

Faculty Senate Minutes April 6, 1998

Present: Anton, Beardsley, Bristow, Cooney (ex-officio), Hale, Hummel-Berry, Kay (ex-officio), Kirchner, Lind, Maxwell, Owen, Sloane, Steiner, A. Wood

Visitors: Baarsma, Fiegenger, Fikes, Jaime, Rundle

1. Approval of Minutes: The minutes of the meeting of March 23, 1998 were distributed for comment.

2. Chair's Report: Kirchner announced that there will be at least three candidates for faculty senate chair. Faculty are encouraged to sign the back of the ballot envelopes to certify their ballots.

3. New Business: Maxwell introduced Fikes and Fiegenger, members of the Ad Hoc Committee on Student Evaluations. Maxwell reviewed the charge to the committee: "To study student evaluations and their use in the faculty evaluation process, and to make recommendations to the faculty senate. In particular, the senate asked the committee to consider how student evaluations might be used better; to address the issue of validity of student evaluations; and to consider the consistency of interpretations of student rating data." Maxwell noted that approximately 100 faculty responded to the survey. He said that the survey was designed to gather data about the student evaluation form (SEF) itself, the use of the SEF form for providing feedback on teaching, and the evaluation's use for appraisal in the evaluation process.

Maxwell described some of the challenges in designing a valid student evaluation form. He noted that their committee tried to assess "face" validity which he defined as: the extent to which those being tested consider the measurement instrument to be valid. As Landy (1985) notes, face validity is

"... not really validity in the technical sense, it refers to a judgment made by those being measured... Face validity is important to ensure a positive attitude toward both the testing and the organization" (p. 67). Maxwell indicated that there are other forms of validity: construct, content, convergent/criterion, discriminant, consequential, empirical and logical validity. These forms of validity are more difficult to assess.

Fikes then presented a flow chart (see appendix 1) depicting the role of student evaluation of teaching in personnel decisions at UPS. He noted the survey attempted to assess opinions about four issues: 1) the structure of the form itself, 2) the usefulness of feedback from the form in changing teaching, 3) usefulness for performance appraisal, 4) external bias - factors other than teaching excellence which may offset teaching excellence.

Fikes observed that the proportion of negative responses by the faculty to survey items (i.e. less than favorable ratings) were viewed by the committee as the most meaningful summary statistic. Fikes reported that workload and expected grade were difficult items to interpret because of their ambiguity. Fikes then reviewed the data, summarizing the main trends of the survey:

- low average ratings
- high variability among respondents
- specific sub-populations (tenured/non-tenured, rank, gender) don't differ significantly

Fikes gave examples of data from each of the four areas that the survey attempted to address:

- Evaluation Form: Items 4 & 5: Fifty-nine percent negative response, suggesting that the numerical ratings on student evaluation forms were not particularly useful for assessing teaching excellence. However, only fifteen percent of the respondents felt that commentaries were not useful (thus 85% felt commentaries were useful). Cooney opined that comments matter more than ratings. He noted that some previous renditions of the student evaluation forms used only quantitative data, while other renditions gave only commentary. He suggested that the interpretation of some items in a survey could be very difficult.
- Changing teaching: Item 8: Nineteen percent of the respondents said student evaluations were useful for pedagogical change.

- Performance appraisal: Item 13: Thirty-seven percent of the respondents indicated that the student evaluation form was not useful for colleagues in assessing performance, 36% indicated that it was not useful in assessing performance by the faculty advancement committee, and 70% indicated that they were not satisfied how the student evaluation form was used by the administration in assessing performance.
- External bias: Item 19: Fikes noted that faculty opinion was that expected grade, workload and course demands, as well as other factors, can bias student ratings.

Fiegener then presented data on the semantic differential (SD) scales that were used to assess "trust." He suggested that a lack of trust in student evaluations could "corrupt the process of evaluation." He indicated that the eleven SD scales used in the survey tapped trust in the student evaluation and trust in the use of that information for evaluative purposes. He noted that part of the mission at UPS was to produce excellent teachers. Fiegener postulated that there were "fast feedback" (e.g. self-correction from student feedback) mechanisms and "slow feedback" (espoused norms) mechanisms to help determine what excellent teachers are.

Fiegener indicated that the SD scales produced large variability and that there was a tendency on the part of the respondents to not trust the data, nor to trust how the data were used. Their analysis attempted to discover whether "trust" varied systematically by any demographic category.

Fiegener's interpretation of the trust data suggested significant differences in the mean levels of trust in the following areas:

- The Humanities faculty are less trusting of the student evaluation numerical ratings and written commentary data than are social science or natural science professors.
- Faculty Advancement Committee members trust the numerical ratings and written commentary data more than do non-Faculty Advancement Committee members.
- Assistant professors are less trustful of student evaluation numerical ratings and written commentary data than are full professors.

Fiegener discussed the implications of low trust to the institution which were drawn directly from comments written by survey respondents and not from interpretation of the numerical data. Several survey respondents expressed their concern about the impacts of low trust (in the data and use of data) on the institution. Fiegener noted that the current evaluations highlighted four types of concern and presented faculty comments to illustrate these concerns. The concerns were: 1) that student evaluations lead faculty to emphasize the importance of intellectual "comfort." However, Cooney noted that almost all faculty said they never change pedagogy on the basis of student evaluations. 2) that the student evaluations lead to reduced rigor (dumbing down) in order to get good evaluations, 3) that the student evaluations lead faculty "to play it safe" to avoid negative feedback and 4) the student evaluations can lead to cynicism.

Fiegener speculated on the sources of low trust in the student evaluation forms. Their analysis attempted to discover factors that were most strongly associated with low trust levels. The factors they uncovered were: concerns that the items may not be valid, the possible biasing influence of other variables (e.g. course workload, grade) or that students may not be thoughtful in filling out evaluation forms.

Fiegener observed that concerns that the evaluation form captures too few of the elements of teaching excellence, and concerns that students lack the ability to assess teaching are not significantly associated with low trust (i.e. they don't appear to be "sources" of low trust.)

Fiegener speculated that low trust levels are associated with the belief that the SEF information would be more useful in faculty evaluations if campus and department norms were made available for comparison.

Cooney said that student evaluations only tap one aspect of the evaluation process. He wondered, "how would students rate the way faculty grade them?"

Maxwell then distributed eight recommendations based on the survey (see appendix 2).

With a glazed look in senators' eyes, the senate adjourned at 5:30.

Respectfully submitted:

Barry S. Anton

Appendix 2. Specific Recommendations

(1) Open the "black box"! This means to assure the faculty and the students, in an openly demonstrable way, of the accuracy and fairness of the process employed in the interpretation and use of evaluation data. Negativity and distrust of the process by faculty (face validity and consequential validity)

(2) We must design different forms for feedback and appraisal.

[According to most research findings, using a single form for both feedback and appraisal is inappropriate. Since teaching improvement and teaching appraisal are, by definition, different constructs, construct validity cannot be assured with a single instrument.]

Dual purposes of the evaluation form. (construct and content validity)

(3) Students must be clearly instructed about the purpose of the particular form they are completing for a particular class. [Currently, students are advised that their responses are used for both feedback to the instructor and for appraisal of the instructors teaching effectiveness for purposes of tenure and promotion. As a result, student responses are ambiguous as to whether they are for feedback or appraisal.] Dual purposes of the evaluation form.

(Construct and content validity)

(4) Questions on the form should take into account whether the class is small or large, team-taught, multi-disciplinary, lab, etc.

[The qualities of effective teaching will differ depending on the class type.] Failure to recognize differing constructs of effective teaching for different types of classes. (construct and content validity)

(5) Questions on the evaluation form should not ask about a particular teaching style. Stereotyping by teaching style (construct validity)

(6) The evaluation form should obtain information which assist in identifying and adjusting for the influences of students' grade expectations, motivation and prior interest, and workload. Each of these have been shown to have at least a moderate confounding effect on evaluations. In order to have discriminant validity, the impact of these biases must be accounted for. Contamination of evaluation information by external variables, or biases. (discriminant validity)

(7) Serious consideration should be given to establishing norms for each criterion of effective teaching we measure. The norms should be determined for each of the course types, i.e., small, large, team-taught, etc. [This is not a recommendation for the use of a strict, or absolute, numerical standard which must be attained or exceeded to receive a positive decision. Rather, these norms should be used to give an instructor a basis for determining the quality of their own performance relative to their colleagues. Lack of "norming" (empirical validity)

(8) An external consultant should be retained to conduct a comprehensive validity study of the student evaluation of teaching process including the design of the form, the manner in which the form is administered to students, and the process by which the information is interpreted and used in arriving at decisions.