# FACULTY FORUM

## Factors Influencing Teaching Evaluations in Higher Education

Ginette C. Blackhart, B. Michelle Peruche,
C. Nathan DeWall, and Thomas E. Joiner, Jr.
*Florida State University*

*Past research indicates several factors influencing teaching evaluation ratings instructors receive. We analyzed teaching evaluations from psychology courses during fall and spring semesters of 2003–2004 to determine if class size, class level, instructor gender, number of publications (faculty instructors), average grade given by the instructor, and instructor rank predicted teaching evaluation ratings. Entering predictor variables into a multiple regression analysis concurrently, results indicated that only average grade given and instructor rank significantly predicted instructor ratings. Specifically, higher average grades given by the instructor predicted higher ratings, and graduate teaching assistants received higher overall ratings than faculty instructors.*

Teaching evaluations are widely used in assessing courses and the quality of teaching in higher education. In addition, university administrations often use ratings an instructor receives on these evaluations in personnel and merit decisions. Many educators have raised concern about this method of evaluation, however, as previous research has indicated there are several factors that affect the ratings students give instructors on teaching evaluations. Variables such as physical attractiveness of the instructor (e.g., O'Reilly, 1987), research productivity (Allen, 1996), class size (e.g., Hamilton, 1980; Haslett, 1976), gender of the instructor and student (e.g., Lueck, Endres, & Caplan, 1993), class level (e.g., Haslett, 1976), instructor rank (e.g., Harper, 1991; Schuckman, 1990) as well as expected grades (Goldberg & Callahan, 1991; Hamilton, 1980) or grades given by the instructor (Howard & Maxwell, 1980, 1982) have all significantly predicted teaching evaluation ratings.

Given the importance of teaching evaluations for instructors and their respective educational institutions, the fact that these variables substantially affect the ratings students give instructors on teaching evaluations should be of concern. The majority of these studies, however, have only examined the impact these factors have on teaching evaluations separately. Our study therefore evaluated how many of these variables, such as class size, class level (graduate vs. undergraduate course), instructor gender, average grade given by the instructor, instructor rank (graduate teaching assistant [TA], assistant professor, associate professor, and full professor), and number of publications (for faculty members only) concurrently affected ratings on teaching evaluations, specifically in psychology courses. The purpose of using this method was to allow these factors to control for one another to evaluate how each variable uniquely affected instructor rank, thus detangling unique predictors.

### Method

#### *Participants*

Students taking courses within the psychology department at Florida State University (FSU) completed the teaching evaluations ($N = 9{,}240$). Information on student characteristics, such as gender or age, was not gathered to maintain anonymity of the ratings.

#### *Materials*

The evaluation instrument consisted of a questionnaire (the State University System Student Assessment of Instruction) used to evaluate all courses and instructors at FSU. We used ratings on the item "overall assessment of instructor" as the dependent variable; students could rate instructors on this item as *excellent, very good, good, fair,* or *poor*.

#### *Procedures*

We collected data from teaching evaluations obtained from all courses (137 undergraduate courses, 30 graduate courses) offered by the psychology department during two consecutive semesters (Fall 2003 and Spring 2004) at FSU. Students completed teaching evaluations at the end of the semester while the instructor was out of the classroom. Faculty members taught 103 of the courses (19 by assistant professors, 40 by associate professors, and 44 by full professors), and graduate TAs taught 64 of the courses offered in the department.

#### *Statistical Analyses*

We first conducted correlational analyses between each variable. Next, we entered class size, class level, instructor gender, average grade given by the instructor, number of publications, and instructor rank into a multiple regression analysis (predictor variables were entered simultaneously to control for one another). We then calculated and entered the average rating on "overall assessment of instructor" as the dependent variable on a scale ranging from 5 (*excellent*) to 1 (*poor*).

### Results

Results indicated that there were significant correlations between average grade given and instructor rating, $r(143) = .30$, $p < .01$, and between instructor rank and instructor rat-

ing, $r(166) = -.22$, $p < .01$. Although the overall regression analysis was significant, $F(6, 134) = 3.51$, $p < .01$, class size, class level, instructor gender, and number of publications (for faculty only) did not significantly predict the overall assessment of the instructor. The average grade given by the instructor ($\beta = .30$, $p < .01$) and instructor rank ($\beta = -.22$, $p < .05$) did, however, significantly predict ratings of instructors (together accounting for 12% of the variance in instructor ratings). Higher average grades given predicted significantly higher ratings, and this effect was similar for both graduate TAs and faculty instructors. For instructor rank, lower rank predicted significantly higher ratings; graduate TAs consistently received better ratings on overall assessment of the instructor than faculty instructors. Among faculty instructors, however, instructor rank had almost no effect on evaluation ratings (see Table 1).

## Discussion

This study investigated how class level, class size, instructor gender, number of publications, average grade given to students, and instructor rank affected teaching evaluation ratings. Results indicated that average grade given by the instructor and instructor rank both significantly predicted instructor ratings on teaching evaluations given to students taking psychology courses at FSU. When evaluating average grade given by the instructor, the higher the average grade for the class, the better the ratings of the instructor. In addition, instructor rank significantly predicted instructor ratings, such that graduate TAs received higher ratings than faculty instructors. Faculty rank (full, associate, or assistant professor), however, did not have a significant effect on instructor ratings.

Although past research has shown that average grade given by the instructor affects teaching evaluations (Howard & Maxwell, 1980, 1982), it remains unclear why. Many educators have concluded that the more lenient an instructor is in grading, the better ratings that instructor will receive. In fact, a number of educators believe student evaluations of their instructors have contributed significantly to grade inflation (Hamilton, 1980). Higher grades given by an instructor, however, may indicate that the instructor was effective in teaching the material to his or her students, thus resulting in higher grades and higher ratings. If students are doing well and getting good grades in the class, they may attribute their performance, in part, to the quality of the instructor. In addition, those students who do well in a class might be more engaged in the course topic and the material being learned, leading to better evaluations of the instructor. Thus, although student evaluations of instructors may well lead to grade inflation, providing instructors assurance of better ratings on their evaluations, there are additional reasons, some directly relating to teaching effectiveness, that could also explain why the grades assigned to students significantly affect teaching evaluations.

**Table 1. Means and Standard Deviations of Teaching Evaluation Ratings for Faculty Instructors and Graduate TAs on the Question "Overall Assessment of Instructor"**

| | *M* | *SD* |
|---|---|---|
| Graduate TA | 4.48 | 0.43 |
| Assistant professor | 4.26 | 0.94 |
| Associate professor | 4.13 | 0.51 |
| Full professor | 4.18 | 0.76 |

*Note.* For average rating ratings were based on a scale ranging from 5 (*excellent*), 4 (*very good*), 3 (*good*), 2 (*fair*), to 1 (*poor*). TA = teaching assistant.

This study also found that graduate TAs received better evaluations than faculty instructors. Graduate TAs receiving higher evaluations is not attributable to the level of the courses they taught, as graduate TAs taught introductory lecture courses and labs as well as upper level courses within the major. There are, however, several other possible reasons for these findings. Some educators have suggested that the more productive professors are in research, the more their teaching suffers (see Allen, 1996). Our study found, however, that the number of publications faculty members had did not significantly predict ratings on teaching evaluations, suggesting that productivity does not affect teaching evaluation ratings. In addition, Allen found that research productivity correlated positively with teaching effectiveness.

Another reason graduate TAs may have received better teaching evaluations is due to the enthusiasm, effort, and energy they often bring to teaching. Teaching college-level courses is often a novel experience for graduate TAs, possibly leading to greater enthusiasm about teaching as well as greater effort and energy being expended toward the task. In addition, graduate TAs may receive better ratings from their students due to the fact that they typically are closer in age to the students, with perceived similarity being greater between graduate TAs and students than between professors and students. Harper (1991) also suggested that graduate TAs might better express positive attitudes toward their students than faculty members.

Finally, it should be noted that within the psychology department at FSU, all graduate TAs receive training before teaching. Graduate students learn the fundamentals of teaching and gain hands-on experience by giving guest lectures, preparing course materials in advance, and having these materials evaluated and approved by faculty. They receive additional supervision by faculty when teaching, and the department encourages graduate TAs to attend teaching seminars and conferences offered by the university. Thus, it may be the training graduate TAs received that led to higher teaching evaluation ratings.

In conclusion, our study suggests that, within psychology courses, the average grade given by the instructor and instructor rank are among those variables significantly affecting teaching evaluations when controlling for class size, class level, instructor gender, and number of publications (for faculty instructors). Although several studies have reported class size, class level, and instructor gender as significant predictors of instructor ratings, our study found that these variables failed to significantly predict instructor ratings when we concurrently entered our predictor variables into a regression analysis. Taking each of these variables into account when predicting teaching evaluation ratings, only average grade given and instructor rank emerged as significant predictors of instructor ratings. Although the average grade given by the instructor

and instructor rank together accounted for only 12% of the variance in teaching evaluation ratings, these results suggest that teaching evaluation ratings are significantly affected by factors that may be unrelated to teaching effectiveness (e.g., average grade given or instructor rank). University administration should therefore take these factors into consideration when using ratings from teaching evaluations to assess an instructor's teaching effectiveness and overall performance.

## References


Allen, M. (1996). Research productivity and positive teaching evaluations: Examining the relationship using meta-analysis. *Journal of the Association for Communication Administration, 2,* 77–96.

Goldberg, G., & Callahan, J. (1991). Objectivity of student evaluations of instructors. *Journal of Education for Business, 66,* 377–378.

Hamilton, L. C. (1980). Grades, class size, and faculty status predict teaching evaluations. *Teaching Sociology, 8,* 47–62.

Harper, B. (1991). Who does the better job teaching the basic course: Teaching assistants or faculty? *ACA Bulletin, 77,* 68–70.

Haslett, B. J. (1976). Student knowledgeability, student sex, class size, and class level: Their interactions and influences on student ratings of instruction. *Research in Higher Education, 5,* 39–65.

Howard, G. S., & Maxwell, S. E. (1980). The correlation between grades and student satisfaction: A case of mistaken causation? *Journal of Educational Psychology, 72,* 810–820.

Howard, G. S., & Maxwell, S. E. (1982). Do grades contaminate student evaluations of instruction? *Research in Higher Education, 16,* 175–188.

Lueck, T. L., Endres, K. L., & Caplan, R. E. (1993). The interaction effects of gender on teaching evaluations. *Journalism Educator, 48,* 46–54.

O'Reilly, M. T. (1987). Relationship of physical attractiveness to students' ratings of teaching effectiveness. *Journal of Dental Education, 51,* 600–602.

Schuckman, H. (1990). Students' perceptions of faculty and graduate students as classroom teachers. *Teaching of Psychology, 17,* 162–165.


## Note


Send correspondence to Ginette C. Blackhart, Department of Psychology, Florida State University, One University Way, Tallahassee, FL 32306; e-mail: blackhart@psy.fsu.edu.


# Effect of Crib Card Construction and Use on Exam Performance


K. Laurie Dickson and Michelle D. Miller
*Northern Arizona University*



*Both normal classroom use and research typically confound crib card construction with crib card use, making it unclear whether students benefit from the process of creating crib cards. We compared the effect of self-constructed and other-constructed crib cards (written by a student research assistant) on undergraduates' (N = 32) multiple-choice exam performance. Performance was better with other-constructed cards than with self-constructed cards. Crib card construction did not facilitate student learning, nor did the use of self-constructed crib cards enhance exam performance.*


College instructors occasionally allow students to use a "cheat sheet" or crib card during an exam. Some educators (Dorsel & Cundiff, 1979; Hindman, 1980; Trigwell, 1987) have argued that the process of constructing a crib card may help students learn the material, whereas others (Dickson & Miller, in press; Whitley, 1996; Whitley & Keith-Spiegel, 2002) have argued that crib cards are disadvantageous for students.

According to the coding hypothesis, the active processes of selecting, organizing, and representing information during crib card construction benefit students (Dorsel & Cundiff, 1979; Hindman, 1980; Trigwell, 1987). In contrast, Whitley (1996; Whitley & Keith-Spiegel, 2002) argued that crib card construction encourages students to study less, as they assume that they do not really need to know or understand the information. Trigwell (1987) stated that crib cards are not helpful because they merely reflect students' knowledge at the time they construct the card and thus do not enhance student exam performance. Dickson and Miller (in press) found that most students reported that the amount of time they spent studying did not increase when they constructed a crib card compared to exam preparation without a crib card. Thus, the time spent constructing a crib card may have diverted students from more productive study activities. Some students reported that they did not learn the course material as well when they had a crib card, which suggests that the process of constructing a crib card interfered with learning. In contrast, a minority of students thought that the process of constructing a crib card was a useful study strategy. Crib card construction and use are typically confounded because they are not examined as separate processes. Thus, it is unclear whether creating a crib card is beneficial for student learning and exam performance.

The purpose of this study was to separate the effect of crib card construction from the effect of crib card use on multiple-choice exam performance. Students in an upper division psychology course used self-constructed crib cards for two of the four exams. For the other two exams, they unexpectedly used a crib card constructed by another student.

## Method

### *Participants and Procedures*

Thirty-two child and adolescent development students (82.4% women; 14.7% sophomores, 35.3% juniors, 50.0% seniors; 73.5% psychology majors) constructed and used crib cards (one side of a 5 × 8 in. [12.5 × 20 cm] index card) for Exams I and III (self-constructed crib card). The instructor directed students to draw on the course learning objectives and the textbook when constructing the cards. For Exams II and IV, students were not aware that they would be able to use a crib card. As the instructor handed out these exams, she also distributed premade crib cards to students for use during the exams. A student research assistant developed crib cards based on the learning objectives and textbook (other-constructed crib card). The research assistant had neither taken the course nor seen the exams prior to developing the crib cards. The first author reviewed the other-constructed crib card for accuracy and found no errors.