TO: President O'Leary  
Dr. Ramaley  
Dr. Desfosses  
Dr. Hartmark  
Dr. Volkwein  

FROM: Patrick T. Terenzini  

DATE: October 12, 1982  

SUBJECT: Student Ratings of Instruction

This memorandum, requested by President O'Leary, describes the procedures (and rationale) we discussed at our September 20 meeting on student ratings of instruction in personnel policies and decision-making. It is intended as a basis for further discussion.

Purposes

The institution of a campus-wide set of items is intended to serve three purposes: 1) as a statement to the University community that teaching is important at Albany; 2) to provide department heads and deans with a means for dealing with (or for forcing them to deal with) singularly poor instructors; and 3) to provide standardized information on teaching ability for use in making decisions on merit increases, promotion or tenure.

Procedures

We seemed to agree on the following procedures:

A. Some small number of items (perhaps no more than 3 or 4) will be selected on the basis of their substantive and technical merit. These items will be recommended to departments for their use. They will be asked to try these items out and report the results.

B. Ratings based on these items will be required of all faculty members being considered for promotion, tenure or a merit salary increase, or teaching awards.

C. Campus-wide norms for these items will be developed based on (1) the college/school in which a course is offered; (2) the size of the course, and (3) the level of the course (the need for norms by course size and level is suggested by the SIRF pilot test). Norms will be established such that an instructor's rating in each course can be classified into one of three categories that differentiate the highest, middle and lowest ratings within the norm category. Category cut-off points might be established by agreement with governance committees at something like the 15th and 85th percentiles.
D. Once norms have been developed, when and how ratings are collected will be the responsibility of individual departments. Pre-printed, optically-readable forms containing the selected campus-wide items and space to add additional items might be available, but responsibility for distribution, administration and collection of the forms would rest with the departments. It may be necessary and desirable to establish some central point for receiving and scoring completed forms.

E. Departments would also be required to forward to appropriate individuals and committees the ratings on the campus-wide items for their candidates for promotion, tenure, merit increases and teaching awards. Ratings on additional items or other evidence of teaching ability would be welcome and encouraged.

Rationale

A decision to use only a small number of items for personnel decision-making can be defended on several grounds. First, the items are not intended to be diagnostic or used for instructional improvement, and experts in the field seem to agree that for personnel decisions, only a few items are needed. For example, Wilbert McKeachie, Professor of Psychology and Director of the Center for Research on Learning and Teaching at the University of Michigan, has written that

"a lengthy standard scale [is not] likely to be helpful for personnel decisions since it is unlikely to be equally well suited for different disciplines or different courses within a discipline. Since comparisons between instructors in different courses can at best be only very general, one should probably not attempt much more than to determine whether students rate an instructor as excellent, adequate, or poor. An item or two on the degree to which a course stimulated interest or curiosity, and perhaps another item or two on general effectiveness, should be sufficient for personnel purposes."

Second, there is sound precedent. The University of Illinois at Champaign-Urbana has an Office of Instructional Resources (within its Division of Measurement and Research) that has developed the "Instructor and Course Evaluation System." ICES has three items (common to all forms) that are used for personnel purposes ("Rate the course content," "Rate the instructor," and "Rate the course content in general"). Illinois has been in this business for over a decade and their measurement staff are recognized nationally as among the best. Closer to home, our School of Business currently makes personnel decisions and recommendations on the basis of two items: "Overall rating of the instructor," and "Overall rating of the course."

Finally, the most detailed and systematic review available on the validity of student ratings finds that, on the average, overall ratings of courses, instructors and instructors' skills correlated .47, .43 and .50, respectively, with student achievement. While natural and physical scientists may find such results unimpressive, social scientists in the house will recognize the evidence as compelling.
Some concern was expressed at our meeting about the reliability of a composite scale based on only two or three items. Using the SIRF pilot test data, we calculated the internal consistency reliability of various combinations of four SIRF items. Results indicate that a scale based on only the two, overall rating items ("Instructor, overall" and "Course, overall") would have a reliability coefficient of .93. The highest reliability was found for a scale comprising overall ratings of the instructor, the course, the instructor's ability to communicate material, and the instructor's ability to stimulate students' interest in the course material. The reliability of this scale was .96.

Both the Trustees' policies and our Faculty Handbook provide clear, formal bases for the use of student ratings in personnel decision-making. The picture is cloudier on the matter of whether such ratings shall come from some uniform, campus-wide instrument. University Senate Bill No. 197172-40 (approved) offers the first formal grounds for a uniform instrument. This bill charged the Executive Committee of the Senate to name an ad hoc committee "which shall recommend to the Senate adoption of a uniform instrument to be used for the purpose of student evaluation of teachers throughout the University." Promotion or tenure candidates would be required to submit ratings from that form. It would appear that the appointment of the Tompkins Committee in 1978 was the first (albeit, perhaps, unknowing) response to that piece of legislation. The EPC action in 1980 reaffirmed the recommendation to develop a uniform set of campus items to be used in personnel decisions and recommended that an evaluation instrument be designed which could include items added by individual instructors, by departments, and by students for other purposes (e.g., teaching improvement and course selection). This led to the SIRF pilot test.

The development of norms will be a central step in the installation of any campus-wide rating system for personnel decision-making. This stage will probably be the most difficult—both politically and technically. At the least, norm development will require a fairly large, representative sample of faculty members. Being representative, such a sample will doubtless include faculty members opposed to the plan who will decline to participate, thereby biasing the norms. Considerable political spadework will be needed to bring this off.