

**The University of Washington
Faculty Council on Instructional Quality**

The Faculty Council on Instructional Quality met Tuesday, January 8, 2002, at 12:30 p.m. in 36 Gerberding Hall. Chair Jan Carline presided.

PRESENT: *Professors* Carline, Coutu, Devasia, McGovern, Reinhall, Simpson
 Ex officio Bridges, Conquest, Jacobson, Lowell

ABSENT: *Professors* Bierne, Hoffer, Kyes, Mulligan, Wells
 Ex officio Evans, Lewis, McCracken, Pitre, Stromberg

Jan Carline called the meeting to order at 12:35 p.m.

Agenda

The agenda was approved.

Minutes

The minutes of the December meeting were approved.

Discussion on Use of Student Evaluation of Teaching

UW faculty members are evaluated in several ways, including student evaluations. Students rate faculty in various categories; the resulting numeric ratings are posted on the Web to aid students in making course selections. The Council is interested in the interpretation, use, relevance, and validity of these evaluations and has asked Nana Lowell and Tom Taggart, Assistant Director of the Office of Educational Assessment, to share their expertise in a discussion of student evaluations of teaching.

Carline opened the discussion by referring to the 1994 book "Assessing Faculty Work" by Braskamp and to a large body of literature on the subject of student evaluations, all of which he used to compile the eleven-point email he sent to the Council (included later in these minutes).

Taggart cited the periodical American Psychologist (November 1998 or 99 issue) as another good source for this kind of information, and handed out several OEA publications as well.

Student evaluations have been in use at the UW since 1920, said Taggart, and were codified as "required" in 1975 (each faculty member must have a student evaluation at least once per year). Since the inception of the modern evaluation, software upgrades and refinements have transformed the evaluation process and reporting - this quarter alone, OEA will evaluate 4000 courses. Instructors can have their results sent to their Chairs, and selected results (typically six to eight items) are published on the ASUW Website. Teaching Assistants are also evaluated, but their results are not published.

Each year, OEA compiles Departmental Profiles for courses evaluated the previous academic year and sends them to Deans and Directors. OEA has little control over how the reports are actually used, but intended them to be used for three purposes:

- By instructors, as a measure of the effectiveness of their own teaching
- By administrators, as a measure of the effectiveness of their teaching staff
- By students, as a way to match the instructor's style to their own personal learning styles

Taggart agreed with Carline that student evaluations should not be used alone in evaluating merit, and that more than one evaluation should always be considered in any assessment. In tenure cases, for example, departments will typically ask to see all the evaluations a candidate has had.

In response to Carline's question, Taggart said the President of the University owns the student evaluation information according to a 1975 opinion by the Attorney General. The yellow comment sheets are privileged information between instructor and student, but everything else is subject to public disclosure laws. Evaluations are held until grades have gone out to ensure that there is no "retaliation" possible for low evaluations, but they are public records. However, only information that is available need be made public - no one can make undue demands to reconfigure data.

At present, student evaluation ratings are carried out to two decimal places and published. This is by custom; there is no requirement about how many decimal places to publish. Is there a real qualitative difference between an evaluation rating of 4.73 and 4.74 - and should faculty really be ranked this way? Some departments are making this kind of distinction for merit reviews. Nana Lowell said that the number of decimal places used is meaningful in a measurement sense, because the number of cases surveyed warrants two decimals. However they are not warranted on the basis of reliability in talking about a range of values - it might be better to use just one decimal. Carline agreed - error measurement shows that surveying the same students in the same course on a different day will produce different results - sometimes the differences are large. If student evaluations are the only instrument used to evaluate faculty, said Taggart, that's a misuse - and perhaps even an abuse - of the evaluations, whether the number of decimal places is one or ninety

Carline asked George Bridges who owns the student evaluation information; i.e., who can recommend changes and make decisions about the way the student evaluations are set up, published, or used. Who makes decisions about how much is public information and what must be made available? If FCIQ were to make recommendations about changes, said Bridges, everyone including OEA would listen - it's a faculty issue. "We're always looking for advice," he said, "so recommend away - let's make it better if we can."

Lowell commented that it would be desirable and valuable to develop a set of recommendations that could be used as a standard for chairs and faculty to use in reviewing and using the evaluation scores. Gerry Gillmore and the Council on Faculty Affairs, said Taggart, developed a Website that is a faculty resource for grading; a similar site might be developed for to help decision-makers and how faculty in the faculty evaluation process. Wayne Jacobson agreed and suggested that a site for junior faculty members might also be developed, to help them present their student evaluations in an informed way. Bridges suggested a meeting with one or two decision-makers from the various colleges, to discover what they think would be most helpful in revising the way the evaluations are reported and used.

Another question, said Carline, is how meaningful the survey data is when reported as numbers. It might be just as effective to report faculty ratings as "top 25%, middle, and bottom" or "excellent, good, etc.," rather than numerically. Lowell agreed that changes to the present system could certainly happen, but there should be a lengthy and considered discussion about them.

Carline polled the group on Bridges' suggestion to gather college decision-makers for a meeting on possible changes to the evaluations - there was consensus that this should be done. Bridges suggested inviting the College Council reps for Arts and Sciences, Engineering, and Business. The Council members agreed with this plan.

The Council discussed the relevance and substance of each of the eleven discussion items Carline proposed:

1. Student ratings must not be used alone in making decisions about merit of a professor's teaching. Other sources such as peer review of classroom sessions, peer review of curricular materials, and faculty self reflection must be reviewed in addition to student evaluations in order to gain a true sense of the teaching skills of a faculty member.

Council: Discussed and agreed (see preceding discussion).

2. More than one set of evaluations should be used in making any decision about merit. Ratings from at least five classes over several quarters should be included for analysis. If possible, courses should include a variety of formats, from lecture based to seminar format.

Council: Discussed and agreed.

3. Consider course characteristics when interpreting results. For example, large lecture courses typically receive lower ratings than smaller courses, and introductory courses for non-majors receive lower ratings than higher division courses for majors. Adjustments for course type should be made in order to have a fairer sense of the professor's teaching skills.

Council: Discussed and agreed. Taggart said adjustments are now made based on three factors - whether the course is in or out of the student's major, class size, and relative grade.

4. Students must be informed in advance how ratings will be used and who will have access to the ratings.

Council: Discussed and agreed. At present, the standardized instructions do not advise students the ratings are used; this could be changed and it might make a difference if they knew. Perhaps "who will see these ratings" should be included.

5. Use standardized procedures for administering forms. Forms should be completed during class time at least one week before the final testing period. Instructors should not be present during the time that the forms are completed, nor should they collect the forms themselves. Forms should be collected at the end of the session, rather than submitted by mail. Results of the evaluation should not be returned to the professor until after the grades for the course have been submitted.

Council: Discussed and agreed. Student responses will be more thoughtful if not done during finals week, although there may be fewer respondents.

6. In order for evaluations to be valid, a minimum of a third of enrolled students must be included in the results. Results may not be representative of the full class if less than a third respond. Alternatively, if the class is randomly sampled, all students included in the sample must respond.

Council: Discussed and agreed. Taggart said this is a good idea, and suggested the process might be changed so data would not be published unless it meets the minimum response rate. Devasia commented that a 100% response to a random sample might be difficult to do.

7. Use comparative data, but with caution. Colleges and departments should compile information about evaluations for faculty in their courses over time. Without comparative norms, professors can easily overestimate or underestimate this effectiveness. Comparative data should be calculated from more than a narrow population of instructors. Departments with fewer than thirty professors should not rely on departmental norms, but use norms calculated for a number of similar departments or for the school as a whole.

Council: Discussed and agreed in principle, with some reservations and many questions. Lowell said departmental norms are provided as a substitute for deciles and provide a basis for comparison. If there are no norms, and no deciles, there is no context. How big does the group have to be to give reliable norms? Does the norm relate directly to faculty members, or to the "pool" of many courses taught by the faculty? Bridges said in the accreditation process the unit of analysis is the faculty/course - a combination of the faculty member and the course taught.

8. Do not overestimate small differences. There is a tendency to overestimate the precision of student ratings because they are quantified. A faculty member at the sixtieth percentile on an item does not differ in any practical way from another at the fiftieth percentile. It is better to deal with much broader classifications, such as Excellent or Very Good.

Council: Discussed and agreed to further discuss the issue. This is the crux of faculty complaints and should be looked at. A goal would be to get meaningful information for improving teaching, without being too granular about published ratings. There is no consensus as to whether the tool is faulty, or whether the use of the tool is faulty, or both. Lowell said that too much change could cloud historical results.

9. Only global ratings of teaching effectiveness, such as "Overall teaching effectiveness," should be used in personnel decisions. Other, more specific, items should be used by the professor for review of specific skills and areas for improvement.

Council: Discussed and agreed to revisit and reword.

10. Give professors an opportunity to respond to evaluation results. Professors should have an opportunity to discuss the objectives of the course and how the teaching methods were used to meet that objective, and how circumstances in the course might have affected evaluations. For example, students tend to rate less traditional teaching methods lower than more traditional ones, even though they might be most appropriate for the objectives.

Council: Discussed and agreed.

11. Limit the use of evaluation forms. For stable courses, ratings tend to change very little over time and repeated evaluation may result in little new information while over burdening students and inhibiting faculty from attempting new teaching methods.

Council: After discussion, it was decided to drop this item.

For the next February meeting, Carline will invite representatives of the Colleges to discuss the issue. He will send the revised discussion items to them as talking points.

Wayne Jacobson is trying to book a computer lab for a meeting with the Council and Rick Ells of Computing and Communications, but so far has not been successful. He will advise.

The meeting adjourned at 2:00 p.m. *Minutes by Linda Fullerton, Recorder.*