A DESCRIPTION OF EFFECTIVE COLLEGE TEACHING IN FIVE DISCIPLINES AS MEASURED BY STUDENT RATINGS

John T. Pohlmann, Ph.D., *Department of Guidance and Educational Psychology, Southern Illinois University, Carbondale, IL*

Student rating of teachers in five disciplines (science and math, education, social sciences, humanities, and business) were analyzed to determine which teacher attributes were important in predicting ratings of teaching effectiveness. Rating results from 1,439 courses taught at Southern Illinois University, Carbondale from 1973 to 1974 were used as data for this study. The results indicated that the instructor attributes rated as characteristic of effective instruction were highly consistent across disciplines, and the effective instructor was described as (1) knowing when students understood him, (2) increasing students' appreciation of the subject matter, (3) answering impromptu questions satisfactorily, (4) achieving the objectives of the course, and (5) giving several examples to explain complex topics.

Key words: student ratings, teaching effectiveness, instructor evaluation

The evaluation of faculty members is probably one of the most difficult decisions required of college administrators. Ideally, evaluative decisions should allow for objective and reliable comparisons of instructors on their contribution to the mission of their department. Most would agree that the primary mission of an institution of higher learning is to educate its students, hence student learning and professional growth is the ultimate criterion of teaching effectiveness. There are, however, methodological difficulties associated with the use of student achievement data for the evaluation of faculty. For example, Astin (1971) has found that much of the variance in college performance is predictable from input variables, such as performance in high school, socio-economic status, and scores on standardized achievement tests. Consequently, instructors teaching students

with low values on these input variables would tend to be evaluated poorly, if student learning was an evaluative criterion. It is also difficult to compare instructors teaching different disciplines or even different courses within a discipline on achievement data. Therefore, student achievement is a critical outcome of education, but methodological and measurement problems effectively preclude its widespread use for faculty evaluation.

As a result of the difficulties associated with the use of student achievement data, ratings by various reference groups (administrators, peers, students) are being widely used, and student ratings are one of the most frequently used means of evaluating college instructors. In many institutions results obtained from student ratings affect decisions regarding pay raises, retention, and promotion. If student ratings are to be used for such important decisions, faculty members should be apprised of the teaching characteristics which contribute to favorable student evaluations. Further, if there are discipline differences on what constitutes effective teaching using student ratings, this would have implications for the development of discipline-specific rating forms.

The purposes of this study were (1) to describe the differences in teacher attributes in five academic disciplines, (2) to identify the characteristics of teachers which account for variation in a general rating of teaching effectiveness, and (3) to determine if the teacher characteristics that contribute to high student ratings differ among five disciplines (science and mathematics, education, humanities, social sciences, and business).

In their recent review of the literature on student ratings, Costin et al. (1972) summarized a number of studies which attempted to assess the criteria used by students in their evaluations of faculty. The attributes of teachers which were most commonly mentioned by students as evidence of excellent teaching were preparedness, clarity, and stimulation of students' intellectual curiosity (Costin et al., 1972, p. 530). Deshpande et al. (1970), in a study of student perceptions of engineering instructors, found that the effective engineering instructor received high student ratings on motivation, structure, content mastery, and instruction skill. Issacson et al. (1963) related selected personality characteristics, as assessed by instructor self-reports and peer group nominations, to student ratings of the overall ability of teaching fellows in introductory psychology. These authors found that the effective teaching fellow possessed a personality structure which was described as artistic, polished, effectively intelligent, and imaginative. More recently, McKeachie et al. (1971) reported the results of a number of studies which examined the relationship between teacher warmth and effective teaching, as assessed by student achievement residualized for academic ability. Mixed results were obtained. In some courses, teacher warmth correlated positively with student achievement, while in other courses the relationship was negative. Turner (1970), upon reviewing the mixed results obtained in the McKeachie et al. (1971) study

and other studies, concluded that contextual variables, course type, student sex, etc., are potent factors in determining which instructor characteristics will prove to be effective.

Another group of studies which attempted to identify the characteristics of the effective teacher used simple student descriptions of the effective teacher. Downie (1952), in a survey of 16,000 college students, found that the attributes of a teacher that were listed as important were: comprehensive knowledge of subject matter, interest in the subject, being prepared for class, and motivating students to do their best. Crawford and Bradshaw (1968) subjected a number of teacher characteristics to a paired-comparison scaling analysis by various groups (student, administrators, and teachers) and those characteristics which obtained the highest scale values in terms of being essential for "effective University teaching" were: a thorough knowledge of subject matter, giving well-planned and organized lectures, enthusiasm and interest in teaching, and a student orientation and willingness to assist outside of the classroom. Gadzella (1968) asked a group of students to list criteria they would use for selecting the ideal professor. The four most important criteria selected were: knowledge of subject matter, interest in the subject, flexibility, and preparation. Costin (1968) had over 200 students rate the frequency of occurrence of various classroom behaviors exhibited by the "best lecturer" they had ever had. The attributes that received the highest ratings of frequency of occurrence were: acted interested in the material, was well prepared, used relevant examples, followed a logical sequence of thought, and explained clearly.

This series of studies suggested that college students equate effective teaching with three broad clusters of instructor attributes, knowledge of subject matter, organization of that subject matter for a clear and logical presentation, and a demonstration of an interest in the subject matter. These clusters indicate a strong subject matter orientation of students in the selection of effective college instructors.

METHOD

The Rating Form

The rating form used in this study was the Instructional Improvement Questionnaire (IIQ) (Pohlmann, 1973). The IIQ is designed to collect student evaluations of instructors and courses. Approximately 30,000 student evaluations contributed to the results. The students responded to the IIQ items using a 5-point scale (5 = exceptional performance, . . . 1 = improvement definitely needed). For the purposes of this study, only those items relating to instructor performance were analyzed. The IIQ items used in this study are presented in Table I.

Data

The data for this study consisted of the results obtained on the IIQ for 1,439 courses at Southern Illinois University, Carbondale in 1973 and 1974. These courses came from virtually every department on campus and every course level. The great majority of the courses analyzed in this study were of the lecture-discussion variety. The results for each course and the results used in these analyses consisted of item means on each of the items. Consequently, only between course rating variation was analyzed.

Each of the 1,439 courses that contributed data for this study was classified into 5 disciplines, science and mathematics, education, social sciences, humanities, and business.

The departments that were classified into each of the disciplines and the number of courses for each discipline were as follows:

I. Science and mathematics (N = 349)
- Biological sciences
- Botany
- Chemistry
- Geology
- Mathematics
- Physics
- Zoology
- Computer science

II. Education (N = 157)
- Elementary
- Secondary
- Special
- Physical
- Health

III. Social sciences (N = 596)
- Anthropology
- Economics
- History
- Psychology
- Sociology
- Political science

IV. Humanities (N = 249)
- Dance
- Language arts
- Music
- Philosophy
- Speech

            Spanish
            French
            English
V. Business (N = 88)
            Accounting
            Administration
            Business administration
            Finance
            Marketing

Statistical Analyses

The statistical analysis was conducted in two phases. The first phase compared the disciplines according to those attributes which received the highest and lowest evaluations. The second phase of the analysis compared the disciplines according to what teacher attributes the students in each discipline felt were important for effective teaching. The following section presents a step by step description of the analyses:

Phase I

Step 1. The item means on the IIQ for each course were converted to standardized T-scores. Each item mean was transformed to a scale where the university-wide normative course mean was 50 and the standard deviation was 10 (Table I).

Step 2. The T-scores obtained in Step 1 were then averaged for each discipline (see Table I).

Step 3. The items were then ranked in each discipline to determine the teacher attributes that were rated highest and lowest in each discipline.

Step 4. The five disciplines were then intercorrelated over their item means. The elements in the data matrix were the T-scores on each item for each discipline. The rows of the data matrix were the T-score means for each of the 21 HQ items. The columns of the data matrix were the T-score means on the 21 items for each discipline. This 21 X 5 (item means by disciplines) matrix was then intercorrelated by columns. The resulting R matrix was then interpreted as a discipline similarity matrix based on the discipline profiles across rating items (see Table II).

Phase II

Step 1. The item means on items 1−20 were correlated with item 21, the general rating item. If an item correlated highly with item 21, it was assumed that the teacher attribute assessed by that item was a good discriminating attribute to distinguish between effective and ineffective teaching from the students' perspective (see Table III).

Step 2. The items from the IIQ were ranked according to their correlation with the general rating item (item 21) in order to describe the important and

unimportant teacher attributes for each discipline (see Table III).

   Step 3. The disciplines were then correlated over the item correlations with item 21. The elements of the data matrix were the correlations between item means on items 1–20 of the IIQ and item 21. The rows of the data matrix were the items 1–20, and the columns of the data matrix were the five disciplines. This 20 X 5 data matrix was then processed to obtain a 5 X 5 matrix of discipline similarity coefficients based on student perceptions of what constituted effective instruction (see Table IV).

RESULTS

   The standardized rating scores on each of the IIQ rating items, for each discipline, are presented in Table I. Humanities courses received the highest ratings, followed in order by education, social sciences, business, and science and mathematics.

   Within each discipline, instructors tended to generate different rating profiles. Science and mathematics instructors received their highest ratings on (1) promptly returning homework and tests, (2) being dependable in holding class as scheduled, and (3) making clear assignments. Science and mathematics instructors received their lowest ratings on (1) encouraging student participation, (2) speaking understandably, and (3) knowing if students understood them. Education instructors received their highest ratings on (1) specifying the objectives of the course, (2) encouraging student participation, and (3) achieving the objectives of the course. Education instructors received their lowest ratings on (1) making clear assignments, (2) grading fairly, and (3) being prepared for class. Social science instructors reveived their highest ratings on (1) giving several examples to explain complex topics, (2) being prepared for class, and (3) showing an interest in the course. Social science instructors received their lowest ratings on (1) specifying the objectives of the course, (2) achieving the objectives of the course, and (3) showing an interest in students. Humanities instructors received their highest ratings on (1) answering impromptu questions satisfactorily, (2) being prepared for class, (3) making clear assignments, and (4) encouraging student participation. Humanities instructors received their lowest ratings on (1) specifying the objectives of the course, (2) promptly returning homework and tests, and (3) being available outside of class. Business instructors received their highest ratings on (1) making clear assignments, (2) being dependable in holding class as scheduled, (3) speaking understandably, and (4) promptly returning homework and tests. Business instructors received their lowest ratings on (1) showing an interest in students, (2) achieving the objectives of the course, and (3) increasing students' appreciation of the subject.

   In the next analysis, the disciplines were intercorrelated over their mean ratings shown in Table I. The resulting correlations may be viewed as profile similarity coefficients. A high correlation between two disciplines would suggest a similar

TABLE 1. Mean Standard Scores (T-scores) Obtained on 21 IIQ Items in Five Disciplines*

Item	Science and math (N = 349)	Education (N = 157)	Social sciences (N = 596)	Humanities (N = 249)	Business (N = 88)
1. Prepared for class	48.9	50.4	50.6	54.0	50.4
2. Made clear assignments	50.1	49.1	49.9	54.0	50.8
3. Set clear standards for grading	48.9	50.9	49.0	51.1	49.2
4. Graded fairly	49.1	50.1	49.3	53.2	48.5
5. Knew if students understood him	46.1	52.5	49.3	53.9	48.3
6. Spoke understandably	45.9	53.5	48.9	53.7	50.2
7. Answered impromptu questions satisfactorily	47.3	52.1	49.4	54.6	50.0
8. Showed an interest in the course	46.8	52.6	50.0	53.6	48.6
9. Gave several examples to explain complex ideas	46.4	51.7	50.8	53.3	48.5
10. Accepted criticism and suggestions	47.8	50.6	49.7	53.6	47.0
11. Increased your appreciation for the subject	46.6	53.7	49.6	53.4	48.0
12. Was dependable in holding class as scheduled	50.6	52.3	48.8	53.1	50.8
13. Specified objectives of the course	47.5	54.9	48.1	52.6	48.3
14. Achieved the specified objectives of the course	48.2	53.8	48.4	53.4	47.8
15. Promptly returned homework and tests	51.5	51.8	49.0	52.7	50.2
16. Showed an interest in students	47.3	52.6	48.8	52.8	47.7
17. Knew his subject matter	48.9	50.7	50.7	53.3	50.2
18. Was available outside of class	48.3	50.9	49.9	52.7	50.1
19. Encouraged student participation	44.6	54.1	48.7	54.0	49.4
20. The course was well organized	48.0	52.7	49.6	52.9	49.8
21. In general, taught the class effectively	47.7	52.0	49.2	53.9	49.1
Mean rating for all items	47.9	52.0	49.4	53.3	49.2

*A high mean connotes a favorable evaluation.

profile of mean ratings over the 21 items, and conversely a low correlation would imply a dissimilar profile of mean ratings over items. The results of this profile analysis is presented in Table II.

TABLE II. The Pearson Correlations Among Disciplines Based on Mean Ratings Across 21 IIQ Items (N = 21)

Discipline	2	3	4	5
1. Science and math	−0.56*	0.06	−0.32	0.44*
2. Education		−0.62*	0.00	−0.30
3. Social sciences			0.25	0.21
4. Humanities				0.08
5. Business				

*r significantly different from 0, α = 0.05, two-tailed.

The correlations in Table II indicate that education and social science instructors are the most dissimilar. Education instructors tended to receive high ratings on items where social science instructors received low ratings. A significant (p = 0.05) negative correlation was also observed between education and science and mathematics. A significant positive correlation was observed between science and mathematics instructors and business instructors, suggesting that a common set of strong and weak attributes was exhibited by instructors in those disciplines.

The next analysis consisted of correlating the mean ratings on items 1–20 with the mean rating on the general rating item (#21), "In general, the instructor taught the class effectively." This analysis was conducted separately for each discipline. If an item correlated highly with item 21, the attribute assessed by that item was assumed to be an important teaching attribute. This analysis is presented in Table III.

The correlations in Table III indicate that the great majority of the items on the IIQ correlated highly with item 21. Previous factor analyses of the IIQ (Pohlmann, 1973) revealed a strong general factor running throughout the questionnaire, so the high correlations in Table III were not too surprising. While the correlations in Table III tended to be high, they were not uniformly high, and the variation in the correlations was of primary concern in this study.

For the entire sample, the items which made strong contributions to item 21 were items 5, "Knew if students understood him," 7, "Answered impromptu

TABLE III. Correlations Between Item 21, "In general, the instructor taught the class effectively", and Items 1–20 from the IIQ for Five Academic Disciplines

Item	Science and math (N = 349)		Education (N = 157)		Social sciences (N = 596)		Humanities (N = 249)		Business (N = 88)	
	r	Rank	r	Rank	r	Rank	r	Rank	r	Rank
1	.75	7.5	.81	11.5	.81	6	.79	11	.79	6
2	.64	16	.81	11.5	.76	11	.75	14.5	.66	15
3	.60	17	.62	18	.64	17	.63	19	.49	19
4	.73	9	.71	17	.69	8.5	.74	16	.68	14
5	.85	1	.91	1	.84	3.5	.86	3.5	.89	1.5
6	.68	13.5	.76	13	.76	11	.83	6	.78	7.5
7	.82	3	.89	3	.84	3.5	.88	1	.84	5
8	.72	11	.83	9	.79	8.5	.82	7.5	.74	10
9	.82	3	.86	4	.81	6	.78	12	.85	4
10	.72	11	.84	6.5	.73	13	.82	7.5	.69	12
11	.82	3	.90	2	.87	1	.87	2	.86	3
12	.55	19	.46	20	.52	19	.56	20	.51	18
13	.68	13.5	.72	16	.79	8.5	.69	17	.69	12
14	.79	5	.84	6.5	.86	2	.86	3.5	.89	1.5
15	.42	20	.55	19	.48	20	.64	18	.45	20
16	.78	6	.84	6.5	.76	11	.84	5	.76	9
17	.67	15	.75	14	.71	14	.81	9	.69	12
18	.57	18	.73	15	.61	18	.75	14.5	.61	16
19	.75	7.5	.84	6.5	.67	16	.80	10	.59	17
20	.72	11	.82	10	.81	6	.77	13	.78	7.5

questions satisfactorily," 14, "Achieved the specified objectives of the course," and 9, "Gave several examples to explain complex topics."

The results in Table III also revealed a highly consistent pattern of correlations across the five disciplines. This implied that the students in the various disciplines tended to agree on the attributes that were indicative of effective teaching. In order to examine the interdiscipline similarity issue further, the disciplines were intercorrelated over the item correlations presented in Table III. The resulting R matrix is presented in Table IV. The correlations in Table IV further demonstrate the strong agreement among students taking courses in the various disciplines. Students taking courses in the five disciplines examined in this study did not differ materially in their opinions of what teacher attributes characterize effective teaching.

TABLE IV. The Pearson Correlations Among Disciplines Based on Correlations Between IIQ Item 21 and the Other IIQ Items Appearing in Table III (N = 20)

Discipline	2	3	4	5
1. Science and math	0.86*	0.88	0.79	0.87
2. Education		0.86	0.92	0.83
3. Social sciences			0.78	0.93
4. Humanities				0.83
5. Business				

*All correlations were significantly different from 0, α = .05, two-tailed.

DISCUSSION

The purposes of this study were to describe differences in instructor attributes in five disciplines, to identify those instructor characteristics which account for variation in general student rating of teaching effectiveness, and to determine if the teacher characteristics that contribute to high student ratings differ among five disciplines.

The results relating to the first purpose indicated that different teaching styles were exhibited by instructors in the five disciplines examined in this study. Science, humanities and business instructors received the highest ratings on making clear assignments and being prepared for class. Social science instructors obtained high ratings on knowledge of subject matter and giving several examples to explain complex topics. Education instructors received their highest ratings on specifying the objectives of the course and increasing students' appreciation of the subject matter.

The students who participated in this study rated highly those instructors who knew when students understood them, increased students' appreciation of the subject matter, answered impromptu questions satisfactorily, achieved the specified objectives of the course, and gave several examples to explain complex ideas. In general, students rated instructors favorably if they were perceived as effective in communicating subject matter to students and in a way that helped stimulate student interest in the material.

There were no substantial differences among students in the five disciplines in their perceptions of the importance of teacher attributes that characterized teaching effectiveness. The interdiscipline similarity coefficients reported in Table IV reveal consistent patterns in the importance attached to the various teacher attributes by students. Students, regardless of their discipline, tend to agree on the teacher attributes indicative of effective instruction.

These results suggested that there are teacher attributes which are consistently

perceived as important by students, regardless of their disciplines, while teaching styles differed among the disciplines. It was also apparent that those disciplines that received the highest ratings on the items students perceived as important also received the highest overall student ratings. Consequently students tended to reward, with good ratings, those teachers who exhibited the attributes students perceived as important determiners of effective instruction.

These results also have implications for universities who rely on student ratings to evaluate their instructors. This study found considerable differences among five disciplines on both the general level of ratings received (Table I) and the profiles of mean ratings (Table II). It would therefore behoove institutions to allow for discipline differences in the use and interpretation of student rating results. Allowances for discipline differences could be made by comparing faculty members only with their discipline peers or developing discipline-specific student rating instruments.

Finally, it should be stressed that lecture-discussion method courses contributed data for this study. Therefore the instructor attributes identified here may not be appropriate for describing effective instructors teaching in different settings such as self-instruction courses or lab courses.

Student ratings, like any assessment tool, can provide very meaningful information only if they are properly interpreted. This study indicates that a proper interpretation of student rating results must include allowances for expected discipline differences.

REFERENCES

Astin, A. (1971). "Predicting Academic Performance in College." New York: Free Press.

Costin, F. (1968). Survey of opinions about lecturers. University of Illinois, Department of Psychology. (Mimeograph.)

Costin, F., Greenough, W. T., and Menges, R. J. (1972). Student ratings of college teaching: reliability, validity and usefulness. Review of Educational Research 41:511–535.

Crawford, P. L., and Bradshaw, H. L. (1968). Perception of characteristics of effective university teachers: A scaling analyses. Educational Psychological Measurement 28:1079–1085.

Desphande, A. S., Webb, S. C., and Marks, E. (1970). Student perceptions of engineering instructor behaviors and their relationships to the evaluation of instructors and courses. American Educational Research Journal 7:289–305.

Downie, N. W. (1952). Student evaluation of faculty. Journal of Higher Education 23:495–496, 503.

Gadzella, B. M. (1968). College student views and ratings of an ideal professor. College and University 44:89–96.

Isaacson, R. L., McKeachie, W. J., and Milholland, J. E. (1963). Correlation of teacher personality variables and student ratings. Journal of Educational Psychology 54:110–117.

McKeachie, W. J., Lin, Y., and Mann, W. (1971). Student ratings of teacher effectiveness: validity studies. American Educational Research Journal 8:435–445.

Pohlmann, J. T. (1973). Evaluating instructional effectiveness with the instructional improvement questionnaire. Southern Illinois University, Testing Center. (Unpublished technical report.)

Turner, R. L. (1970). Good teaching and its contexts. Phi Delta Kappan 51:155–158.