Chapter III—Supporting Data: Collection & Presentation

Since the documentation of teaching must be selective to avoid unnecessary bulk in a tenure file and to increase the efficiency of the review process for all involved, it is imperative that the supporting data presented be collected in a way that reduces bias, is representative and yields valid measurements. It is also important to understand that the manner in which the data is presented can itself bias the evaluative outcome. Fortunately this is an area which has stimulated a very prolific amount of research. This chapter will provide a synthesis of some of that work, focusing on two major sources of evaluation data: students and peers.

Use of Student Evaluation Data

It has been argued that students are not valid sources of evaluation information, that their numerical and written responses on questionnaires used to make tenure and promotion decisions are based on superficial criteria, like appearance and popularity. This assumption has not been empirically supported. "Based on the findings of the meta-analysis, we can safely say that student ratings of instruction are a valid index of instructional effectiveness. Students do a pretty good job of distinguishing among teachers on the basis of how much they have learned. Thus, the present study lends support to the use of ratings as one component in the evaluation of teaching effectiveness. Both administrators and faculty should feel secure that to some extent ratings reflect an instructor's impact on students."[1] Key in the use of student data is the notion that as a data source, it is only one component which is available for a committee to make an informed judgment. When incorporated into a thorough analysis, student evaluation data is useful not only because it represents the learner's perspective, but it can stand to round off a picture of a candidate's teaching quality when presented in relationship to peer data and data supplied from the candidate's own perspective.

Student data can be solicited and presented in both quantitative and qualitative forms. Quantitative data, in the form of numerical student evaluation questionnaire scores, was the more prevalent form that appeared in the tenure files that were analyzed during the preparation of the report *Evaluation and Recognition of Teaching*. Qualitative student evaluation data, in the form of written letters of evaluation, appeared in less abundance,

---

[1]Cohen, Peter. "Student Ratings of Instruction and Student Achievement: A Meta-analysis of Multisection Validity Studies." *Review of Educational Research*, Fall 1981, Vol. 51, No. 3, Pg 305.

although when such data was provided it took up a lot of space in the file, as in one case where a tenure file contained 88 student evaluation letters.

The potential to statistically manipulate quantitative data is a mixed blessing. Assuming the data was gathered properly—using questionnaire items which were validated to insure what was being measured by the question was, in fact, what the question purported to measure—quantitative data can be very efficient in that many teaching factors from many individual perspectives can be presented in relatively little space. However, in the analysis of tenure files which was conducted, many of the guidelines listed below were not followed which resulted in a students' picture of the candidate's teaching which was not only limited, but possibly biased.

Instrumentation:

Research in the area of student evaluation of instruction has resulted in the publication of over 2,500 studies. Much has been learned about proper questionnaire design. One finding is that the purpose of the evaluation should determine the format and kinds of questions included in the evaluation instrument. In general, summative evaluation questionnaires designed for tenure and promotion decisions contain fewer items than formative questionnaires. Summative instruments focus on global items ("Overall, how would you rate the quality of the instructor's teaching?") and use evaluative scales (Excellent, Good, Fair, Poor... or Strongly Agree...Strongly Disagree) rather than frequency scales (Frequently, Somewhat Frequently, Rarely, Never...)

The use of "core" items allows an individual's scores to be compared to scores determined from a group aggregation, such as interdepartmental, or across a college's faculty. Core items are more generic aspects of teaching which are not influenced as much by course design or size. Core items enable the development of normative scores so an individual can be validly compared to his or her peers. Examples of such core items which have been validated through controlled quantitative methods include:

> The instructor is well prepared for class.
> The instructor has a thorough knowledge of the subject.
> The instructor communicated his/her subject well.
> The instructor stimulated interest in the course subject.
> The instructor is one of the best Cornell teachers I have known.
> The instructor clearly interprets abstract ideas and theories.
> The instructor demonstrates a favorable attitude towards students.

The instructor is willing to experiment and be flexible.
The instructor encourages students to think for themselves.[2]

Administration of Questionnaires:

Research on questionnaire validity suggests that if the following guidelines are followed for administering student evaluation questionnaires, reliability and validity of results will be improved.

- response format should be clear and consistent
- students should remain anonymous
- students should be given adequate time to complete the questionnaire
- students should not be allowed to discuss their ratings while they are being administered
- questionnaires should be administered during the last 2 weeks of semester (but not the last day and not during or after an exam)
- someone other than the one being evaluated should administer the questionnaire, or at the very least, the one being evaluated should leave the room
- a student should collect the questionnaires and mail them to an independent office for scoring
- 80% minimum attendance of the student population in a course is necessary on the day an evaluation is administered
- don't use a numeric questionnaire in courses with fewer than 10 students (use open-ended, written response items instead.)[3]

Reporting scores:

How summative evaluation scores are reported in a tenure file or in the tenure/promotion process can bias that process, either positively or negatively. Some general principles for proper questionnaire score reporting include:

- report frequency distribution for each item
- don't carry mean scores beyond one decimal place
- <u>multiple</u> sets of scores should be provided for each type of course (survey, lab, seminar) and collected <u>over a period of time</u>
- narrative (qualitative) data from the candidate, colleagues or chair about the contextual circumstances of the quantitative student rating scores is an aid in their interpretation.
- normative data sets should be established yearly for course type (elective, required,

---

[2]Sell, G. R. & Chism, N. *Assessing Teaching Effectiveness for Promotion and Tenure: A Compendium of Reference Materials*, Center for Teaching Excellence, Ohio State University, January, 1988.
[3]*Ibid*.

lecture, lab, etc.) on both a department level and college level for comparison with a tenure candidate's own scores.
• appropriate normative data should be provided wherever possible

Figure 7 below is an example of a simple format for reporting student evaluations scores for a single course. Figure 8 is an example of a visually clear way of reporting a candidate's relative standing in relation to departmental normative data.

Instructor: Professor X	Course: 100, sec 22	Spring '92		# choosing each response				
		Standard		Strongly		neutral		strongly
Item	**Mean**	**Deviation**	**Count**	disagree	disagree	opinion	agree	agree
1. well prepared for class		4.6	0.8	27	0	1	1	7	18
2. knowledge of the subject		4.1	0.5	27	0	0	2	19	6
3. communication effectiveness		4.6	0.6	27	0	0	1	10	16
4. stimulated interest in course subject		4.5	0.6	27	0	0	1	11	15
5. one of the best Cornell Professors		4.5	0.7	27	0	0	3	8	16
6. clearly interprets abstract ideas and theories		4.4	0.6	27	0	0	1	13	13
7. favorable attitude toward students		4.5	0.5	27	0	0	0	14	13
8. willing to experiment & flexible		4.4	0.6	27	0	0	2	12	13
9. encourages students to think for themselves		4.3	0.6	27	0	0	2	15	10
10. You found the course intellectually challenging and stimulating?		3.8	0.8	27	0	2	6	14	5
11. Instructor's explanations were clear.		4.4	0.6	27	0	0	1	15	11
12. Instructor presented the point of view other than his/her own when appropriate		4.1	0.9	27	0	1	6	10	10
13. Feedback on examinations/graded material was valuable.		4.4	0.9	27	1	0	2	7	17
14. Methods of evaluating student were fair and appropriate.		4.1	0.7	27	0	0	4	15	8
15. Instructor gave lectures that facilitated taking notes.		3.8	0.8	27	0	2	5	16	4
16. Instructor contrasted the implications of various theories.		2.8	0.9	27	1	10	11	4	1
17. Students were encouraged to participate in class discussion.		4	0.8	27	0	1	6	13	7
18. You have learned and understood the subject materials in this course.		3.7	0.6	27	0	0	9	16	2
19. overall rating		4.6	0.6	27	0	0	1	9	17
					poor	fair	average	good	excellent

Figure 7

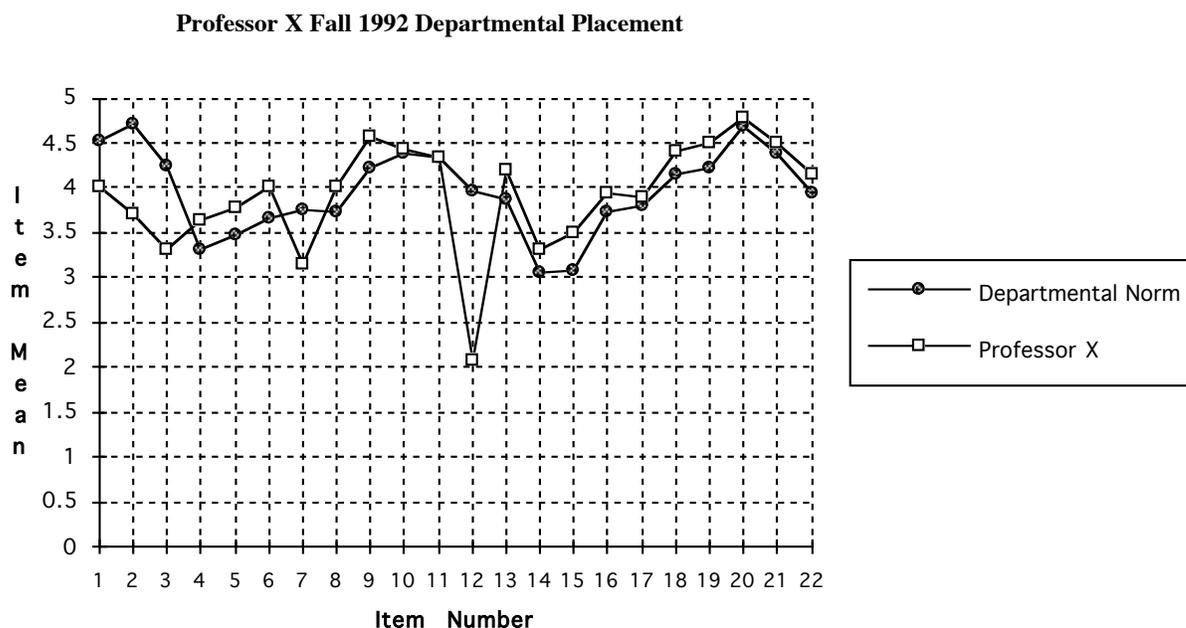**Professor X Fall 1992 Departmental Placement**

Figure 8

Qualitative data is less generalizable and harder to aggregate due to the fact that it is in a more open-ended form. Its potential bulkiness can be reduced through a synthesis by an objective individual familiar with the case, such as a department head. For others who must review this kind of synthesized data and who are less familiar with the candidate's situation, like deans and administrators outside of the department, a supplementary reflective statement from the candidate synthesizing the student letters can be useful in concert with the department head's report. If these reports are well written and address major developmental issues in the candidate's teaching practice, the time necessary to write them is well justified, especially if their creation leads to improved practice. The work of synthesizing can be spread out over time, on a year-by-year basis, as part of an annual review process.

During the preparation of the report *Evaluation and Recognition of Teaching*, the deans of the colleges were interviewed. One dean raised the issue of anonymity of student evaluation data. Quantitative questionnaire scores, combined with letters of recommendation, provide a good balance of general and specific information. However, letters in their original form do not preserve the anonymity of the student. While students, either undergraduate or graduate, are still working with the candidate they are in what one dean called the candidate's "power web." This may prevent students from being as candid in their written remarks if they know they may be identified at some point by the candidate during the tenure decision process. If letters by students are returned to someone other than the candidate—the department head or *Ad Hoc* chair, for example—and if they are then keyboarded on a computer and summarized by an independent person (a member of a departmental standing committee on teaching) and students are informed that these precautions are being taken when they are asked to write their letters, the validity of their responses will be enhanced.

An example of the department chair synthesizing the relevant comments from undergraduate student reviewers who were asked to write letters of recommendation is included below.

> ...undergraduates uniformly describe him as an unusually effective, conscientious, enthusiastic teacher who enables students to do their best work, master difficult subject-matter, and gain confidence in their own intellectual abilities.

> This [student quote from a review letter] clear and convincing testimony describes the experience of all the students who wrote to us from the courses he taught in spring 1988 and in fall 1989. Since the most disturbing aspect of some of the student responses two years ago was the suggestion that he could be authoritarian and coercive in his teaching, we are reassured by all these letters which suggest precisely the opposite.
>
> It seems clear that like many young assistant professors [candidate] was too demanding in his first dealings with graduate students, imposing admirable but often excessive standards of professionalism both in the classroom and as a special committee member, and expecting his students to share his commitment to his own projects. As the letter from [student] suggests, however, he has since become more realistic and flexible. And all the letters attest that he is always extremely conscientious and helpful.
>
> One should conclude, I think that [candidate] is an intellectually stimulating and enabling graduate teacher, with an expertise and commitment that many of our students find particularly valuable, one who has had trouble finding the appropriate mode in which to exercise authority, but who has now learned to do so.[4]

The *usefulness* and *reliability* of student letters of evaluation, whether undergraduate or graduate, can be improved if specific criteria are communicated when letters are solicited to help focus the students. If the students are all requested to respond to the same questions, reliability will be enhanced and it will be easier to summarize all the letters. The following is an example of the kinds of questions about teaching which can be used to aid students in writing evaluation letters:

1. Factual Knowledge: how well did the candidate help you acquire and integrate new terms, information and methods? Please give explicit examples where possible.

2. Concepts and Principles: how well did the candidate organize the material covered into a comprehensive whole? Were important concepts and principles from theory interrelated? Please give explicit examples where possible.

---

[4]A Report of the Select Committee, *Evaluation and Recognition of Teaching*. Cornell University, Ithaca, N.Y. Jan. 14, 1992. *Appendices*, pg. 21.

3. Application: Do you feel that the candidate's teaching and course structure enabled you to apply what you learned in the course to concrete problems? Were you able to generalize beyond the text? Please give explicit examples where possible.

4. Motivation: Did you feel the candidate was sufficiently motivated about the subject matter to excite your own interest in it? Describe how the candidate communicated a sense of enthusiasm about teaching.

5. Self Understanding: To what degree did the candidate help you become more aware of yourself as a learner? Describe specific experiences where the candidate contributed to your feeling empowered in your ability to learn.

6. Improvement of Instruction: Did the candidate seek out information from you and experiment with ways of improving his or her teaching? To what degree was the candidate open to feedback on improving the course? How confident are you in the candidate's ability to continually develop as a teacher? Please be as specific as possible.

To avoid biasing faculty opinion of a candidate's teaching effectiveness, student letters, in any form, summarized or not, should not be available to the voting faculty until all file data on both teaching and research has been assembled into the tenure file. This is true for all data: everyone voting on the candidate should have the same data base to make an informed and unbiased decision.

Peer Data

Evaluation of the candidate's teaching by peers is a practice which has become more prevalent in tenure and promotion decisions over the last 20 years.[5] During that time peer review has taken on an increasingly significant role in the tenure process. Effective peer review depends to a large degree on the explicitness of the criteria by which candidates are to be judged. Colleagues and peers are necessary contributors to evaluating a tenure candidate's teaching. They are best qualified to evaluate the candidate's breadth and depth of subject matter knowledge, course design skills and assessment strategies for determining students' learning the material. The information necessary for colleagues and peers to evaluate these kinds of skills must be thorough without being redundant. The

[5]Seldin, P. *Changing Practices in Faculty Evaluation*. Jossey-Bass, 1984.

candidate can help in peer evaluation by supplying the kind of information described in Chapter 2. However, colleagues from within the institution, both within the candidate's own department and outside it, and peers outside the candidate's institution, representing the discipline, will be required to provide their own data.

To be most effective the peer evaluation process should be neutral, open, relatively unthreatening and must be structured, all of which can be enhanced through the use of standardized rating and observation procedures and criteria. Standardization is a precaution stimulated by the evidence that colleagues' ratings may not be statistically reliable. In one study by Centra[6] the average correlation between colleagues was .26 per item. Another study[7] revealed the potential for positive bias of peer evaluation: fifty-four teachers were evaluated on the basis of 2 classroom visits by each of 3 different colleagues, which showed that 94% of all ratings were in the top 2 categories of a 5 point scale.

Several factors are critical in insuring a valid and fair peer review process. What questions are asked and answered by the reviewers is central. Some sort of replicable protocol is necessary to insure fairness and accountability for the process. This is true for whatever data is being reviewed, whether course materials, classroom performance or student learning. Developing a set of questions to focus the reviewer can make the task less arbitrary and subjective. The entire review process by peers should be governed by a set of procedures established within the department. Examples of such procedures include:

> ...peer ratings should be used in conjunction with student ratings. . . . dimensions [of teaching] should be decided upon in advance. . . [the] procedure should guarantee the anonymity and independence of the rater. . . . at least three colleagues be chosen to rate an instructor's teaching... these raters...may come from...an elected committee of the college faculty whose function is to evaluate teaching....raters do not meet together and preferably do not know who else is involved in the evaluation process. Rather, each judge independently rates the instructor on the preselected dimensions and submits the ratings to the dean [or department head], who then compiles a pooled rating for each dimension.[8]

---

[6]Centra, J. *Determining Faculty Effectiveness*. Jossey-Bass, 1979.
[7]Centra, J.A. "Colleagues as Raters of Classroom Instruction". *Journal of Higher Education*. 46, 1975, 327-337.
[8]Cohen Peter & McKeachie, Wilbert. "The Role of Colleagues in the Evaluation of College Teaching." *Improving College and University Teaching*. Vol. 28, no. 4 pg. 150.

Developing these procedures and the questions used to review the candidate can be a useful accomplishment of a departmental standing committee on teaching.

Qualification of Peer Reviewers

How peer reviewers are selected is another critical factor in establishing validity in peer review. No one should be placed in a position to review or observe a colleague for tenure or promotion decisions who is not qualified to carry out that task. A very consistent finding of peer observation studies is that observers should have some kind of training which prepares them for that responsibility. Peers are typically capable of evaluating subject matter knowledge, what must be taught by the candidate, whether the appropriate methodology is being employed for teaching specific content areas, the degree to which the candidate has applied adequate and appropriate evaluation techniques for course objectives, and the degree to which professional behavior has been exhibited according to current ethical standards. The following questions may be useful to determine to what degree an individual is qualified to undertake a specific peer review task:

- Do you believe you can properly judge the teaching-learning process in the classroom visited?
- Would you recommend this instructor to students advised by you? Why or why not?
- What specific changes are needed to strengthen teaching performance?
- How would you rate this instructor against others teaching similar courses in the department?