlike the previous version of the SOTE, the current version presents each item in a separate box of its own. The form was designed in this way to maximize the likelihood that each item would be read and considered on its own, and to reduce the likelihood that students would simply endorse the same rating for each item by marking the same number in a straight line.

Scale

The rating scale for the current SOTE consists of five points in a Likert type scale with ratings of (5) Very strongly agree; (4) Strongly Agree; (3) Agree; (2) Disagree; and (1) Strongly Disagree. There is also a sixth option, (NA) Not Applicable/No Opportunity to Observe. Note that in the previous version of the SOTE, the ratings ranged from (5) Excellent to (1) Far Below Average, with (3) rated as Average. In interpreting the previous version of the SOTE, there were, in essence only two points (ratings of 4 and 5) that signified teaching excellence. In the current version there are three points (agree, strongly agree, very strongly agree) that signify a positive evaluation of teaching effectiveness. Students now have the option of choosing among a greater range of "good" evaluations. When looking at dossiers that contain both the previous and the current version of the SOTE, RTP committee members should consider that **the two sets of ratings are not directly comparable**. **In interpreting SOTES collected using both the old and the current SOTE, instructors' scores should be evaluated in comparison to the corresponding Department, College, and University norms for each item (see below for an explanation of new norms).**

Number of items

Unlike the previous version, the current SOTE contains 13 items, the last of which is comparable to the "old" item 14, and refers to overall teaching effectiveness. This item shows a very high correlation with most of the other items and therefore is a good index of overall effectiveness. Nonetheless, **RTP committees should carefully examine ratings for all 13 items and not solely rely on ratings for item 13**.

Subjective Evaluations

The new rating form is formatted as a 2-page booklet. The first page contains standardized rating items, the second page of the new form contains questions in which students are asked to provide subjective written comments regarding the teaching effectiveness of their instructor. Subjective ratings of "officially" rated classes must be included in the dossier. **In interpreting these responses, members of RTP committees should take into account the majority of comments, rather than focusing on atypical responses. However, if comments are repeatedly observed for the same instructor across sections and time, then the RTP committees should consider further evaluations for that instructor.**

SOTE Report Display

The SOTE Report remains basically unchanged from the previous version. However, item **medians** now appear alongside the item means and standard deviations. Each report is also identified as "OFFICIAL" or "PERSONAL" along the top and bottom margins. Also appearing on the bottom of Page 2 of "OFFICIAL" reports is the number of written comment pages that are on file in the PAF for the class.

<div align="center">

The Statistics

</div>

Explanations for the various statistics used on the SOTE report, how to interpret them, and potential caveats are described herein.

The **mean** is the arithmetic average of student responses. Means are reported to the first decimal place. As noted below, caution should be used in interpreting means based on fewer than 10 students' responses.

The **standard deviation** is a measure of agreement among respondents. It indicates the variability among the responses. That is, how much, on the average, student responses vary from the mean. Standard deviations for most items are very close to 1.0. A large standard deviation (greater than approximately 1.3) indicates that students frequently do not agree about what rating should be assigned (i.e. students use three or more descriptors for a single item). A small standard deviation (less than approximately .7) indicates that students generally agree about what rating should be assigned (i.e. students usually use only two adjacent descriptors for a given item). We do not expect to often see 100% agreement among students – an excellent teacher for one student may be only average for another student given differential preparation or experiences of the two students.

A caveat in interpreting means and standard deviations is that both statistics are highly influenced by even one or two aberrant scores if the number of ratings is fewer than about 10. Thus classes and/or items where fewer than 10 students have responded have been flagged with an asterisk and the following sentences will be printed directly below the rating items - *ITEM STATISTICS ARE BASED ON 10 OR FEWER STUDENTS. RESULTS SHOULD BE INTERPRETED WITH CAUTION*. Great caution should be used when interpreting means and standard deviations of such classes and/or items because the statistics may be unstable – check for consistency across classes and across rating occasions. In addition, when more than 30% of the students in a class leave an item blank or mark it "not applicable," that rating probably should not be interpreted.

The **median** is the middle ranking. A median of 3.5 indicates that half the students gave ratings higher and half lower than 3.5. The median is helpful in cases where outliers might influence the mean and standard deviation; e.g. cases in which a few extremely high or extremely low ratings might push the mean score in a direction that is not representative of the class as a whole. This is particularly likely in smaller classes or classes with large numbers of blanks or "not applicable" ratings.

The Norm Data

Following the introduction of the current SOTE form, new norms were computed based on the administration of the SOTE to all classes during the Fall 2003 semester (SOTE forms were returned for 93% of classes that were subject to evaluation). Norms for each item are provided at the Department/School, College, and University level (except in cases where there were there were 12 or fewer classes evaluated in the Department/School). At each level, responses were aggregated to compute the means, medians, and standard deviations that serve as norms or a referent point for making comparisons. Comparisons between the class data and norm data are best made using the graphic display on the second page of the report.

Norm data for the College and University levels only are graphically displayed on page 2 of the printout. For each item the middle 60% of ratings received by instructors was determined. This range is displayed as a line of dashes. This line represents the usual range of ratings received by instructors for that item. The class mean is printed as an asterisk on the same line. Only if the class mean falls below the norm (represented by an asterisk to the left of the dashes) or above the norm (represented by an asterisk to the right of the dashes), can SOTE data can be used to identify exceptional teachers (those with rating means outside the norm average.) The usefulness and validity of the ratings will be degraded if ratings within the norm area are interpreted as anything other than typical. It

should be noted that students tend to "agree" with the statements on the SOTE, indicating a highly favorable evaluation of the typical SJSU instructor.

SOTE interpretation should be done using trends across classes and semesters. If one item mean is <u>consistently</u> below (or above) the norm then the item should be noted as important. If an item mean is <u>inconsistently</u> above or below the norm, RTP committee members should request further information from the faculty member about the classes. It is especially important to note consistencies or inconsistencies in the same course preparation on different occasions. Thus it is possible to note steady improvement or decline.

Factors Affecting the Ratings

Several factors were found to systematically influence SOTE ratings in the Fall 2003 pilot. Each is described below and references to similar findings from research on faculty evaluation conducted elsewhere are provided. These factors should be considered in any RTP evaluation of SOTE data. It is the responsibility of the faculty member to assure that information about any of these factors is included in the dossiers along with the ratings.

Expected Grades

Ratings are slightly but positively related to both expected and received grades (Theall, 2002). Students are asked to report their expected grade at the time of the SOTE administration. Frequencies for each possible grade are noted on the SOTE report, as is the actual average final GPA grade for the class. In general, one would expect to see expected grades distributed across the range of possible grades. Data from the Fall 2003 norming sample indicate that students expecting higher grades tend to rate their instructors more highly than students expecting lower grades. When interpreting SOTE ratings RTP committees should note the distribution of expected grades. Classes in which the majority of students expect either low or high grades should be fairly rare (exceptions to this would be graduate and credential classes in which a grade lower than a "B" is

often considered equivalent to a failing grade, and some classes in the Colleges of Science and Engineering in which grades are often lower than in other subjects).

In addition to reporting students' expected grades, the average grade for each class is also reported. In general, ratings tend to be slightly but positively related to grades (In the Fall, 2003 sample the correlation between expected grade and the SOTE score given by the student is approximately .24). In general, it should be expected that average grades for a class show some relationship to expected grades. In cases where there is a wide discrepancy (e.g. 80% of the class expects a grade of "A" while the actual average grade for the class is a 2.3) RTP committees should request further information from the instructor.

Class Size

Ratings in small or moderate sized classes (<20) classes are higher than large (>20) classes (Mateo and Fernandez, 1996). Those interpreting SOTEs should consider average class sizes at the department, college and university levels when comparing a candidate's scores to the norms, as class size may influence SOTE scores.

Student Level

Faculty evaluation ratings can be influenced by student level. Ratings in graduate and credential classes tend to be higher than in undergraduate classes (Arreola, 2000). Freshmen in the current norming sample tended to give slightly higher ratings, while seniors in the current sample gave lower ratings. The literature in this area is mixed in its findings, (Arreola, 2000; Aleamoni and Thomas, 1980; Stewart and Malpass, 1966).

Course Choice

 Students who have taken a class because of either an interest in the class or because of the instructor's reputation tend to rate their instructors more favorably

than students who take a course because it is required. Ratings given by students who are required to take a class are often lower than ratings by students for whom the class is an elective (Arreola, 2000).

College Level Comparisons

Instructors in the Colleges of Science and Engineering tend both to give lower grades and to be rated lower than instructors in the other colleges. There were also significant differences in rating between departments within colleges as well. In light of this, it is important that RTP committees evaluating candidates from different departments and colleges (University level RTP) compare instructors to colleagues within their own departments and colleges. (Arreola, 2000).

Instructor "Responsiveness"

The current SOTE includes a question about instructor responsiveness to diversity in the class (item #7). As indicated in the results of the Fall 2003 pilot, ratings for this item tend to have somewhat higher correlations with items 4, 5, and 6 (responsive to questions, established an atmosphere that facilitated learning, and approachability of instructor) and lower correlations with the other items. These correlations suggest that as a group, these items may measure students' perceptions of the instructor as approachable and responsive and that instructor responsiveness to diversity is equated with the instructor's general responsiveness.

APPENDIX A

Student Opinion of Teaching Effectiveness Rating Form (Adopted Fall 2003)

STUDENT OPINION OF TEACHING EFFECTIVENESS

Use only a #2 pencil to mark this form.
Completely darken the ovals you select.

San José State UNIVERSITY

INSTRUCTOR	CLASS	SECTION

Individuals needing assistance in reading or filling out this form due to a disability, please contact Institutional Planning and Academic Resources in ADM 112 to arrange for accommodations.

This instrument is designed to evaluate only your instructor's teaching performance. It is NOT designed to measure your reaction to the subject or to the facilities, such as the physical conditions of the classroom. Your individual ratings will be anonymous and a summary of items 1-13 will be available to your instructor after grades are turned in. This summary may enhance your instructor's teaching. It will also be used in the evaluation of your instructor for personnel matters such as retention, tenure and promotion.

This Instructor:

1. Demonstrated relevance of the course content:
- Very Strongly Agree
- Strongly Agree
- Agree
- Disagree
- Strongly Disagree
- Not Applicable/No Opportunity to Observe

2. Used assignments that enhanced learning:
- Very Strongly Agree
- Strongly Agree
- Agree
- Disagree
- Strongly Disagree
- Not Applicable/No Opportunity to Observe

3. Summarized/emphasized important points:
- Very Strongly Agree
- Strongly Agree
- Agree
- Disagree
- Strongly Disagree
- Not Applicable/No Opportunity to Observe

4. Was responsive to questions and comments from students:
- Very Strongly Agree
- Strongly Agree
- Agree
- Disagree
- Strongly Disagree
- Not Applicable/No Opportunity to Observe

5. Established an atmosphere that facilitated learning:
- Very Strongly Agree
- Strongly Agree
- Agree
- Disagree
- Strongly Disagree
- Not Applicable/No Opportunity to Observe

6. Was approachable for assistance:
- Very Strongly Agree
- Strongly Agree
- Agree
- Disagree
- Strongly Disagree
- Not Applicable/No Opportunity to Observe

7. Was responsive to the diversity of students in this class:
- Very Strongly Agree
- Strongly Agree
- Agree
- Disagree
- Strongly Disagree
- Not Applicable/No Opportunity to Observe

8. Showed strong interest in teaching this class:
- Very Strongly Agree
- Strongly Agree
- Agree
- Disagree
- Strongly Disagree
- Not Applicable/No Opportunity to Observe

9. Used intellectually challenging teaching methods:
- Very Strongly Agree
- Strongly Agree
- Agree
- Disagree
- Strongly Disagree
- Not Applicable/No Opportunity to Observe

10. Used fair grading methods:
- Very Strongly Agree
- Strongly Agree
- Agree
- Disagree
- Strongly Disagree
- Not Applicable/No Opportunity to Observe

11. Helped students analyze complex/abstract ideas:
- Very Strongly Agree
- Strongly Agree
- Agree
- Disagree
- Strongly Disagree
- Not Applicable/No Opportunity to Observe

12. Provided meaningful feedback about student work:
- Very Strongly Agree
- Strongly Agree
- Agree
- Disagree
- Strongly Disagree
- Not Applicable/No Opportunity to Observe

13. Overall, this instructor's teaching was:
- Very Effective
- Effective
- Somewhat Effective
- Ineffective
- Very Ineffective

OFFICE USE ONLY

SEQ#	COURSE ID#	#WC

-Continue on back-

SOTE1(Revised S04)

1

Please answer the following informational items.

14. What is your current estimate of your expected overall grade in this course?

- ○ A
- ○ B
- ○ C
- ○ D or F
- ○ Other (Credit/No Credit, Audit, Incomplete, etc.)

15. You are a:

- ○ Freshman
- ○ Sophomore
- ○ Junior
- ○ Senior
- ○ Graduate Student
- ○ Credential Only
- ○ Other (e.g., Open University, Audit, etc.)

16. Did you complete this form without undue influence from other students? ○ Yes ○ No

17. Did you complete this form without undue influence from the instructor? ○ Yes ○ No

SUPPLEMENTAL SECTION

This area is for use in responding to any additional rating items. Follow the instructions printed on the additional questionnaire for marking your responses. If no additional questions have been provided leave this area blank.

Item 18	Item 19	Item 20	Item 21	Item 22	Item 23	Item 24	Item 25	Item 26	Item 27
⑦	⑦	⑦	⑦	⑦	⑦	⑦	⑦	⑦	⑦
⑥	⑥	⑥	⑥	⑥	⑥	⑥	⑥	⑥	⑥
⑤	⑤	⑤	⑤	⑤	⑤	⑤	⑤	⑤	⑤
④	④	④	④	④	④	④	④	④	④
③	③	③	③	③	③	③	③	③	③
②	②	②	②	②	②	②	②	②	②
①	①	①	①	①	①	①	①	①	①

SOTE2

San José State
UNIVERSITY

This page will be returned to the instructor only after course grades have been released.

Instructor	Class	Section

Semester: ⬭ Fall ⬭ Spring   Year: ⬭ 2004 ⬭ 2005 ⬭ 2006 ⬭ 2007

● Discuss the strengths of this instructor's teaching.

● Discuss the weaknesses and/or areas in need of improvement for this instructor's teaching.

3

SOTE3

11

● **Please provide any other comments you feel would be helpful to the instructor regarding his/her teaching performance/ability.**
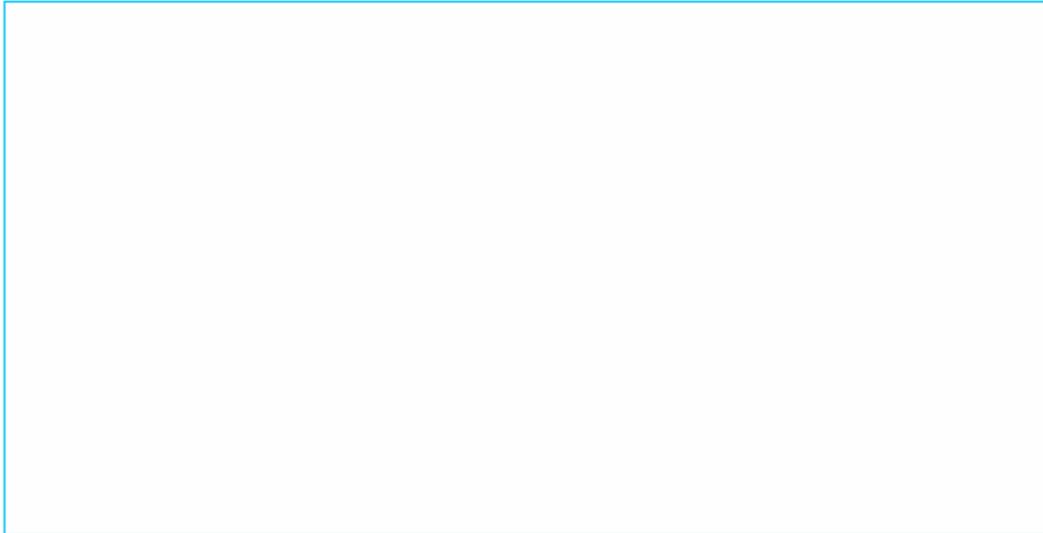
You may also supplement this anonymous questionnaire with a formal signed letter about this course to the department chairperson.

APPENDIX B

Student Opinion of Laboratory and Activity Teaching Effectiveness Rating Form (SOLATE)

STUDENT OPINION OF LABORATORY AND ACTIVITY TEACHING EFFECTIVENESS

Use only a #2 pencil to mark this form.
Completely darken the ovals you select.

San José State UNIVERSITY

INSTRUCTOR _____ CLASS _____ SECTION _____

Individuals needing assistance in reading or filling out this form due to a disability, please contact Institutional Planning and Academic Resources in ADM 112 to arrange for accommodations.

You are being asked to provide your opinion regarding the effectiveness of your instructor in the attitudes or behaviors listed below. Only the ratings of the class as a whole will be reported; individual student ratings will not be identifiable. You may also supplement this anonymous questionnaire with a formal signed letter to the department chairperson. DO NOT MAKE WRITTEN COMMENTS ON THIS FORM.

This evaluation form is designed for use in a course where laboratory, practicum, or field activities are the major part of the course requirements.

Use the rating scale below to respond to items 1 - 14. For each item below fill in the lettered circle you select for that item in the appropriate space to the right of the item.

A = Excellent
B = Above Average
C = Average
D = Below Average
E = Far Below Average
NA = Not applicable or no opportunity to observe

This instructor:-

1. Made course requirements clear.
2. Collected enough relevant information to assign grades.
3. Used fair and impartial grading methods.
4. Increased my understanding of the subject.
5. Was well prepared for class or activity.
6. Used class or activity time effectively.
7. Helped me apply theory or concept to class activity.
8. Showed concern for students.
9. Helped me learn the material.
10. Provided individual assistance as necessary.
11. Demonstrated proper and safe use of the equipment or technique.
12. Demonstrated or explained technical skills as necessary.
13. Was accessible to students during the class activity.
14. The overall effectiveness of this instructor is:

Please answer the following informational items.

15. Based on mid-term exams, projects, and graded assignments to date, what is your current grade in this class?
- A
- B
- C
- D or F
- Other(Credit/No Credit, Audit, Incomplete, etc.)

16. You are a:
- Freshman/Sophomore
- Junior/Senior
- Graduate Student
- Other (e.g., Open University, Audit, etc.)

SUPPLEMENTAL SECTION

The area below is for use in responding to additional rating items. Follow the instructions printed on the additional questionnaire for marking your responses. If no additional questions have been provided, leave this area blank.

17.   22.
18.   23.
19.   24.
20.   25.
21.   26.

OFFICE USE ONLY

SEQ #   COURSE ID#

References

Aleamoni, L. M., & Thomas, G. S. (1980). A review of the research on student evaluation and a report on the effect of different sets of instructions on student course and instructor evaluation. *Instructional Science, 9, 67-84.*

Arreola, R.A. (2000). *Developing a comprehensive faculty evaluation system.* Bolton, MA, Anker Publishing.

Mateo, M.A., & Fernandez, J. (1996). Incidence of class size on the evaluation of university teaching quality. *Educational and Psychological Measurement, 56(5)*, 771-778.

Stewart, C. T. & Malpass, L. F. (1966). Estimates of achievement and ratings of instructors. *Journal of Educational Research, 59, 347-350*.

Theall, M. (2002) *Student ratings: Myths vs. research evidence*: Focus on Faculty, Faculty Center newsletter article, BYU.

1998 INTERPRETATION GUIDE FOR SJSU SOTE AND SOLATE SURVEYS

Prepared by The SJSU Student Evaluation Review Board (September 1997)
Revised by the Instruction and Student Affairs Committee (January 1998)

This guide should be used by all SJSU faculty, all Department/School, College, and University Retention, Tenure, and Promotion (RTP) committee members, and all others who make judgments based on data from Student Opinion of Teaching Effectiveness (SOTE) and Student Opinion of Laboratory and Activity Teaching Effectiveness (SOLATE) surveys. Information from SOTE and SOLATE surveys are but one source of information for assessing teaching effectiveness (see Senate Policies S91-9 and S94-6). Other sources of information about teaching effectiveness should be employed before reaching an RTP decision.

Responsibilities of Candidate Faculty

It is the responsibility of individual faculty members and their colleagues to ensure that other sources of teaching effectiveness, e.g., peer evaluations, departmental or individual instructor's course evaluations, and student letters, are collected and included in personnel dossiers (in accordance with University policy). If an item mean is consistently above or below the norm range, the faculty member should provide further information about that rating. It is the responsibility of the faculty member to ensure that information about factors that influence student opinion ratings be included in the dossier, along with the ratings. For example, the faculty member should note whether the class was composed primarily of people required to take the course. (See discussion below: Factors to be Considered When Interpreting the Ratings.)

Statistical Display

The upper half of the first page of SOTE and SOLATE reports provide shortened versions of each of the 14 questions. To the right are four double columns of numerical data. The first double column lists means and standard deviations for the class in which opinions were gathered. The second, third, and fourth double columns list the norming data collected in F89/S90 aggregated across the Department/School, the College, and the University. The lower half of page 1 displays frequencies of responses for the individual items from which the above means and standard deviations are calculated. Also included are self-reported students' expected grades, self-reported class level, and actual average class GPA for the class at the time the SOTE/SOLATE report was generated. Page 2 is a graphical display of class means superimposed upon
College and University rating "norms."

The Statistics

The mean is the arithmetic average of student responses in which a score of 5 is assigned to the rating of "excellent," 4 to "above average," 3 to "average," 2 to

"below average," and 1 to the rating of "far below average." It is important to remember these descriptors when interpreting ratings. Means are reported to the first decimal place. Interpretation. The extent of agreement or disagreement on an item can be seen
directly from the frequency distribution for that item displayed at the bottom of the page. A less sensitive gauge of agreement is provided by the standard deviation. Most standard deviations are very close to 1.0. A large standard deviation (e.g., 1.3) indicates that students often do not agree about what rating should be assigned. A small standard deviation (e.g., 0.7) indicates that students generally agree about what rating should be assigned.

Ranges of Typical Values ("Norm Data")

"Norms" for each item are provided at the Department/School, College, and University levels. At each level, responses are aggregated over a specified norming period (most recently, F89/S90 for SOTE) to compute means and standard deviations which serve as reference points for making comparisons. Comparisons between the class data and norm data are best made using the graphic display shown on page 2 of the report.

Ranges of typical values ("norm ranges") for the College and University levels are graphically displayed on page 2. For each item, the middle 60% of ratings, from the 20th to the 80th percentiles, was determined for all classes surveyed during the norming period. This range is displayed as a line of dashes. The class mean is printed as an asterisk on the same line. Interpretation. If the asterisk is printed within the line of dashes, the class mean should be interpreted as no different from the norm group. If the class mean clearly falls outside the line of dashes, it can be concluded that the rating was below (to the left of the dashes) or above (to the right of dashes) that of typical scores. The usefulness and validity of a rating will be degraded
if ratings within the norm area are interpreted as anything other than typical. It is also important to remember the initial response descriptors when interpreting ratings (e.g., a score of "4" indicates "above average"). The mean score of most items is approximately "4." SOTE and SOLATE interpretation should use data across classes and semesters.

Factors to be Considered When Interpreting the Ratings

Many factors are known, through statistical research, to influence student opinion ratings. Therefore, ratings should always be interpreted with caution. Several, but by no means all, of the factors which have been shown to be consistently related to ratings are listed below.

1.   On the whole, research suggests that ratings are highly correlated to expected grades. Therefore two items are provided; both a distribution of the students' self-reported expected grades and the actual class GPA given by the instructor are to be found in SOTE and SOLATE report printouts on the bottom of

page 1. A distribution of the actual class grades given can also be routinely added to the printout by candidate faculty.

2.   Ratings in small classes tend to be higher than in large classes.

3.   Ratings in graduate classes tend to be higher than in undergraduate classes, and ratings in upper division classes tend to be higher than in lower-division classes. Self-reported class level is reported on the bottom of page 1.

4.   Ratings given by students who are required to take a class are often lower than ratings by students for whom the class is an elective.

5.   When a significant number of students in class leave an item blank or mark it "not applicable," that rating should be interpreted with caution. The number of students indicating these responses is reported in the frequency distribution on the bottom of page 1.

6.   Ratings from team-taught courses should be cautiously interpreted as students may be unable to separate their experiences from one instructor to the next.