

University Senate Task Force on Faculty Evaluations
FINAL REPORT
January 9, 2015

Task Force Members:

Andrew Baker, Marketing (Fall)
Marcie Bober-Michel, Learning Design and Technology (Senate Officer Representative)
Nola Butler-Byrd, Counseling and School Psychology
Morgan Chan, Associated Students (Spring)
Doug Deutschman, Biology (Chair of the Academic Resources & Planning Committee)
Philip Greiner, Nursing
Chad Karczewski, Associated Students
Glen McClish, Rhetoric and Writing Studies (Task Force Chair)
Cezar Ornatowski, Rhetoric and Writing Studies (Senate Officer Representative, Spring)
Paula Peter, Marketing (Spring)
Cory Polant, Associated Students (Fall)
Hongmei Shen, Journalism and Media Studies

The general charge for this Task Force, which was constituted by the Senate Officers in January 2014, was “to study and discuss current issues related to the use of student course evaluations and recommend changes to the current process.” The Task Force was also asked to consider several more specific issues:

1. Recommendations on core questions to be included in all evaluations: one each on instructor and course, up to 3 other standard questions, and a standard 5-point scale with standard scale point definitions.
2. A procedure which offers guidance on how to use course averages in the PDS with comparisons to comparable classes: lab, lecture, team, GE, core/major, graduate/undergraduate, upper/lower division.
3. Suggestions on dealing with the relationship of course evaluations to instructor effectiveness (if any), data reliability issues, and relationships to grades and popularity.
4. The importance of comments to augment the numbers, including perhaps standardized instructions for comments.
5. Recommendations for the discussion of evaluation results in the PDS.
6. As time and resources allow, a search of best practices at SDSU and at other universities and the scholarly literature on the use of student course evaluations.

The Task Force met frequently during the spring semester. We began by discussing the general issues before us, studying the forms currently in use across the colleges, and surveying the relevant literature on the topic. We soon broke into subcommittees to handle more specific tasks, and—when needed—consulted with Edith Benkov (Associate Vice President for

Academic Affairs) and Tom Packard (Chair of the Faculty Affairs Committee). We continued to work through Fall 2014, although mostly in subcommittees.

The Task Force's recommendations are presented below in four subsections:

- 1) Faculty Evaluation Forms and Questions;
- 2) Presentation of Statistical Results from Faculty Evaluations;
- 3) Relevant Criteria for Interpreting Faculty Evaluations at the Department, College, and University Levels; and
- 4) Student Involvement and Investment in Faculty Evaluation.

1. Faculty Evaluation Forms and Questions

In order to clarify the ultimate focus of the instruments—which is neither students nor courses, but faculty performance in courses—evaluation forms should be titled “Faculty Evaluation.”

For the purpose of clarity comparability across campus, responses to all quantitative items should be rated from 1 to 5, with 5 the highest (best) and 1 the lowest (worst). These numbers should correspond to the following descriptors: 5 = outstanding; 4 = very good; 3 = average; 2 = needs improvement; 1 = poor. Responses of “not applicable” or “does not apply” should be placed at the far right (after the “five” descriptor).

Items should emphasize criteria that are credibly evaluated by students (such as clarity of instruction, the organization of a course, perceived fairness, punctuality and reliability, ability to stimulate student interest, ability to communicate one's subject matter or expertise, and problem-solving ability), rather than criteria that students are not particularly well qualified to judge (such as the instructor's knowledge of the subject matter or teaching methodology).

If included on the form, demographic items (such as class standing, major, and so forth) should be listed first and should be clearly distinguished from evaluative items.

Evaluative items should be limited in number—no more than approximately ten quantitative items and no more than approximately three qualitative items.

Although evaluation forms will naturally vary from academic unit to academic unit and from college to college, each form should contain a subset of five common questions and a composite mean (or overall average) that together constitute universal reference points or common ground across the university's faculty evaluation process. We recommend these common questions:

- Rate the course's overall organization and structure.
- Rate the instructor's focus on the student learning outcomes listed in the syllabus.
- Rate the usefulness of the instructor's feedback on assignments and/or exams.
- Rate the clarity of instruction.
- Considering the criteria featured above, rate the instructor's teaching overall.
(This question is intended as the final or summative item).

In addition to these quantitative items, each form should contain at least two open-ended, qualitative items prompting students to provide substantive written comments. These items should solicit both positive commentary and critique/suggestions for improvement, as the following items indicate: “What were the instructor's particular strengths?” and “In what ways might the instructor improve this course?”

2. Presentation of Statistical Results from Faculty Evaluations

Evaluations of faculty provide important feedback concerning their teaching. In addition, they are used in the RTP and Periodic Evaluation processes. As a result, it is important that the reporting of the quantitative results represents the information in an accurate and concise manner. We propose a simple and robust way to present an instructor's evaluations with respect to a well-defined comparison group.

Faculty evaluations reflect students' experience in the class relative to their expectations. Evaluations are influenced by many factors, including class grades, class size, course modality, and course level. These factors are often intertwined. Lower division classes are often larger and have lower average grades than upper division or graduate classes. An appropriate comparison group needs to be broad enough to provide robust and stable information while being narrow enough to avoid comparisons among dissimilar classes. We propose that the comparison group be defined at the department level based on three categories: lower division, upper division, and graduate courses. This approach distinguishes between levels of instruction (which is correlated with class size and expected grade) while still allowing for a reasonably sized comparison groups.

Currently, an instructor's averages (more specifically, arithmetic means) are compared to the departmental means for an unspecified number of courses. There are two important weaknesses with this approach. First, these averages will be calculated even if the comparison group is very small group (e.g. $n = 2$). Second, such an average is an inadequate description of the distribution of values in the comparison group. The average provides no information on variability and is strongly influenced by unusual (extreme) values. We recommend that the entire distribution of scores be presented using a box plot based on the distribution of scores within the department (**Figure 1**) provided that the comparison group is 10 or larger in size. An illustration of how this would look is presented in **Figure 2**. If there are fewer than 10 comparable courses, only the median for the comparable courses will be displayed.

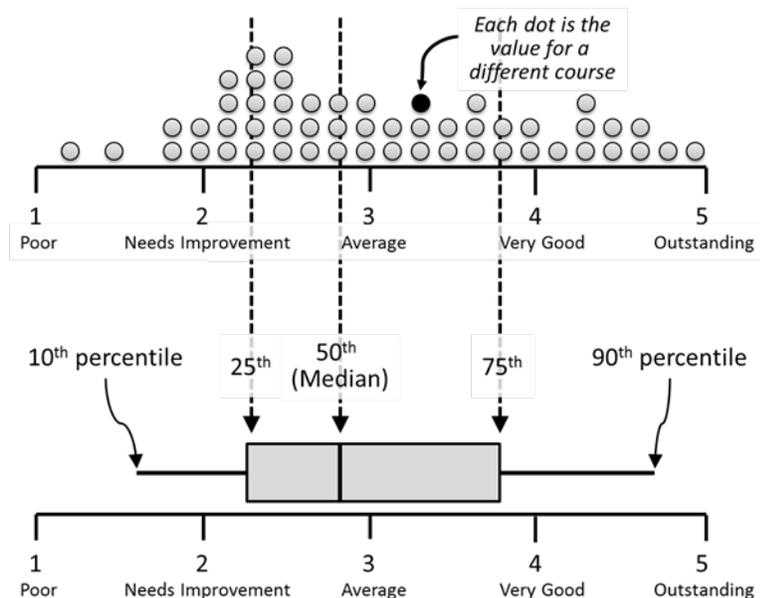


Figure 1: Information about the distribution of values for the comparison group (dots on the upper panel) is represented by **five different quantities (percentiles)** of the distribution.

The box of the box plot is defined by the **1** 25th percentile (left side of box) and the **2** 75th percentile (right). The **3** 50th percentile (= the median) is denoted as a vertical line inside the box. Finally, the whiskers (horizontal lines) are drawn to the **4** 10th percentile and **5** 90th percentile. In this particular box plot, the median falls at approximately 2.8, the 25th percentile begins at ~2.3, and the 75th percentile begins at ~3.8.

Illustration of Box plot display: Student evaluations for a set of universal questions for Instructor X. The graphic is annotated to show the richness of information that is depicted (Figure 2).

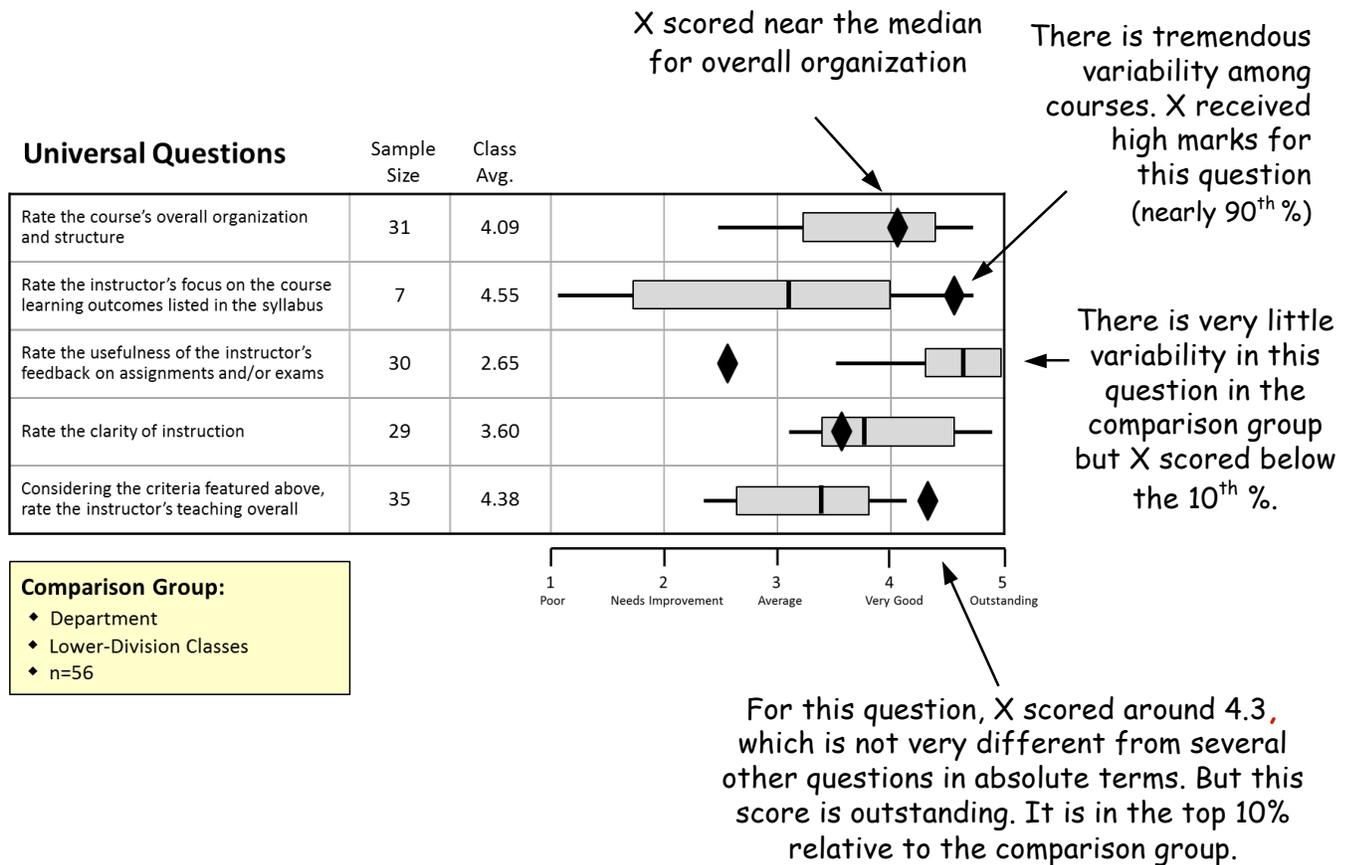


Figure 2: Illustration of the use of box plots to display student evaluation of the course/instructor. The size and composition of the comparison group is shown (lower left). The box plot depicts the distribution of the comparison group. The diamond marks where the mean score for this course/instructor. The graphic is annotated (text and arrows around the graphic) to aid interpretation. A graphic like this would be accessible both to the instructor and to those evaluating his or her teaching.

3. Relevant Criteria for Interpreting Faculty Evaluations at the Department, College, and University Levels.

The following criteria should be considered by committees and individuals who use faculty evaluations to assess the performance of faculty. They are also designed to help instructors better understand the strengths and weaknesses of their teaching.

- Course modality (face-to-face, hybrid, online)

Online courses might yield lower faculty evaluations than face-to-face courses because of possible difficulties raised by the use of technology (e.g. connection problems or lack of personal contact with faculty).

- **Course types (seminar/lecture/lab/studio)**

Seminars, labs, and studios have a tendency to be evaluated higher than lecture-based courses because of their relatively small class size and the interactive nature of the course type. In addition, generally speaking, the smaller the class, the higher the variance across terms.

- **Course levels (lower division/upper division/MA, MS/ PhD)**

Students' motivation may be greater in upper-division (more specific) than lower-division (more general) classes, which may affect the students' evaluation of the instructor.

- **Class function (prerequisite/major/elective)**

Students' motivation may be greater in elective/major than prerequisite classes, which may affect the students' evaluation of the instructor.

- **Class size (e.g., 7/35/150/300/800)**

The larger the class size, the more difficult it is to engage students in the course. Engagement inevitably influences the instructor evaluation. Furthermore, small sample size is highly variable and more extreme.

- **Academic discipline**

Disciplines engage students differently and therefore comparisons across disciplines should be avoided.

- **Team taught vs. single instructor**

Team taught courses may create challenges for coherence and consistency, as well as confusion about evaluation. For example, if three instructors collaborate on the teaching of a course, it may be difficult to sort out which student comments and assessments correspond with which instructor. In addition, if an instructor is in charge of a large class that includes laboratory sections, teaching assistants may be the ones supervising those labs. A distinction should be made in terms of evaluation of the instructor and evaluation of the teaching assistants.

- **Student experience with evaluation process**

Lower-division students and new transfer students have less experience with courses than seniors have and this may affect the students' evaluation of the instructor.

- **Student response rate to questions**

Low response is not necessarily an indicator of bad teaching; it simply does not allow generalizing results reliably to the whole class.

- **Difficult issues or challenging topics**

Faculty who teach courses related to cultural diversity and other challenging subjects often receive low evaluations, as do faculty of color who teach predominately Euro American classes.

4. Student Involvement and Investment in Faculty Evaluation

Currently, students are asked to complete faculty evaluations at the end of each semester. However, many students have realized the evaluations are not required. Also, they have found little immediate value in completing these evaluations since many faculty keep students' grades updated on Blackboard. Many students, especially freshmen, are unaware the course evaluations even exist. Furthermore, students are generally unaware of how faculty evaluations are used by the colleges, by deans, and by other committees and administrators. Having investigated other universities' evaluation systems, we believe that increasing student understanding of the process and earning their buy-in will lead to a higher quantity of student responses and a better academic environment.

In order to improve student awareness of and buy-in concerning the faculty evaluation process, we provide two recommendations:

A. Communicating the Importance of Evaluations to Students

In order to obtain thoughtful and constructive responses from students, it is important that they understand *why* they are being asked to take time at the end of each semester to complete the optional evaluation surveys about their instructors. We recommend that Academic Affairs (in cooperation with Associated Students) organize a "campaign" each semester (somewhat along the lines of "no adds, no drops, no kidding!") targeted at students to promote the evaluations and communicate the value for students in completing them. The campaign should include efforts to educate students on how the evaluations are used for reappointment, tenure, and promotion purposes. The campaign should also work with faculty and colleges to encourage their students to complete the evaluation for their course.

B. Implementing Mid-Semester Evaluations

We recommend the practice of anonymous mid-semester course evaluation (most likely administered in class or through Blackboard). This evaluation would be encouraged, but not mandated for either faculty or students. The evaluation would provide a way for students to give feedback at the midpoint of the semester on what they find works and what parts of the course they would like to see improved, thus giving instructors a way to strengthen *ongoing* courses. We believe this procedure would provide students with a clear connection between evaluation and improved learning.

C. A Further Consideration

In addition to developing these two recommendations, we seriously considered a proposal to develop a list of supplementary evaluation questions that could be made available to students. These supplementary questions could provide more accurate information to students about faculty and courses than currently available sources such as RateMyProfessors.com.

The difficulty of such a proposal is that at San Diego State, faculty evaluations are inextricably tied to the collectively bargained procedures of RTP and Periodic Evaluation. Simply put, it is infeasible to attempt to include questions in faculty evaluation that can be made available to students. Although we greatly value the principled case made by the students who served on the Task Force, we are ultimately unable to forward such a recommendation. It is possible that students may be able to create a parallel evaluation site online that could ask general questions such as the following:

1. Your class level?

2. Your reason for taking this class?
 - a. GE
 - b. Major (required)
 - c. Major (elective)
 - d. Minor
3. What grade do you expect in this class?
4. I learned a great deal from this course. (Likert scale)
5. Do you recommend this course overall? (Y/N)
6. Do you recommend this professor overall? (Y/N)