

Guidelines for Use of Student Evaluations of Faculty

The current instrument and procedures for student evaluation of faculty, based in part on what is referred to as the Endeavor system, has been utilized at FDU since 1980 when it was selected and approved by a joint faculty/administration committee. Early in the development process there was extensive discussion, training, and reference material provided by the Administration to support faculty personnel committees and department chairs in their efforts to interpret and make use of the evaluation system. In recent years almost no interpretive material has been available. In an effort to reorient those involved, this brief compendium of relevant information is offered.

Components of the Evaluation Form

The standard form has four sections:

The *Overall Evaluation* section (Q1 through Q7) is an exact reproduction of a student evaluation developed at Northwestern University by Professor Peter Frey as part of an extended research project. The form was shaped by Frey through twelve versions in a six year period. The final version, consisting of seven questions, has been shown to be reliable and valid in a large number of different circumstances. The questions have been carefully designed to avoid asking the student to *judge* the faculty member. The items can be answered from direct classroom experience. The two factors, *rapport* and *pedagogy*, that are computed¹ from responses to these seven questions are as reliable as any in general use in higher education for the purpose of evaluating faculty for merit or status purposes. The results of Part I of the form are appropriate for direct inclusion in faculty files. Inter-faculty comparisons may be made based on the data as long as direct comparisons are limited to generally similar courses and those context-specific influences affecting the results are kept clearly at the forefront.

The *Course Evaluation* section consists of five questions (Q8-Q12) that tend to be more course oriented, and will provide some useful information beyond that from the first seven questions, but are not as appropriate for numerical comparisons between faculty. The data from these questions should be analyzed and evaluated by the faculty member, department chair and others with the detailed local knowledge necessary to validate and confirm the information.

The section headed *Instructor's Questions* (Q13-Q15) provides a limited vehicle for individual faculty to ask three questions of their own devising. The questions asked need not be made public by the instructor. All that is required is that a hand-out be prepared with questions A, B, and C listed, and given to the individual administering the evaluations.

¹The formulas for these items, determined by factor analysis, are: $Rapport = -0.2 \times (Q2, \text{ advanced planning}) + 0.5 \times (Q3, \text{ class discussion}) + 0.5 \times (Q4, \text{ personal help}) + 0.2 \times (Q6, \text{ grade accuracy})$, $Pedagogy = 0.1 \times (Q1, \text{ hard work}) + 0.4 \times (Q2, \text{ advanced planning}) - 0.2 \times (Q3, \text{ class discussion}) + 0.3 \times (Q5, \text{ presentation clarity}) + 0.4 \times (Q7, \text{ increased knowledge})$. The result of these formulas is to define a measure (*rapport*) of an instructor's ability to establish a good working relation with a class and a second independent measure (*pedagogy*) of an instructor's class room skills. The *pedagogy* score is designed and has been shown to be largely independent of the instructors "popularity". As a result, a demanding teacher who will generally score lower on *rapport* because of rigorous technique will not be penalized by also getting low *pedagogy* scores. Conversely, a popular, charismatic teacher will not receive an inflated *pedagogy* score solely as a result of being well liked. The two negative coefficients are necessary for this result.

The *Written Response* section provides an opportunity for students to provide short written responses to three general questions. The part of the evaluation form containing the written answers is separated from the answer sheet by the administering department, and given directly to the faculty member after the completion of the course. Since these comments can be quite idiosyncratic they should remain confidential, and not be used in faculty status decisions.

The evaluation should be administered about two-thirds of the way through a term or semester. This is late enough that all faculty will (or should) have returned some graded or evaluated work and late enough that the semester's patterns will have been established. It is far enough away from finals that no conflict will arise with either remaining class time or final exam anxiety. The evaluation should never be administered with the faculty member present in the room. However, use of a reliable student from the class or a staff person is acceptable.

Use and Interpretation of Evaluations

Student evaluations should be analyzed at the local level *before* inter-faculty comparisons are made. It is at this level that the department chair can examine and, where necessary, note additional factors to be considered when interpreting the evaluation results on a comparative basis. In general, context influences, but does not significantly alter, the results. Among the contexts to keep in mind when interpreting instructional ratings are the following:

- Averages based on small samples are unreliable; when instructional ratings are based on 10 or fewer students they should be viewed with considerable caution. Small differences (10% or less) in scores are not significant.
- Evaluation decisions based on ratings from three or more classes are more reliable than those based on ratings from only one or two classes.
- Students who major in different departments have different backgrounds and different expectations and often use nonuniform standards in making their ratings; therefore, the instructional ratings for individuals should be compared only when similar courses are involved.
- People tend to become more tolerant of others as they grow older; research demonstrates that freshmen rate instructors more harshly than do upperclassmen.
- Ratings on items which ask about the personal relationship between instructors and their students are influenced by class size; the larger the class, the lower the ratings.
- Instructors who grade stringently (i.e., many Bs and Cs) tend to receive lower ratings on items which ask about class discussion, student-instructor interaction, and satisfaction with grading. This influences *Rapport* not *Pedagogy*.
- Experienced instructors tend to be rated higher than less experienced instructors on presentation clarity and organizational skill but lower on student-instructor interaction.