# RESEARCH ARTICLES

# Student Evaluations: Are They Valid Measures of Course Effectiveness?

Robert S. Kidd, PharmD, MS, and David A. Latif, PhD, MBA

Bernard J. Dunn School of Pharmacy, Shenandoah University

**Objectives.** This study was designed to assess the relationship among students' grade expectations, actual grades, and evaluations of courses.

**Methods.** A total of 5,399 individual student evaluations from 138 course offerings taught over 4 academic years were compiled and analyzed. In addition to questions evaluating the course effectiveness, the instrument included a query of the grade the student expected to receive in the course.

**Results.** In this study, the students' grade expectations for a course had a strong positive correlation with the mean course evaluation score ($P < 0.001$, Pearson's correlation coefficient was 0.508).

**Conclusions.** The validity of using students' course evaluations as the primary measure of teaching performance for faculty performance reviews, recognizing and rewarding excellence in teaching, promotion, and tenure considerations, and other purposes is questionable.

## INTRODUCTION

Since the publication of a 1976 study[1] examining the methods used for teaching evaluations in schools and colleges of pharmacy in the United States, the use of students to evaluate faculty members' teaching effectiveness has increased by 60%, and is used in 100% (72 of 72) of schools and colleges of pharmacy responding to a survey concerning faculty evaluation.[2] Student course evaluations can be used for a variety of purposes within schools and colleges of pharmacy including recognizing and rewarding excellence in teaching, faculty performance reviews, assignment of teaching responsibilities, academic promotion, award of tenure, and merit salary increases. A recent study determined the type of evidence most frequently used to select faculty members at schools and colleges of pharmacy for teaching awards was students' rating of faculty instruction.[3] Another study of student rating determined the "Teacher of the Year" awards "…were generally determined by student vote."[4]

In spite of the widespread uses of student evaluations, a survey of attitudinal statements regarding student evaluation of faculty members found the question with the lowest respondent mean score on a 19-item survey was, "Students rating represent the best procedure for evaluating instructors' classroom teaching," which received a mean score of only 2.4 out of a possible 5.0 on a Likert scale.[5] These responses may result from questions about the validity, reliability, and utility of student evaluations since the widespread adoption of student evaluations in the early 1970s. Several pioneering experiments performed in actual courses taught in successive semesters have found that the manipulation of grades resulted in corresponding changes in course evaluations.[6-9] In other words, the higher the perceived or actual final grade, the more favorable the student rated the course and vice versa. This relationship has been challenged by others, and the debate continues as to the reliability, validity, and proper interpretation and use of student evaluations in other health professions including medicine, nursing, and dentistry.[10-12] This study was conceived and designed to assess the relationship among students' grade expectations, actual grades, and evaluations of courses at Shenandoah University's Bernard J. Dunn School of Pharmacy.

## METHODS

Prior to beginning our research, this study was approved by the Human Subjects Review Board of Shenandoah University. This investigation used a convenience sample and was a blinded retrospective record review of course evaluations. A power analysis, based on an estimated moderate effect size, used a 0.80 convention to determine the proper sample size. Using an a priori 0.05 significance level, 80 evaluated courses would be needed for this investigation.[13] A total of 5,399 individual student evaluations from 138 course offerings taught between the fall semester of 1999 to the spring

**Corresponding Author:** Robert S. Kidd, PharmD, MS. Address: Department of Biopharmaceutical Sciences, Bernard J. Dunn School of Pharmacy, Shenandoah University, 1775 North Sector Court, Winchester, VA 22601. Tel: 540-545-7217. Fax: 540-665-1283. E-mail: rkidd@su.edu.

Table 1. Summary of Sample

| Variable | |
| --- | --- |
| Number of years assessed | 4 academic years (1999-2003) |
| Number of course offerings evaluated | 138 |
| Number of student evaluations | 5,399 |
| Mean evaluation scores | 4.05* |
| Mean expected grade | 3.23† |
| Mean actual grade | 3.09† |
| Professional years assessed | P-1 through P-3 |

*Likert scale anchored at 5=Strongly Agree, 1= Strongly Disagree
†A= 4, B=3, C=2, D=1, E(fail)=0

semester of 2003 were compiled and analyzed. The fall of 1999 was selected as a start date for the study based on the power analysis, and because the same instrument had been used for course evaluations since the fall of 1999. The total number of students completing each course, and the mean grade obtained in each course were obtained from the university registrar. The course evaluations were typically administered in the final week of a course. The mean course evaluation scores were calculated as a numerical average of the 9 questions in the course evaluation (Appendix 1) that pertained to the students' perception of the course using a 5-letter Likert scale anchored at A = strongly agree and E = strongly disagree. The numerical means of these responses were determined using A = 5.0, B = 4.0, C= 3.0, D = 2.0, and E = 1.0. The mean grade the students expected to receive was calculated by taking the numerical mean of only the course grade question using the coding A = 4.0, B = 3.0, C= 2.0, D = 1.0, and E (fail) = 0. A Pearson r correlation test was used to determine if the mean grade the students expected to receive in a course was significantly correlated to the mean course evaluation score. In addition to the primary question, a Pearson r correlation test was also used to determine if the mean grade the students actually received in a course was significantly correlated to the mean course evaluation score. Finally, a Pearson r correlation test was performed to determine if there was a significant correlation between expected course grade and the actual grade received, and a t-test to determine if the means were significantly different. *SPSS* (version 11, SPSS, Inc, Chicago, Ill) was used to evaluate the data for statistical significance.

## RESULTS

Of the 217 didactic course offerings at the School from the fall semester of 1999 to the spring semester of 2003, evaluations were performed in 138 (64%). The 138 course offerings were comprised of 58 individual courses. Of the 7,474 students who completed the cours-

es, 5,399 (72%) completed a course evaluation. Table 1 summarizes the sample data.

The first analysis examined the correlation between the students' grade expectation and the mean course evaluation score for all the course evaluations compiled. The mean evaluation score was $4.05 \pm 0.54$, the mean expected grade was $3.23 \pm 0.52$, and the resulting Pearson correlation coefficient was 0.508. As shown in Figure 1, the Pearson r correlation revealed that students' grade expectations were significantly correlated to course evaluation scores ($P < 0.001$).

The second analysis examined the relationship between students' actual grades and their course evaluation scores. The mean actual grade was $3.09 \pm 0.61$, and the resulting Pearson correlation coefficient was 0.409. As shown in Figure 2, the Pearson r correlation revealed that students' grade expectations were significantly correlated to course evaluation scores ($P < 0.001$).

The final analysis examined the relationship between students' expected grades and their actual grades. The Pearson correlation coefficient was 0.854. As shown in Figure 3, students' grade expectations were significantly correlated with the actual grade they received ($P < 0.001$). The mean expected grade of 3.23
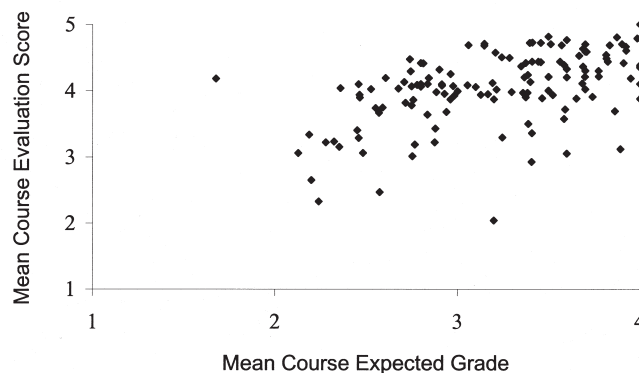


Figure 1. Mean course expected grade versus mean course evaluation score for 138 individual course offerings (r = 0.508).
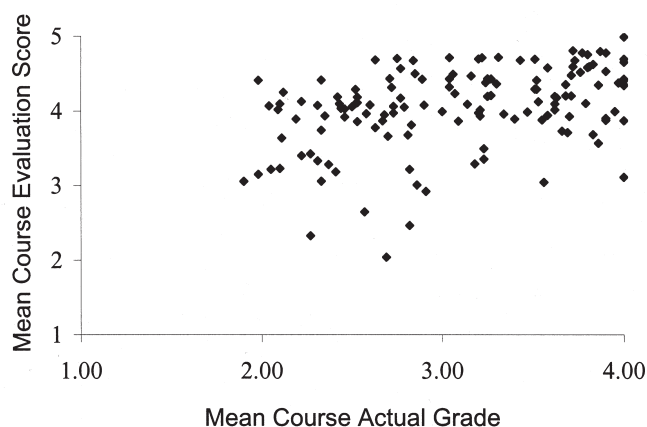
Figure 2. Mean course actual grade versus mean course evaluation score for 138 individual course offerings (r = 0.409).

was statistically significantly higher than the mean actual grade of 3.09 ($P < 0.05$). However, this represents less than a 5% difference between students' expected grades and their actual grades.

## DISCUSSION

The correlations of the first 2 analyses between expected grades, actual grades, and course evaluations were 0.508 and 0.409, respectively. These are quite high for social science research and may invite the question, "Are students assessing course effectiveness or are those assessments a function of the perceived or actual grades received?" Although disturbing, these results corroborate several prior studies that reported a significant correlation between students' grades and students' evaluations of a course. For example, one study found that there was a positive correlation between test performance and course evaluations.[14] However, this is the first study the authors are aware of in pharmacy education that has examined students' expected final course grades, and found them to be significantly correlated with the students' evaluations of a course ($P < 0.001$).

Not every previous investigation of student evaluations and course grades corroborates our findings. For example, a study failed to demonstrate that pharmacy students' course grades influenced their evaluation of the course. Al-Achi et al did not find a significant correlation between percentage of A and B grades and mean evaluation scores in courses offered by a pharmaceutical sciences department.[15] However, the total number of courses and evaluations examined was not given, and no statistical analysis was provided. Our study appears to be more comprehensive in terms of number of evaluations, number of courses, and breadth of courses examined.

The third relationship assessed was that between stu-

dents' expected grades and their actual grades. Interestingly, in this investigation, the mean expected grade was statistically significantly higher than the mean actual grade ($P < 0.05$). However, the difference between mean expected grade and actual grade was only 0.14, or less than a 5% difference, so there was not a practical difference between the students' expected grades and their actual grades. Therefore, students' expectations of their grades were realistic, which accounts for the strong positive correlation shown between both the expected grade and course evaluation score and between the actual grade and course evaluation score.

Although one cannot infer causality due to this investigation's research design, one can offer plausible explanations for the results. Kerr provided compelling evidence that management must reward the behavior it desires.[16] If faculty members at schools and colleges of pharmacy are rewarded with promotion, tenure, and merit raises based on student evaluations, it seems logical that some may lower their grading standards if they think it will result in more favorable evaluations. This widespread use and significance placed on student evaluations has been cited as one of the causes of grade inflation.[17] One study found that faculty members who were classified as "easy graders" were given better evaluations by students.[18] Goldman also concluded that good evaluations could be partially "bought" by assigning higher grades to students.[19] A recent study evaluated the overall mean grade point average of all pharmacy graduates from 1982 to 2002, and reported that grade point averages have consistently risen at a rate of approximately 1% per year for the past 20 years.[20] However, there was not a similar rise in PCAT scores or preprofessional grade point averages over the same time period.

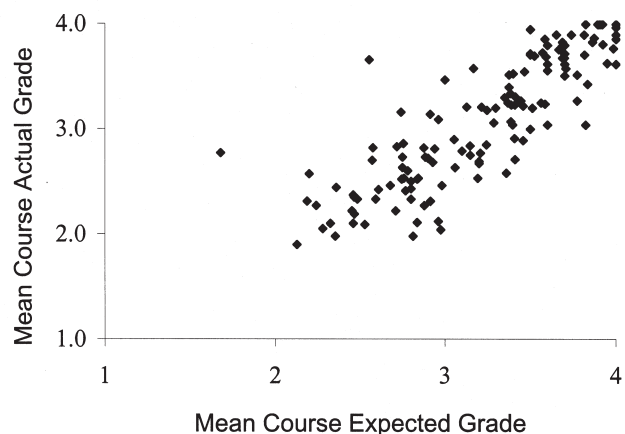One explanation for why students may rate courses more favorably if they expect better grades is that they



Figure 3. Mean course actual grade versus mean course expected grade for 138 individual course offerings (r = 0.854).

understand the material and have mastered it and therefore feel good about the course. An alternative explanation emanates from social psychology, in which praise, especially if greater than expected, promotes a liking of the praiser. In the context of student evaluations of teaching, it is known as the *leniency* or *grade-satisfaction theory* where the instructor assigns high grades and praise to the students, and in return, the students provide high rating for the instructor, who is the praiser.[6] When student evaluations play a significant role in performance evaluations and in recognizing and rewarding excellence in teaching, faculty members may be biased toward improving their evaluations of students, whether through grading leniency or other methods.

The determination of the effectiveness of faculty members in the classroom is a required component for many faculty evaluation processes. Based on the results of this study and others, the use of students' evaluations to evaluate instructor effectiveness may not be appropriate. However, these results should not be interpreted as a reason to completely abandon student course evaluations. There are good reasons to conduct student course evaluations, but the interpretation and use of the information needs to be carefully scrutinized. Even if student evaluations are biased by achievement, they still may contain helpful information for both formative use and student motivation. The formative use of student evaluations can benefit both the faculty member and future course offerings. Regarding the formative use of student evaluations, Barnett and Matthews found that over 97% of schools and colleges of pharmacy surveyed stated that faculty members used student evaluations for self-improvement and continued development of courses.[2] This formative use of students' course evaluations may be the most appropriate.

Another reason to use student course evaluations, even if they are biased by achievement, relates to student motivation. Greenwald and Gillmore likened the relationship to the patient's evaluation of a physician's bedside manner.[21] A physician's bedside manner does not determine his or her absolute worth as a physician, but it may be a good predictor of a patient's likelihood to return for follow-up visits and adhere to the physician's recommendations. Similarly, student course evaluations may reflect a student's willingness to attend class, do the assigned coursework, study the course materials, etc, which should have resulted in better student performance in the course.

From a simplicity standpoint, few assessment methods can match student evaluations. This is likely the reason 100% of schools and colleges of pharmacy use student evaluations, while only half use peer evaluation of classroom teaching, less than 20% use faculty self-appraisal, and less than 3% use all other methods surveyed.[2] Based on our findings, schools and colleges of pharmacy that do not use methods other than student evaluations should consider the incorporation of peer, administrative, and self- and/or expert evaluation into the faculty evaluation protocol as additional assessment tools. The utilization of multiple methods of assessment may provide more valid assessments, albeit at the cost of additional time and money. There are at least 3 limitations to applying the results of this study to other schools and colleges of pharmacy. The first limitation is this investigation used a convenience sample at one private school of pharmacy. However, the sample size for this study was relatively large: 5,399 individual student course evaluations for 138 course offerings were analyzed, and the correlation between students' expected grades and course evaluations scores was highly significant ($P < 0.001$).

The second limitation is that course assessment rather than teacher assessment was the dependent variable in this study. The instrument queried students about the course effectiveness, and not specific questions about the instructor's effectiveness (instructor's accessibility, helpfulness, teaching style, ability to explain difficult material, etc.); therefore, one may not generalize responses on course evaluations to instructor effectiveness.

Finally, the instruments used for course evaluations vary among schools and colleges of pharmacy, and only a single instrument was evaluated in this study (Appendix 1). Thus, it may be difficult to generalize these results to other schools and colleges of pharmacy whose course evaluation instrument may differ significantly.

## CONCLUSIONS

Because student course evaluations are commonly used in schools and colleges of pharmacy it is imperative that we understand the factors that influence students' course evaluations. In this study, students' grade expectations and actual grades had a strong positive correlation with the students' evaluations of the courses. Therefore, it is essential for schools and colleges of pharmacy to evaluate the proper and improper use of students' course evaluations. The formative use of course evaluations, and the use of additional methods of evaluation are encouraged. To confirm these results and answer the questions raised, these findings should be evaluated at other schools and colleges of pharmacy.

## REFERENCES

1. Henderson ML, Speedie S, Zanowaik P. Final report of the ad hoc committee on Teacher/Teaching evaluation of the council of faculties. *Am J Pharm Educ*. 1980;44:285-292.

2. Barnett CW, Matthews HW. Current procedures used to evaluate teaching in schools of pharmacy. *Am J Pharm Educ*. 1998;62:388-391.

3. Chism N, Bender E. Teaching awards: an analysis of criteria and evidence. Presented at the American Association of Higher Education 9th Annual forum on Faculty Roles and Reward Conference, February 2, 2001.

4. Draugalis JR. Teaching award and recognition programs in U.S. schools and colleges of Pharmacy. *Am J Pharm Educ*. 2002;66:112-118.

5. Barnett CW, Matthews HW. Student evaluations of classroom teaching: a study of pharmacy faculty attitudes and effects of instructional practices. *Am J Pharm Educ*. 1997;61:345-350.

6. Holmes DS. Effects of grades and disconfirmed grade expectancies on students' evaluations of their instructor. *J Educ Psychol*. 1972;63:130-133.

7. Powell RW. Grades, learning, and student evaluation of instruction. *Res High Educ*. 1977;7:193-205.

8. Vasta R, Sarmiento RF. Liberal grading improves evaluations but not performance. *J Educ Psychol*. 1979;71:207-211.

9. Worthington AG, Wong PTP. Effects of earned and assigned grades on student evaluations of an instructor. *J Educ Psychol*. 1979;71:764-775.

10. Fallon SM, Croen LG, Shelov SP. Teachers' and students' rating of clinical teaching and teachers' opinions on use of student evaluations. *J Med Educ*. 1987;62:435-438.

11. Cohen PA. Effectiveness of student ratings feedback and consul-tation for improving instruction in dental school. *J Dent Educ*. 1991;55:145-150.

12. Kirschling JM, Fields J, Imle M, et al. Evaluating teaching effectiveness. *J Nurs Educ*. 1995;34:401-410.

13. Hair JF, Anderson RE, Tatham RL, Black WC. *Multivariate Data Analysis*. 5th ed. Upper Saddle River, NJ: Prentice Hall;1998:11-13.

14. Berchou R, Sayenga E. Faculty evaluations in team teaching. *Am J Pharm Educ*. 1984;48:1-7.

15. Al-Achi A, Greenwood R, Junker J. Evaluating students' course evaluations: a retrospective study. *J Pharm Teach*. 1994;4:53-62.

16. Kerr S. On the folly of rewarding A while hoping for B. *Acad Manage J*. 1975;18:769-783.

17. Rosovosky J, Hartley M. Evaluation and the Academy: Are We Doing the Right Thing? Grade Inflation and Letters of Recommendation. Cambridge, MA: American Academy of Arts and Sciences;2002:9.

18. Wilson R. New research casts doubt on value of student evalua-tions of professors. *Chron High Educ*. 1998;44:A12-A14.

19. Goldman L. The betrayal of the gatekeepers: grade inflation. *J Gen Educ*. 1985;37:97-121.

20. Granberry MC, Stiegler KA. Documentation and analysis of increased grade point averages at a college of pharmacy over 20 years. *Am J Pharm Educ*. 2003;67(3):Article 77.

21. Greenwald AG, Gillmore GM. Grading leniency is a removable contaminant of student rating. *Am Psychol*. 1997;52:1209-1217.

Appendix 1. Course evaluation instrument.

**Bernard J. Dunn School of Pharmacy**
**Course Evaluation**

Please take a few minutes to seriously consider & complete this form. Your responses will be used as a part of the process of faculty evaluation of this professor(s) and this course.
A = Strongly Agree,  B = Agree, C = Neutral, D = Disagree, E = Strongly Disagree

Course:
1. The resources (e.g., textbook, notes, slides) used in this course contributed to my learning.
2. Integrated teaching was effectively used in this course (If Applicable). If the question is not applicable to this course, please do not select a choice.
3. I understood the subject matter of this course.
4. The content of the laboratory or recitation was a worthwhile part of this course (If Applicable). If the question is not applicable to this course, please do not select a choice.

Student Expectations of the Course:
5. Grade I expect to receive in this course. If you expect to receive the grade of F in this course, please select choice E.

Examinations/Grades:
6. Exams/Assignments accurately assessed what was taught in this course.
7. Complexity and length of course assignments were reasonable.
8. Methods of evaluation were fair.
9. Feedback on evaluations was valuable.
10. Graded assignments and examinations were returned in a timely fashion.