AC 2008-1915: THE RELATIONSHIP OF INSTRUCTOR RATINGS WITH TA RATINGS IN HIGH ENROLLMENT, LECTURE/LAB COURSES: A PRELIMINARY STUDY

Jon Sticklen, Michigan State University

Mark Urban-Lurain, Michigan State University

THE RELATIONSHIP OF INSTRUCTOR RATINGS WITH TA RATINGS IN HIGH ENROLLMENT, LECTURE/LAB COURSES: A PRELIMINARY STUDY

Introduction

Sound evaluation of faculty performance as class instructors is multifaceted, and typically is framed along dimensions such as student ratings, student achievement of objective learning outcomes, enhancement of course methodology or content, and visitation by faculty colleagues. However, at many institutions, student ratings are heavily weighted in faculty evaluation. (Abrami, 2001a, 2001b). Student ratings of instructors have been extensively researched, yet remain controversial. While many faculty feel that student ratings are not reliable, the research consistently shows that they are (Abrami, 2001a; Cashin, 1995; Marsh & Dunkin, 1992). As Cashin (1995) points out, much of the controversy around student "ratings" comes from their misuse as "evaluation." That is, while the research suggests that students' ratings of instructors may be among the best data about the instructor's effectiveness, they should be treated as one *data source*, rather than as the final evaluation of the instructor.

In high enrollment courses that have both lecture and laboratory components, student ratings of the lead faculty instructor may have two components: student attitudes about the course based on perceptions of the lead faculty person, and student attitudes about the course based on perceptions about the student's teaching assistant (TA). It could be conjectured that these two sources of attitudes about the course merge in the perception of a student, and that an important factor in the rating a student gives to a faculty person is the rating the student gives to his TA or vice versa. Certainly, anecdotal evidence is available that if a student is unhappy with a TA, the same student may be unhappy with the course in general, and with the faculty person in charge of the course. Even though the literature on student ratings is extensive, the subtopic of the interaction of TA ratings and faculty ratings is not; indeed there is not specific literature on this topic that we could identify.

In this paper, we explore the relationship between student ratings of the instructor and TAs in a large enrollment lecture/lab course to try to understand what, if any, impact there of a student's TA on student ratings of instructors.

Literature on Student Evaluation of Teaching

Cashin (1995) notes that there were over 1500 references on research of student evaluations of teaching in 1995. That number has grown, so that a recent search in ERIC on the thesaurus term *Student Evaluation of Teacher Performance* results in 2647 references. Narrowing that search to include the phrase "teaching assistant" reduces the literature to 30 references. These include such characteristics: as the TA's GRE scores (Vecchio & Costin, 1977); the TA's affective communication skills (O'Hair & Babich, 1981); the TAs dress and student classroom

behavior (Roach, 1997); TA gender, age and other demographics (Bos & et al., 1980; Rohrscheib, 1993); TA training (Shannon, Twale, & Moore, 1998; Williams, 1991); and the impact of non-native English speaking TAs (Dalle & Inglis, 1989). However, within that literature, there is no research examining the relationship between student ratings of faculty and the teaching assistants within a course. Generally, the research examining ratings of teaching assistants focuses on the attributes of the TA.

Our ultimate goal is (a) to determine if there is a relationship between student ratings of faculty and student ratings of TAs and (b) if a relationship is found, then to understand the factors contributing to the relationship. We could locate no published research on this topic.

Research Setting and Procedure

CSE 131 is an introduction to the use of computing systems for technical problem solving in engineering and science course required by most engineering majors in Michigan State University. Over 85% of the student enrollment in the course consists of freshman and sophomores. Total ending enrollment in the course for Fall, 2007, was 207 students. For Fall, 2007, there were two lecture sections of the course conducted by the same instructor. The two lecture sections were operated as closely as possible to be identical to each other. The course had 23 lab sections that were met by fourteen teaching assistants (TAs.) The course had two types of TAs: graduate students in computer science and undergraduate, upper division engineering students who taken the course and had done well. Graduate student TAs were each responsible for three lab sections of the course; undergraduate TAs were each responsible for one lab section.

The end-of-semester ratings by students were done in the standard manner for our university. One paper form was used for student ratings of TAs. Students identified their lab section on the form. The form consisted of five specific questions (coded on a 5-point Likert scale from Strongly Agree to Strongly Disagree) and one general rating question:

1. The teaching assistant was available and willing to help the student.

2. The teaching assistant was prepared for class sessions and enthusiastic about teaching the course section.

3. The teaching assistant organized and explained the materials for this section well and generally displayed a high level of competency in the subject matter of the course.

4. The teaching assistant communicated in both written and oral models, well and with ease.

5. The teaching assistant was fair in the grading of assignments and tests.

6. Rate the teaching assistant on a scale of 4 -- 3 -- 2 -- 1 -- 0 (four is best).

A second form was used for student ratings of the course faculty. Students indicated the lab section they attended on the form. The faculty rating form had four specific questions (coded on

a 5-point Likert scale from Strongly Agree to Strongly Disagree) and one general rating question as follows:

1. The instructor was available and willing to help the student.

2. The instructor explained course material clearly.

3. The instructor was well prepared for classes and other related course activities.

4. The instructor organized the course well.

5. Rate the instructor on a scale of 4 -- 3 -- 2 -- 1 -- 0 (four is best).

The questions on both forms address the standard dimensions of student ratings forms as derived from research on student ratings of instructors (Cashin, 1995). Both forms also had space after each question for students to elaborate on their response. The student open responses were not considered in this study.

The TA rating forms were administered on the last class day in lecture. Care was taken that the TAs did not handle the forms to emphasize to the students that their TA would not see the ratings until after the end of the Fall Term. The faculty rating form was administered with the final examination. Neither the faculty nor the TAs handled the completed forms; a student volunteer returned the forms to the administrative offices following standard University procedures. This was to emphasize to students that the faculty rating form would not be seen by faculty until the term was over.

Data Aggregation and Analysis

Following the submission of all grades for the course to the Registrar's Office, the responses from each student were transcribed into a spreadsheet program. The student record for each TA survey turned in consisted of the student's lab section, the student's responses to each of the five specific questions on the TA rating form (coded as indicated above), and the student's response for the TA summary question. A similar student record for each faculty survey turned in consisted of the student's lab section, the student's response to each of the four specific questions on the faculty rating form (coded as indicated above), and the student's response for the faculty summary question. Any form that did not include the student's lab section was discarded from this study because without the lab section identification including the data in an aggregate lab section is not feasible. Likewise, any student TA rating that included a response of "not applicable" on any of the specific questions was discarded from this study.

Because the rating forms do not have student identifiers, we cannot link one student's ratings of his/her TA with the same student's ratings of his/her faculty instructor. That severely limits the types of analyses and conclusions we are able to make, as noted below. But because we do know the lab section numbers for all submissions - both for TA ratings and for faculty ratings- we can aggregate the results across sections, and then compare the aggregated values.

Results

Our dataset contained individual student ratings of the instructor and TA organized by TA and by lecture day; recall that we cannot connect any individual students' ratings of the instructor or TA, so we must aggregate the data by lab section. Since each lab section was associated with one of the two lecture periods, we can analyze the data for each lecture section by collapsing the lab sections based on which day that lab's section met for lecture. This makes it possible to compute statistics at the lab section level. Since each lab is conducted by a TA, but some TAs teach more than one section, we combined the data for all sections taught by each TA to examine the results by TA.

Each lab section has a maximum enrollment of 11 students. Therefore, it is very likely that the data are not normally distributed. For each variable, we used the Kolmogorov-Smirnov test to determine if the data were normally distributed by TA. The data were not normally distributed, so we computed ranks for the data and performed all subsequent statistical comparisons on the ranked variables.

Results by Lecture Sections

We compared the ratings of both the instructor and TAs by lecture section. There were two lecture sections; a Thursday lecture section and a Friday lecture section, both taught by the same instructor. First, we compared the (ranked) mean ratings of the instructor across sections (Table 1)

Table 1

ANOVA for Ranked Instructor Ratings: Compare Thursday and Friday Lectures

Question	df	F	η^2	p
1. The instructor was available and willing to help the student.	1, 187	2.112	.011	.148
2. The instructor explained course material clearly.	1, 187	1.698	.009	.194
3. The instructor was well prepared for classes and other related course activities.	1, 187	13.970**	.070	.000
4. The instructor organized the course well.	1, 185	.045	.000	.833
5. Rate the instructor on a scale of 4 - 3 - 2 - 1 - 0	1, 169	1.299	.008	.256

** $p < .01$

The only scale on which there were significant differences was scale 3: *Instructor was well prepared for class and other course activities.* Here, the Friday section rating was higher (Friday

mean = 3.98; Thursday mean = 3.54, p < .001, η^2 = .07). It is possible that the instructor, having already conducted the class on Thursday had practiced the class and was better prepared on Friday. Another possibility is differing demographics between students in the Thursday lecture section (taught at 7 p.m.) and the Friday morning lecture section (taught at 10 a.m.). However, class day accounts for only 7% of the variance on this dimension of the student ratings on this item (η^2 = .07).

Next we compared the student ratings of their instructor based on the TA the students had. That is, for each TA's students, we compared the ratings of the instructor to see if there was a difference in student rating of the instructor based on which TA the students had (Table 2.)

Table 2

ANOVA for Ranked Instructor Ratings: Compare Instructor Ratings Across TAs

Question	df	F	η^2	p
1. The instructor was available and willing to help the student.	14, 174	1.252	.091	.243
2. The instructor explained course material clearly.	14, 174	1.443	.104	.138
3. The instructor was well prepared for classes and other related course activities.	14, 174	1.808[*]	.127	.041
4. The instructor organized the course well.	14, 172	.960	.072	.497
5. Rate the instructor on a scale of 4 - 3 - 2 - 1 - 0	14, 156	1.342	.107	.189

* p < .05

Again, the only scale on which there was a difference was scale 3: *Instructor was well prepared for class and other course activities.* Here, the ANOVA across 14 TAs was significant (p < .05) accounting for 13% of the variance on this scale (η^2 = .127)

Results by TA

For this analysis, we compared ratings by TA using ANOVA on the ranked ratings for the TAs on each scale. The results are shown in Table 3. There were significant differences among the TAs on every scale (p < . 001) The amount of variance accounted for by TA on each scale is shown in the η^2 column. Generally, about 1/3 of the variance in rating of the TA is accounted for by which TA the student had. This is an understandable result since some TAs are more experienced and have more skill teaching than others.

Table 3

ANOVA Comparing Ratings of TAs

Question	Mean Rating for all TAs	Range of Mean Ratings (Min / Max)	df	F	η^2	p
1. TA available and willing to help students.	4.24	1.57/2.44	13, 166	6.720[**]	.345	.000
2. TA was prepared for class and enthusiastic about teaching.	3.91	2.43/4.43	13, 166	8.031[**]	.386	.000
3. TA organized and explained the work of the session well and generally displayed a high level of competence in the course material.	4.01	2.43 /4.76	13, 165	6.183[**]	.328	.000
4. TA communicated both in oral and written modes well and easily.	4.01	2.57/4.57	13, 165	4.833[**]	.276	.000
5. TA was fair in grading assignments and tests.	4.02	3.43/4.63	13, 166	2.922[**]	.186	.001
6. Rate the teaching assistant on a scale of 4 -- 3 -- 2 -- 1 -- 0	3.30	2.17/4.00	13, 150	6.496[**]	.345	.000

** $p < .01$

Discussion and Future Research

Many STEM classes have large lecture sections taught by faculty, with several small lab sections met by a range of TAs. While there is a substantial literature on student ratings of instructor and TAs and the factors contributing to their ratings, we found no research on the relationships among ratings of faculty and TAs in these settings. In this study, we compared the student ratings of the instructor and TAs in a large, introductory computing class for engineering students at Michigan State University to uncover any relationship among the ratings of the instructor and the ratings of TAs.

In this course, we found that the student ratings of the instructor were stable, regardless of the TA that a student had. In contrast, students' ratings of their TAs was highly variable,

regardless of which lecture section they attended. This suggests that the students' ratings of the instructor and TA were independent of each other. That is, students of TAs who were rated lower did not tend to rate the instructor any lower (or higher) than students of TAs who were rated highly.

It is important that the results of our preliminary study be correctly understood. We cannot make statements based on our research reported here that speak to the actual student-level correlation between TA ratings and faculty ratings. That, of course, is what we ultimately want. In this preliminary study our statements must be couched in terms of the section-aggregated data, and the relationships among entire sections of students. Correlation between section-aggregated data for faculty evaluation versus section aggregated data for TA evaluation, although interesting in passing, cannot be used to assert student level correlation.

We are now in the process of applying for IRB approval to undertake a study in late Spring, 2008, that will allow us to collect student data containing student ratings of TAs and faculty, final term grade, and student demographics (gender, ethnicity, ACT score, GPA since matriculation). After approval is granted, we will be in position to gather linked data, and to analyze that data at the student level to better understand the nature of the relationship among student ratings of the instructor and TAs in large lecture/lab courses.

Citations

Abrami, P. C. (2001a). Improving judgments about teaching effectiveness using teacher rating forms. *New Directions for Institutional Research, 2001*(109), 59-87.

Abrami, P. C. (2001b). Improving judgments about teaching effectiveness: How to lie without statistics. *New Directions for Institutional Research, 2001*(109), 97-102.

Bos, R. R., & et al. (1980). Teaching Assistant Traits: Their Influence on Student Ratings. *Improving College and University Teaching, 28*(4), 179-185.

Cashin, W. E. (1995). *Student ratings of teaching: The research revisited* (No. 32): Center for Faculty Evaluation & Development Kansas State University.

Dalle, T. S., & Inglis, M. J. (1989). *What Really Affects Undergraduates' Evaluations of Nonnative Teaching Assistant's Teaching?*

Marsh, H. W., & Dunkin, M. J. (1992). Students' evaluations of university teaching: A multidimensional perspective. In J. C. Smart (Ed.), *Higher education: Handbook of theory and research* (Vol. VIII, pp. 143-233). New York: Agathon Press.

O'Hair, H. D., & Babich, R. M. (1981). *The Evaluation and Prediction of Affective Response to Graduate Teaching Assistants' Classroom Communication.*

Roach, K. D. (1997). Effects of Graduate Teaching Assistant Attire on Student Learning, Misbehaviors, and Ratings of Instruction. *Communication Quarterly, 45*(3), 125-141.

Rohrscheib, L. (1993). *Vestals or Victims: Suggestions for the Female Teaching Assistant.*

Shannon, D. M., Twale, D. J., & Moore, M. S. (1998). TA Teaching Effectiveness: The Impact of Training and Teaching Experience. *Journal of Higher Education, 69*(4), 440-466.

Vecchio, R., & Costin, F. (1977). Predicting Teacher Effectiveness from Graduate Admissions Predictors.

Williams, L. S. (1991). The Effects of a Comprehensive Teaching Assistant Training Program on Teaching Anxiety and Effectiveness. *Research in Higher Education, 32*(5), 585-598.