American Journal of Pharmaceutical Education **2003; 67 (4) Article 117.**

RESEARCH ARTICLES

A Comparison Between Student Ratings and Faculty Self-ratings of Instructional Effectiveness

Candace W. Barnett, PhD, Hewitt W. Matthews, PhD, and Richard A. Jackson, PhD

Southern School of Pharmacy, Mercer University

Objectives. This study compared the results of traditional student evaluations of classroom teaching with those of faculty self-evaluations and with the results of evaluations by smaller, representative subsets of students.
Methods. Students enrolled in required courses completed teaching evaluations, and 31 faculty members self-evaluated their instruction using the same 12 evaluation items given to the students. Students used a 5-point, ordinal response scale, and faculty used a visual analog scale. Within each professional year, representative subsets of 24 students were selected.
Results. There were no overall differences between the scores for faculty members' self-evaluations and the scores for evaluations by the whole class of students, with one exception: responses to the evaluation item "the pace of presentation." At the level of individual instruction, there was no significant difference between responses given by faculty members on self-evaluations and those given by whole-class ratings for a mean of 7.31 items. There were no differences between the overall ratings given by the whole class and those given by a subset of students from that class for 91.7% of the instruction sessions.
Conclusion. Faculty self-evaluations and evaluation by representative subsets of students can enhance the evaluation of faculty teaching.

Keywords: students, faculty, teaching, evaluation, self-evaluation, assessment

INTRODUCTION

Student evaluations can be a reliable and valid indicator of effective teaching.[1-5] Despite the documented reliability and validity of student evaluations, whether students have enough content knowledge to effectively evaluate teaching has been the subject of debate in the educational research literature.[6-8] In addressing this concern, faculty member self-evaluations, in addition to student evaluations, have been recommended as a more holistic approach.[9]

Evaluations administered at the conclusion of a course to all students enrolled remains the most common method used to evaluate teaching in schools of pharmacy.[10] A recent national study examined the attitude of pharmacy faculty toward student evaluation of classroom teaching.[11] Despite an overall neutral attitude, the faculty felt that the process by which student evaluations are administered does not allow students to give adequate thought to the items. The authors recommended that alternate methods of administration to students be explored. Problems in the administration of evaluations can be compounded when students are asked to complete evaluations of multiple faculty members involved in teaching a course. Such multiple evaluations occur in team-taught courses in the majority of schools of pharmacy.[10] Sampling a smaller, representative portion of students has been proposed as one alternate sampling method to reduce the number of assessments completed per student.[12]

The following study was conducted to address the methodological concerns about the evaluation of teaching that are outlined above. The objectives were (1) to determine the differences between the results obtained from student evaluations of faculty member performance with those obtained by faculty member self-evaluations of their performance on the same instruction, and (2) to determine the relationships between the

Corresponding Author: Candace W. Barnett, PhD. Mailing Address: 3001 Mercer University Dr., Atlanta, GA 30341. Tel: 678-547-6383. Fax: 678-547-6315. E-mail: Barnett_C@mercer.edu.

overall results of student evaluations completed by a whole-class and the results of evaluations completed by a representative subset of students from each class.

METHODS

Thirty-one faculty members at the Mercer University School of Pharmacy who were the instructors for the 13 required courses that spanned 1 semester within the first 3 professional years of the Doctor of Pharmacy curriculum were asked to participate. The faculty members completed self-evaluations of their teaching within courses. From 113 to 128 students (M =119) were enrolled in each course. Students completed separate in-class evaluations of each of the course faculty members at the conclusion of that faculty member's teaching.

The following customary evaluation procedures were followed: no faculty members were present in the classroom during completion of the evaluations. Two class officers distributed, collected, and submitted the evaluation forms. Per school policy, administration of student evaluations of teaching in all courses is mandatory for faculty members, and completion is voluntary for students. Tabulated results of student evaluations are returned to faculty following the conclusion of the semester in which the course is taught. Department chairpersons discuss the results of student evaluations with faculty members at the midpoint and annual performance appraisals.

Faculty members and students used the same evaluation criteria, which included 11 items measuring specific aspects of instruction and 1 item measuring overall teaching ability (Table 1). The students' evaluation form also contained a place for written comments, but these were not included as part of this study.

The evaluation items had been used for student evaluation of classroom instruction for the past 6 years. The items included those mentioned in the literature as possibly important in evaluating faculty instruction.[13-19] Students used the following 5-point ordinal response scale (1=poor, 2=below average, 3=average, 4=above average, 5=excellent). Faculty members used a visual analog scale (Table 1) consisting of a 4-inch line anchored at the ends by the terms "poor" and "excellent," respectively. Faculty rated each aspect of their teaching by placing a mark on the line provided. The location of the mark was measured for conversion to a numerical response. Faculty responses were not anonymous so that student and faculty member evaluations could be linked with the course to which they pertained.

Systematic sampling was used to identify subsets of students whose grade point averages collectively represented a cross section of the class. The literature indicates there is some relationship between grades and student ratings. Furthermore, there is a common belief among faculty members that the two are highly correlated.[20] From a list of students ranked in order of grade point average for each professional year, a random sample of 24 students was selected by choosing every name occurring at a predetermined sampling interval.[21]

The size of the subset was determined by computing the sample size required to estimate a population mean where the population size was 119, the standard deviation was 1.5, the bound on error was 0.5, and the confidence coefficient was 0.95.[22] Corrections for small populations and nonresponse were factored into the sample size. Subset members agreed to place a code identifying themselves as a member of the subset on the evaluation forms they completed so that statistical comparisons of their responses to those of the whole class could be made. Responses could not be linked to individual students.

Analyses were conducted using the STATISTIX program and included simple descriptive statistics, t tests, analysis of variance, and Scheffe's multiple comparison procedure.[23] Significance was set at $P<0.05$.

RESULTS

All 31 faculty members teaching required courses at the time of the study participated, resulting in a 100% response rate. Six (19.4%) faculty members represented the discipline of pharmacy administration; 12 (38.7%) represented pharmaceutical sciences, and 13 (41.9%) represented pharmacy practice. The faculty members had a range of 0 to 36 years of prior classroom teaching experience, with a mean of 14.1 ± 10.8 years. Thirteen faculty members held the rank of professor, 9 were associate professors, 8 were assistant professors, and 1 was an instructor. Twelve (38.7%) faculty members were women and 19 (61.3%) were men. The faculty members completed a total of 36 self-evaluations: 5 faculty members taught in more than 1 course and completed self-evaluations of their teaching for each course. Nine (25.0%) of the evaluations were for first professional year courses, five (13.9%) of the evaluations were for second professional year courses, and 22 (61.1%) were for third professional year courses.

The average student response rate for each class was 55.6% ±18.0% (range 19.7% to 83.3%). The response rate for the students in the subsets from each class averaged 77.1% ±13.5% (range 41.7% to 95.8%).

Table 1. Comparison of Faculty and Student Responses on a 12-Item Instructional Evaluation (N=36)

Evaluation Item*	Combined Responses of Faculty Members	Combined Responses from Students in Each Class	Combined Responses from a Subset of Students Identified in Each Class	F value.
	Mean (SD)	Mean (SD)	Mean (SD)	F
Teaching ability				
1. Pace at which the material was presented	3.37 (1.10)	3.89 (0.46)	3.84 (0.53)	5.08† ‡ §
2. Encouragement of class participation	3.87 (0.77)	3.99 (0.41)	3.97 (0.46)	0.43
3. Ability of the instructor to stimulate thinking in the content area(s)	3.93 (0.67)	4.08 (0.44)	3.98 (0.48)	0.75
4. Willingness of instructor to discuss content area(s) outside of class	4.37 (0.65)	4.24 (0.32)	4.23 (0.39)	0.97
5. Instructor's concern about the students' comprehension of the content area(s)	4.29 (0.50)	4.16 (0.39)	4.14 (0.45)	1.16
6. Overall teaching ability of this instructor	4.23 (0.67)	4.15 (0.42)	4.10 (0.47)	0.56
Teaching materials				
7. Handouts	4.13 (0.62)	4.09 (0.54)	4.13 (0.54)	0.07
8. Audio-visual aids	3.76 (0.85)	4.03 (0.46)	3.98 (0.47)	1.91
Course organization				
9. How well the objectives represented the content area(s)	4.17 (0.54)	4.16 (0.37)	4.11 (0.42)	0.23
10. Organization of the content area(s)	4.20 (0.59)	4.09 (0.47)	4.03 (0.48)	1.00
11. How well the exam questions related to objectives	4.38 (0.49)	4.07 (0.49)	4.00 (0.56)	5.11† §
Instructor's knowledge				
12. Knowledge of the instructor in the content area(s)	4.32 (0.53)	4.50 (0.23)	4.47 (0.30)	2.39

* Students used a five-point ordinal response scale: 1=poor, 2=below average, 3=average, 4=above average, 5=excellent.
Faculty used a visual analog scale: Poor _____ Excellent

† Significant at $P<0.05$

‡ Scheffe's test revealed a significant difference between the mean for faculty self-ratings and class ratings.

§ Scheffe's test revealed a significant difference between the mean for faculty self-ratings and subset ratings.

Each set of faculty self-evaluation scores for the 12 items was linked with the corresponding mean class score and mean student subset score to form 36 data records. While the data did not pass the tests of normality and homogeneity of variance for some variables, the sample size permitted use of t-tests and ANOVA even if the assumptions for validity were not completely satisfied.[24]

Comparison of Whole-Class Student Evaluations and Faculty Self-Evaluations

Analysis of variance was conducted on the combined data for all instruction to determine if there were significant differences among the mean ratings by the faculty members, whole-class, and student subsets for each evaluation item (Table 1). The only significant

Table 2. Frequency of Significant Differences* in Student Ratings† and Faculty Member Self-ratings‡ for Individual Instruction (N=36)

Item	Significant Differences§ n (%)	Size of Differences Mean (SD)
Pace at which the material was presented	18 (50.0)	1.30 (0.73)
Handouts	16 (44.4)	0.65 (0.37)
Audio-visual aids	15 (41.7)	1.05 (0.47)
Encouragement of class participation	15 (41.7)	0.87 (0.49)
How well objectives represented content areas	13 (36.1)	0.70 (0.50)
Organization of content areas	13 (36.1)	0.73 (0.62)
Knowledge of instructor in content area(s)	13 (36.1)	0.74 (0.46)
Ability of instructor to stimulate thinking in content area(s)	13 (36.1)	1.00 (0.64)
Willingness of instructor to discuss content area(s) outside of class	10 (27.8)	0.77 (0.64)
Overall teaching ability of this instructor	10 (27.8)	0.93 (0.46)
Instructor's concern about the students' comprehension of the content area(s)	8 (22.2)	0.83 (0.54)
How well the exam questions related to Objectives	7 (19.4)	0.83 (0.34)

*Based on t-tests

†Students used a 5-point ordinal response scale: 1=poor; 2=below average; 3=average; 4=above average; 5=excellent.

‡Faculty used a visual analog scale: Poor _____ Excellent

§When significant differences occurred, the faculty self-rating was lower for the item than the student rating.

difference found between the faculty members' self-evaluation scores and the mean scores for whole class evaluations was for the item, "pace at which the material was presented." The mean class rating on this item was 3.89 ±0.46, indicating an assessment by students of "above average," while the faculty members' self-rating on this item was 3.37±1.10, indicating a response of "average."

Analyses were also conducted at the level of individual instruction, defined as each instruction event. For every data record, t-tests were conducted on each of the 12 evaluation items to see how the faculty self-rating compared to the mean class rating. For each item, significant differences and the direction of the differences were recorded. For all 12 items and 36 records, when significant differences between faculty member and class ratings occurred, the faculty members rated themselves lower on the item than did the students in their class. The mean size of the difference ranged from a low of 0.65 ±0.37 for "handouts," and a high of 1.30 ±0.73 for "pace at which the material was presented" (Table 2).

The most frequently occurring difference between the faculty members' ratings of themselves and the ratings by the students in their classes was for the evaluation item, "pace at which the material was presented," for which half (50.0%) of the faculty members rated themselves lower than their students did. Other variables for which the faculty members' self-ratings differed from that of the mean class rating included "handouts," with 44.4% of the faculty members rating themselves lower than their class did, "audiovisual aids," with 41.7% of the faculty members rating themselves lower than their class did, and "encouragement of class participation," with 41.7% of the faculty rating themselves lower than their class did (Table 2).

Some of the items required the student to have knowledge of the course's content in order to complete the evaluation. These items produced differences in 36.1% of the records and included: "how well objectives represented content areas," "organization of content areas," "knowledge of instructor in content areas," and "ability of instructor to stimulate thinking in content area(s)" (Table 2). The fewest differences occurred for the evaluation items measuring "how well the exam questions related to objectives" (19.4%) and "instruc-

Table 3. Comparison of alternative methods of evaluating individual instruction in terms of the average number of evaluation items with and without significant differences* in ratings.† (N=36)

Variable	Mean (SD)	Range
Number of items (out of 12) where there was a significant difference in faculty self-ratings and student ratings.‡	4.19 (3.81)	0 - 12
Number of items (out of 12) where there was no significant difference in faculty self-ratings and student ratings.	7.31 (3.69)	0 - 12
Number of items (out of 12) where there was a significant difference in subset and class ratings.§	0.25 (1.03)	0 - 6
Number of items (out of 12) where there was no significant difference in subset and class ratings.	11.58 (1.08)	6 - 12

* Based on t-tests.

† Students used a 5-point ordinal response scale: 1=poor; 2=below average; 3=average; 4=above average; 5=excellent. Faculty used a visual analog scale: Poor _____ Excellent

‡ When significant differences occurred, the faculty self-ratings were lower than the student ratings.

§ When significant differences occurred, the subset ratings were lower than the class ratings.

tor's concern about the students' comprehension of the content areas" (22.2%)(Table 2).

For every record, tabulations were made of the total number of items (out of 12) for which faculty member and class ratings were significantly different. There were no significant differences between faculty member and class ratings for a mean of 7.31±3.69 items, and there were significant differences for 4.19±3.81 items (Table 3). For every item that a significant difference in responses was detected, the faculty member rated his or her performance lower than the students.

Comparison of Responses on Evaluations Completed by Whole-Class Enrollments and Representative Subsets of Students

Analysis of variance on the combined data revealed no significant differences in the mean ratings by the whole-class and student subsets for each of the evaluation items (Table 1). At the level of individual instruction, for every record, t-tests were conducted on each of the 12 evaluation items to see how the mean class rating compared to the mean subset rating. Significant differences and the direction of the differences were recorded. Also, for every record, tabulations were made of the total number of items (out of 12) where class and subset ratings were significantly different. There was no significant difference between class and subset ratings for a mean of 11.58 ±1.08 items (Table 3).

There was a significant difference between class and subset ratings for a mean of 0.25±1.03 items (Table 3). The subset ratings were lower than the class ratings in these instances. The mean size of the differences was 0.58±0.14). These differences occurred in 3 (8.3%) of the records. Thus, for 33 (91.7%) of the 36 records, there were no differences between class and subset ratings.

Limitations

The number of students completing the evaluation in each class varied. A possible explanation is the large number of evaluations and frequency with which students are asked to complete them, particularly in team-taught courses. The study proceeded on the premise that these results, regardless of response rate, are what are normally accepted as reflective of the class enrollment.

Readers should exercise caution in generalizing the results of the study to all pharmacy faculty members because data were collected at only one school of pharmacy. Analyses at the level of individual instruction were based on multiple t-tests performed on single data sets, which can increase the chance for error.

DISCUSSION

This study has documented that, overall, faculty member self-evaluations of their teaching and student evaluations of the same instruction produce similar results. When differences at the level of individual instruction between ratings from faculty member self-evaluation and student evaluation of teaching occur, the small size of the difference, while statistically significant, may not be meaningful. Furthermore, the lower self-ratings of faculty members seem to counter any implied concern they might have that students would provide artificially low ratings due to insufficient knowledge of the subject matter. Faculty members may simply be their worst critics. In areas where

faculty members are particularly judgmental with themselves, examination of student evaluation responses for the same area could provide balance. On the other hand, students simply may be overwhelmed by the expertise displayed by faculty members and feel unable to provide a useful evaluation. Future research comparing results from faculty peer evaluations with those from faculty member self-evaluations may help clarify this issue.

For now, a more comprehensive evaluation process, incorporating both student evaluations and faculty member self-evaluations should be employed. The importance of student evaluations should continue to be emphasized to the students so that their participation does not decline. With regular input through self-evaluation, faculty may move from feeling noncommittal to favorable toward evaluation by students. This, in turn, may increase instructional improvements and ultimately enhance learning.

Based on the results of this study, subsets of students, selected by statistically sound methods, can be an alternative means of evaluating teaching. This sampling scheme can be particularly useful in team-taught courses where multiple evaluations are done. Several subsets can be identified within the class enrollment and evaluation duties divided, addressing concerns about the number of assessments completed per student. With fewer evaluations to complete, response rates would probably be affected favorably as would the reliability of the evaluation results. If representative subsets were used, some students might feel they were being denied the opportunity to provide input. Therefore, it would be important to share with students how the evaluation data would be used for instructional improvement. Furthermore, it would be important to select new subsets frequently to increase the opportunities for more students to be selected. Faculty member self-evaluations and representative subsets of students are means by which schools of pharmacy can enhance the evaluation of faculty teaching.

REFERENCES

1. Aleamoni LA, Yoner M. An investigation of the relationship between colleague rating, student rating, research productivity and academic rank in rating instructional effectiveness. *J Educ Psych*. 1973;3:274-7.

2. Doyle KO. *Student Evaluation of Instruction*. Lexington, MA: Lexington Books, Inc; 1975.

3. Braskamp LA, Brandenburg DC, Ory JC. *Evaluating Teaching Effectiveness, a Practical Guide*. Beverly Hills, CA: Sage Publications, Inc: 1984.

4. Seldin, P. *Changing Practices in Faculty Evaluations*. San Francisco, CA: Josey-Bass Publishers: 1984.

5. McKeachie WJ, Pintrich PR, Lin Y, Smith D. *Teaching and Learning in the College Classroom: A Review of the Research Literature*. Ann Arbor, MI: Regents of the University of Michigan: 1986.

6. Ley R. Tenure and student evaluations of teaching. *Coll Student J*. 1981;15:147-150.

7. Prisco D. Student evaluation of instruction in higher learning. *Coll Student J*. 1979; 13: 290-3.

8. Rotem A, Glassman N. On the effectiveness of students' evaluative feedback to university instructors. *Rev Educ Res*. 1979;49:497-510.

9. Dennis LI. Student evaluations: Are they an appropriate criterion for promotion? *Nurs Health Care*. 1990;11:79-82.

10. Barnett CW, Matthews HW. Current procedures used to evaluate teaching in schools of pharmacy. *Am J Pharm Educ*. 1998;62:388-91.

11. Barnett CW, Matthews HW. Student evaluation of classsroom teaching: A study of pharmacy faculty attitudes and effects on instructional practices. *Am J Pharm Educ*. 1997;61:345-50.

12. Kasiar JB, Schroeder SL, Holstad SG. Comparison of traditional and web-based course evaluation processes in a required, team-taught pharmacotherapy course. *Am J Pharm Educ*. 2002;66:268-70.

13. Kotzan JA, Entrekin DN. Development and implementation of a factor0analyzed faculty evaluation instrument for undergraduate pharmacy instruction. *Am J Pharm Educ*. 1978;42:114-8.

14. Grussing PG, Valuck RJ, Williams RG. Development and validation of behaviorally-anchored rating scales for student evaluation of pharmacy instruction. *Am J Pharm Educ*. 1994;58:25-37.

15. Popovich NG. Effective teaching: One person's philosophy. *Am J Pharm Educ*. 1987;51:178-80.

16. Wynn JE. The role of good teaching and its recognition in the academic setting: A personal perspective. *Am J Pharm Educ*. 1987;51:186-9.

17. Kinnard WJ. Teaching and its encouragement (some post-decanal ramblings). *J Pharm Teaching*. 1994;4:3-18.

18. Colaizzi JL. Evaluation of faculty members. *Am J Pharm Educ*. 1995;59:75-84.

19. Arreola RA. *Developing a comprehensive faculty evaluation system*. Boston, MA: Anker Publishing Company, Inc; 1995: 129-141.

20. Arreola RA. *Developing a comprehensive faculty evaluation system*. Boston, MA: Anker Publishing Company, Inc; 1995: 86.

21. Marks RG. *Designing a Research Project*. Belmont, CA: Wadsworth, Inc; 1982: 106.

22. Marks RG. *Designing a Research Project*. Belmont, CA: Wadsworth, Inc; 1982: 142.

23. STATISTIX. Analytical Software, Tallahassee, FL (1984).

24. Marks RG. *Analyzing a Research Project*. Belmont, CA: Wadsworth, Inc; 1982: 3, 68-9;92.