
Students' Perceptions of Faculty and Graduate Students as Classroom Teachers

Harold Schuckman

Queens College of the City University of New York

Student ratings of the effectiveness of graduate student teaching versus faculty teaching were examined on seven occasions over an 8-year period. When all courses taught were considered, graduate student and faculty ratings were equivalent for six of the semesters in which student evaluations were conducted; on one occasion, the graduate students received significantly higher ratings than did the faculty. When only the ratings for introductory psychology courses were examined, results for graduate student teachers were significantly higher in four semesters; in three semesters, the ratings of the two groups were equivalent. There was a significant positive correlation of student evaluation ratings for individual faculty over time, but no such correlation for graduate students.

Graduate students teach a significant proportion of the lower division courses at many colleges and universities (McMillen, 1986), and it has been reported that 74% of doctoral programs in psychology assign full responsibility for a course to doctoral students (Lumsden, Grosslight, Loveland, & Williams, 1988). Whether these students are effective teachers is a question that is raised frequently. A report from the Association of American Colleges (1985, p. 45) stated, "As an initiation rite, the teaching assistantship is almost invariably a disaster." An earlier survey of psychology department chairs revealed that two thirds of those sampled did not believe that typical graduates of a conventional PhD

program are well prepared to assume their teaching responsibilities (Williams & Richman, 1971).

One may wonder on what evidence such opinions are based. Some years ago, Droyer (1973) and Miller (1974) reviewed studies that correlated student evaluations of teaching effectiveness with instructor age (years of teaching experience), and both concluded that there was no justification for any generalization about the relation between teaching effectiveness and age (experience). Feldman (1983) noted that 6 of the 12 studies examining the association between teaching effectiveness and age reported no correlation between these variables; however, the other 6 studies reported a negative correlation. When studies relating teaching effectiveness to years of teaching experience were reviewed, Feldman found a similar trend. In this case, 2 studies reported positive correlations, 5 reported negative correlations, and 6 indicated no significant relationship between the variables. Clearly, there is little empirical support for the belief that teaching improves with age or experience: if any relationship is suggested by empirical research, it is that the younger faculty are viewed by their students as better teachers.

Feldman (1983) also reviewed studies examining the relation between teaching effectiveness and academic rank. He found that 23 of the 35 studies indicated that students' over-

all evaluation of teaching effectiveness was unrelated to academic rank. A positive correlation between teaching effectiveness and academic rank was reported in 10 studies, but the correlations were low (.06 to .26). On the basis of his review, Feldman (1983) concluded that academic rank is unrelated to students' overall evaluation of instructor effectiveness. After considering these same studies, Marsh (1987) arrived at a similar, although more tempered, conclusion: There is almost no relation between rank and global ratings.

When attention is focused on the effectiveness of teaching assistants (TAs) and when their student evaluations are compared with those of the faculty, the results are less clear. Citing the large empirical studies of Centra and Creech (1976) and Brandenburg, Slinde, and Batista (1977), Marsh (1987) concluded that "teaching assistants typically receive lower ratings than other faculty" (p. 323). But this conclusion does not take into account such studies as those of McKeachie (1951) and Nevill, Ware, and Smith (1978), which found student ratings of TAs and faculty to be equivalent. Marsh's (1982) data also indicate the equivalence of TA and faculty student ratings. Although he suggested that courses taught by TAs are judged to be less effective than classes taught by faculty, the data are not convincing. In his study, the median student rating of undergraduate courses taught by faculty was 4.0 (on a scale ranging from 1 to 5); the median rating for courses taught by TAs was 3.8. No statistical techniques were used to demonstrate that the differences between the two groups of instructors were significant. Finally, Blount, Stallings, and Gupta (1978) presented data indicating that TAs received significantly higher student ratings than both assistant professors and full professors.

In summary, despite the generally accepted belief that undergraduate students rate TAs lower as classroom teachers, the empirical evidence does not consistently support that belief. I attempted to clarify this issue by examining student evaluation ratings of Psychology Department TAs and faculty at Queens College. The study was designed to take into account the possible confounding effects of the characteristics of the course being taught and to ascertain the stability of differences.

Method

Since 1978, Queens College has conducted seven student evaluations of its courses and faculty. The evaluation instrument consists of a questionnaire, which varies from semester to semester. One question that has appeared on every administration of the survey is: "What is your overall evaluation of the instructor?" Students rate the instructor on a 5-point scale ranging from *poor* (1) to *excellent* (5). Results of the evaluation are published the following semester. The published material lists each course for which evaluation data were collected, the name of the instructor, the number of students who completed the questionnaire, and the mean of the ratings for each question.

I identified and analyzed the undergraduate psychology courses taught by full-time faculty and TAs in each of the seven published evaluations. For each individual, the mean score on the question that assessed overall instructor effective-

Table 1. Student Ratings of Faculty and TAs for All Psychology Courses by Semester of Evaluation

Semester	Faculty			TAs			<i>t</i>
	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	
Spring 1978	3.38	.64	29	3.69	.63	23	1.72
Spring 1981	3.40	.84	26	3.53	.88	20	.50
Fall 1981	3.56	.66	32	3.41	.51	14	.75
Spring 1983	3.51	.65	27	3.93	.32	14	2.27*
Fall 1983	3.65	.71	22	3.55	.57	12	.42
Spring 1985	3.44	.83	16	3.76	.70	11	1.03
Fall 1986	3.70	.70	21	3.93	.55	14	1.05
Overall	3.52			3.69			

* $p < .05$.

tiveness was ascertained. For persons who taught more than one course in any semester, the mean of the mean scores on that question was calculated to yield a single representative score for that semester.

Results

Across the seven semesters, data were collected from courses taught by 41 full-time faculty and 68 TAs. The mean student evaluation ratings of faculty and TAs for all courses taught by these groups are shown in Table 1. Comparison of the two groups revealed no statistically significant differences in student ratings in six of the seven semesters. The single exception was the spring 1983 semester in which TAs received significantly higher ratings ($p < .05$) than did the faculty.

Table 2 presents the mean student ratings for the two groups when only the ratings of those teaching introductory psychology courses were tabulated. Analyses of these data revealed that student ratings of TAs exceeded those of the faculty in six of the seven semesters; on four of the six occasions, the differences were statistically significant.

Sixteen faculty members taught both introductory and more advanced psychology courses. In their introductory courses, the mean student evaluation rating for these faculty was 2.80; in their advanced courses, the mean evaluation rating was 3.26. This difference in mean ratings was statistically significant, $t(16) = 2.96$, $p < .01$.

Table 2. Student Ratings of Faculty and TAs for Introductory Psychology Courses by Semester of Evaluation

Semester	Faculty			TAs			<i>t</i>
	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	
Spring 1978	3.53	.76	7	3.12	.80	6	.95
Spring 1981	3.33	.44	8	3.46	.92	7	.27
Fall 1981	2.99	.80	8	3.74	.33	7	2.27*
Spring 1983	3.13	.68	6	3.84	.66	8	2.55*
Fall 1983	3.05	.85	6	3.65	.58	8	1.58*
Spring 1985	2.38	.67	5	3.84	.66	7	3.95**
Fall 1986	2.95	.65	6	3.67	.55	9	2.31*
Overall	3.05			3.62			

* $p < .05$. ** $p < .01$.

Estimates of the stability of ratings assigned to instructors were determined by computing Pearson r s between the earliest and latest ratings of faculty members and TAs for whom two such ratings were obtainable. The mean interval between the earliest and latest ratings was 5.5 and 1.9 years for faculty and TAs, respectively. The correlation between the earliest and latest ratings was statistically significant for the faculty, $r(33) = .60, p < .05$, but not for the TAs, $r(23) = .23, p < .05$. A Fisher r to z transformation indicated that the two stability coefficients were not statistically different, $z = 1.57, p < .05$.

Discussion

In general, TAs earned ratings equivalent to those received by full-time faculty, but when the two groups' ratings for only introductory courses were compared, the TAs received higher ratings. Several factors may account for these findings. First, in our department, TAs receive teaching assignments only after 2 to 3 years of residence in the doctoral program. (Before that time, their most frequent assignment is that of laboratory assistant.) Furthermore, not all resident doctoral students receive teaching assignments. Our TAs must be recommended by the faculty who are familiar with the students' intellectual and personal talents, and faculty are encouraged to recommend only those who are viewed as potentially excellent teachers.

Graduate students are appointed as TAs based only on their ability to teach. By contrast, teaching ability is one of many factors considered when full-time faculty are appointed, reappointed, and tenured. Scholarly productivity, clinical skill, and service to the institution are examples of prerequisites demanded of the faculty, but not of the graduate students. Consequently, it is not surprising to find that those who are selected exclusively on the basis of their ability to teach are rated as more effective teachers than are those who are not selected exclusively on that basis. Also, graduate students who perform poorly in the classroom are not reappointed.

The data also show that the greatest differences in student ratings between faculty and graduate students appear when their performances in the introductory course are compared. This is because the faculty ratings for teaching the introductory courses were significantly lower than their ratings for teaching upper division courses. Upper division and elective courses tend to receive higher ratings (Braskamp, 1980). Our data indicate that this occurs even when upper and lower division courses taught by the same instructor are being compared. Higher ratings for the advanced courses may reflect students' greater interest in their major courses or that some instructors put more effort into teaching courses in their areas of specialization.

The finding that TA ratings equal or exceed those of full-time faculty is similar to the findings reported by Blount et al. (1978), McKeachie (1951), and Nevill et al. (1978), but contrasts with those of Brandenburg et al. (1977) and Centra and Creech (1976). As noted previously, the latter studies were large-scale examinations of thousands of individual classes, and although such studies are more likely to produce generalizable results, they suffer a disadvantage in that a number of relevant variables may be masked or uncon-

trolled. For example, Centra and Creech (1976) examined student ratings in 10,000 classes derived from a highly mixed sample: community colleges, 4-year colleges, and universities. As Centra and Creech observed, teachers with teaching loads of 13 or more hours were rated higher than any other group, and these faculty are generally located at 2-year or 4-year colleges. Furthermore, although Centra and Creech reported that student ratings are affected by both method of instruction and subject area, they did not examine the possible confounding effects of these variables on TA-faculty comparisons. Perhaps the most significant influence not studied by Brandenburg et al. (1977) and Centra and Creech (1976) is whether TA ratings were derived from independently taught classes, and hence comparable to classes taught by faculty, or whether the TA ratings were obtained from laboratory and recitation sections. In the Brandenburg et al. (1977) study, which was conducted at the University of Illinois at Urbana-Champaign, 41% of 3,355 class sections for which student ratings were obtained were taught by TAs. A proportion of this magnitude suggests that many of the TA sections were either laboratory or recitation sections.

That ratings received by TAs are either equivalent or superior to those received by the faculty does not imply that teaching effectiveness of the two groups is equivalent, because several dimensions of teaching affect student ratings (Feldman, 1976). For example, students seem to agree that faculty of higher professorial rank have a greater knowledge of the subject matter (Downie, 1952; Goodhart, 1948; Stuit & Ebel, 1952), but they also regard instructors of lower ranks to be more tolerant and helpful (Downie, 1952; Stuit & Ebel, 1952). Perhaps factors such as these are weighted or averaged when students judge overall effectiveness.

Correlational data from my study demonstrate that individual faculty ratings are consistent over time. The TA ratings do not show this effect, but the stability coefficient for TAs is not significantly lower than that for the faculty. A high positive correlation of similar magnitude between ratings received by the same instructors in different offerings of the same courses over a 4-year period was reported by Marsh and Hocevar (1984). Thus, it appears that individual faculty ratings are stable over a considerable period of time. To be definitive on this issue would require evaluative measurements of faculty over the entire span of their teaching careers, but such data are not available.

Somewhat surprising, in view of the often-expressed concern over the quality of teaching in higher education institutions (e.g., Watkins, 1989), is the fact that the students, the recipients of that education, generally rate their instructors favorably. The average faculty rating for all courses taught by all instructors evaluated during the 8-year period was 3.6, in other words, between *good* and *very good*. Favorable student opinions of similar magnitude have been reported frequently (Centra & Creech, 1976; Marsh, 1982; McKeachie, 1951). Thus, despite the fact that "college teaching is probably the only profession in the world for which no specific training is required" (Milton & Shoben, 1968, p. xvii), college instructors are not viewed by their students as poor teachers. The data from my study suggest that TAs are rated no less favorably by those same students. One caveat remains: The data from this study may simply reflect the quality of Queens College TAs or may be a testament to the validity of the

department's selection procedures. Further studies will be required to establish if they are generalizable to other institutions.

References

- Association of American Colleges. (1985). *Integrity in the college curriculum: A report to the academic community. The findings and recommendations of the project on redefining the meaning and purposes of baccalaureate degrees* (Report No. ISBN-0-911696-18-0). Washington, DC: Author. (ERIC Document Reproduction Service No. ED 251 059)
- Blount, H. P., Stallings, W. M., & Gupta, V. G. (1978). The effects of different instructions in student ratings of university courses and teachers. *Journal of Educational Research*, 71, 149-152.
- Brandenburg, D. C., Slinde, J. A., & Batista, E. E. (1977). Student ratings of instruction: Validity and normative interpretations. *Research in Higher Education*, 7, 67-78.
- Braskamp, L. A. (1980). The role of evaluation in faculty development. *Studies in Higher Education*, 5, 45-54.
- Centra, J. A., & Creech, F. R. (1976). *The relationship between student, teacher, and course characteristics and student ratings of teacher effectiveness* (Report No. PR-76-1). Princeton, NJ: Educational Testing Service.
- Downie, N. W. (1952). Student evaluation of faculty. *Journal of Higher Education*, 23, 495-503.
- Droyer, F. M. (1973). Selected criteria for evaluating teaching effectiveness. *Improving College and University Teaching*, 21, 51-52.
- Feldman, K. A. (1976). The superior college teacher from the students' view. *Research in Higher Education*, 5, 243-288.
- Feldman, K. A. (1983). Seniority and experience of college teachers as related to evaluations they receive from students. *Research in Higher Education*, 18, 3-124.
- Goodhart, A. S. (1948). Student attitudes and opinions relating to teaching at Brooklyn College. *School and Society*, 68, 345-349.
- Lumsden, E. A., Grosslight, J. H., Loveland, E. H., & Williams, J. E. (1988). Preparation of graduate students as classroom teachers and supervisors in applied and research settings. *Teaching of Psychology*, 15, 5-9.
- Marsh, H. W. (1982). Validity of students' evaluation of college teaching: A multitrait-multimethod analysis. *Journal of Educational Psychology*, 74, 264-279.
- Marsh, H. W. (1987). Students' evaluations of university teaching: Research findings, methodological issues, and directions for future research. *International Journal of Educational Research*, 11, 253-388.
- Marsh, H. W., & Hocevar, D. J. (1984). The factorial invariance of student evaluations of college teaching. *American Educational Research Journal*, 21, 341-366.
- McKeachie, W. J. (1951). A program for training teachers of psychology. *American Psychologist*, 6, 119-121.
- McMillen, L. (1986, October 29). Teaching assistants get increased training. *The Chronicle of Higher Education*, pp. 9-10.
- Miller, R. I. (1974). *Developing programs for faculty evaluation*. San Francisco: Jossey-Bass.
- Milton, O., & Shoben, E. J. (1968). *Learning and the professors*. Athens: Ohio University Press.
- Nevill, D. T., Ware, W. B., & Smith, A. B. (1978). A comparison of student ratings of teaching assistants and faculty members. *American Educational Research Journal*, 15, 25-37.
- Stuit, D. B., & Ebel, R. L. (1952). Instructor rating at a large state university. *College and University*, 27, 247-254.
- Watkins, B. T. (1989, April 26). Colleges urged to avow that teaching is central to their mission and to take steps to upgrade it. *The Chronicle of Higher Education*, pp. A13-A16.
- Williams, J. E., & Richman, C. L. (1971). The graduate preparation of the college professor of psychology: A survey. *American Psychologist*, 26, 1000-1009.

Notes

1. I thank J. Orbach, P. Ramsey, and W. Winnick for their helpful comments.
2. Requests for reprints should be sent to Harold Schuckman, Department of Psychology, Queens College of the City University of New York, Flushing, NY 11367.