Student Perceptions of the Faculty Course Evaluation Process: An Exploratory Study of Gender and Class Differences

Peter Heine
Stetson University

Nick Maddox
Stetson University

ABSTRACT

Student evaluation of teaching (SET) has long been the subject of research, primarily focusing on two areas. The first area addresses the accuracy of students' perceptions regarding their teachers' performance in class. Secondarily, research has focused upon uncovering the source of students' perceptions about teaching effectiveness and quality. Most such studies utilize actual SET data sets generated by student course evaluations. A variety of variables such as class size, gender, and expected grade are used in statistical analyses. This paper, however, describes a study, not of SET data itself, but of student perceptions of the entire faculty class evaluation process. After a series of student focus group discussions, a 16-item Faculty and Course Evaluation Questionnaire was developed. The survey results indicate a number of significant statistical differences in responses related to gender and class ranking, as well as other process issues. For gender differences, female students were found to take the evaluation process more seriously than their male counterparts. Additionally, female students reported believing the evaluation process was more important than males in the sample. Male students, indicating some cynicism about the class evaluation process, were significantly different in a negative way from female students in terms of their perception that the higher the grade projected the higher their evaluation of a professor and their belief that professions adjusted their in-class behavior at the end of the semester to achieve higher evaluations. Discussion of these and other results are provided as well directions for future research.

Keywords: Student evaluation of teaching, student perceptions of class evaluation processes

INTRODUCTION

In this paper, the authors report on an exploratory study of the perceptions that a sample of business students hold regarding the faculty and class evaluation process within a small, private, southern comprehensive university. The School of Business Administration is accredited by AACSB at the graduate and undergraduate levels.

While most of the research on student evaluations of teaching (SET) focuses on various student characteristics and contextual factors that may affect the legitimacy, validity and reliability of SET processes, this research asked students to share their views on the overall process using a 16-item assessment tool. Thus it represents one of the few recent attempts to ascertain students' perceptions of the evaluation process. In addition to considering the aggregate perspectives of students regarding the process, analysis was also performed to determine any existing gender or year-in-school (class) differences within the sample. For this paper, the authors will describe the origins of the assessment tool, the methodology and results of the study and discuss the implications of results. Attention will also be given to future research in this area of great personal and professional significance to educators.

OVERVIEW OF THE LITERATURE

Helterbran (2008) suggests that teaching is a complex process that "involves the interweaving of content knowledge, pedagogy skills and a knowledge and appreciation of the multi-faceted nature of students to, in the end, be able to point to evidence that learning has occurred (p. 126). However, in higher education these days faculty teaching effectiveness is often evaluated through SET processes that have been routinely criticized for being open to many sources of bias and error. Wachtel (1998) is among those who additionally question whether students have the capacity to actually evaluate teaching and teaching effectiveness.

Conversations among university colleagues on the topic of student course and faculty evaluations are typically animated and full of opinions, myths, war stories and frustrations. Almost every business discipline forms professional associations, holds conferences and publishes journals focused only on teaching students in their discipline, not on organizational research within the discipline.

Predictably many articles appear in such journals that focus, not on how to more effectively teach the discipline, but on how faculty teaching is evaluated. Researchers frequently point out that because SET ratings are often used to both establish teaching competence and as a component of overall faculty evaluation, the origins of the ratings and the influences on the ratings are critical to consider. Moore (2008) notes many of these controversies in a recent work that also addresses the perceptions students hold regarding SET processes.

Published studies of course and faculty evaluation by students generally fall into two separate but related areas. The first area addresses the accuracy of perceptions by students on faculty performance, while the second research focuses on the sources of students' perceptions about teaching effectiveness. The first area, accuracy of perceptions, often involves grading leniency as related positively to student evaluation, a commonly held perception among faculty. Cognitive dissonance theory, for example, suggests that students who expect poor grades rate instructor poorly to minimize psychological or ego threat. One study (Maurer, 2006) found support for cognitive dissonance as a significant variable affecting accuracy of student perceptions.

Another such study (Heckert, Latier, Ringwald-Burton & Drazen, 2006), however, found that student perceived "effort appropriateness" was more positively related to faculty evaluation than was simple expected grade. That is, students who extended effort, learned more, and were subsequently rewarded, rated instructors more highly than simply expected grade could explain. The role of the scale used and question sequencing used on the rating form was another approach studied (Sedimeier, 2006) That author concluded both of these had significant effects on the accuracy or reliability of student evaluations.

The second, and perhaps more common, research approach in analyzing student evaluation processes is to assess the source of the student perceptions about teaching effectiveness. Such studies often focus on the students' demographics as well as the characteristics of the course or delivery method. Females were found to have more effective "evaluation abilities" than males and evaluated some aspects of courses more favorably than males, particularly in open-ended formats. Females, for example, had more accurate recall (Darby, 2006a) in articulating course events. Another study by the same author found that elective courses were more highly rated than required courses (Darby, 2006b). In a complex longitudinal study, McPherson (2006) found that many variables such as student class level, and even time of day, related significantly to student evaluation scores. Using the same course and instructor teaching one section of a traditional course format and a section of a two-way interactive television delivery, results suggested that the traditional course format was more highly rated (Mintu-Wimsatt, et al., 2006).

A third, and more unique approach, is presented in this paper and involves the study of students' perceptions of the faculty and class evaluation process. There is relative scant research in this area although logically it is an important orientation to consider regarding SET. Moore (2008), using eight items to assess student perceptions, completed a study of student perceptions of teaching evaluations and found while students perceived SETs as effective measures of teaching effectiveness, they exhibited doubt about whether either faculty or students take the process seriously. The students also were indifferent about whether the SET process was a useful process. Another study addressed some concern about how students perceived the evaluative process. It suggested that faculty and students should partner in evaluating courses and pedagogy to overcome erroneous perceptions, such as why faculty emphasized certain topics. The study found that course effectiveness was improved through such partnering (Giles, et al., 2006).

In another study of student perceptions of SET processes, Sojka, Gupta and Deeter-Schmelz (2002) compared student perceptions with faculty perceptions regarding students' evaluations (3 items), effects of student evaluations (6 items) and seriousness of the SET process (2 items). Results from that study indicated that on seven of the 11 items there were significant differences between student and faculty SET perceptions. In each of those comparisons, faculty rated the particular item as more positive than did students.

The intention of this study is to extend consideration of student perceptions of SET processes and to look at gender and year-in-school differences that may exist in the perceptual schema that students collectively retain about faculty and class evaluation processes.

METHODOLOGY

The original idea for this research evolved from the authors' concerns about the faculty/class evaluation and its legitimacy as related to performance evaluation of faculty at their university. As part of a project-focused undergraduate human resource management course,

students were assigned to a focus group to discuss faculty and class evaluation issues and to identify perceptions that business students, in general, held about the SET processes within the School of Business Administration.

The student focus group met four times to dialogue about the mandatory faculty/course evaluation process. Focus group dialogue was facilitated by the lead author and also included discussion of the standardized SET forms that were and remain in use. Additional discussion focused on semester timing of the evaluations, use of the SETs by faculty and administrators, student candor and comfort level with the process, and faculty behavior surrounding the SET process. At the time of the focus group, the authors had not discovered related work done by the researchers noted above. Hence, the generation of potential items for the FCEQ (Faculty and Course Evaluation Questionnaire) was independent of and not informed by any other sourcing than the aggregate perceptions of students participating in the focus groups. Final item selection and composition of the FCEQ was completed by the lead author of this article based on the project work of the student focus group.

The Faculty and Course Evaluation Questionnaire (FCEQ) included 15 perceptual items and one item addressing students' overall perceptions on the effectiveness of the SET process. The instrument was divided into three sections addressing: 1) Students Responses about Themselves; 2) Student Responses about Professors; and 3) Student Responses about the Evaluation Process. Item 16 was an omnibus "effectiveness" item which served as the dependent variable in the regression analysis of the other 15 items included in the FCEQ. All 16 items were scored on a 1 – 4 scaling from Strongly Disagree (1) to Strongly Agree (4). Demographic data was collected on gender, school (Business versus Arts and Sciences) and class ranking (freshman through senior). The FCEQ items are identified in Table 1 below. Table 2 presents a summary of sampling characteristics.

TABLE 1

Faculty and Course Evaluation Questionnaire

Directions: Please complete all of the questions below as truthfully as possible and fill in the information at the bottom of the page. Use the following scaling key for your responses

Strongly Disagree/Never **
Disagree/Seldom
Agree/Sometimes
Strongly Agree/Often

Student Responses about Themselves

1) I take evaluating the professors in my courses seriously.
2) I tend to evaluate female professors higher than male professors.
3) I feel comfortable giving a negative evaluation for a bad professor.
4) I rate professors based on their personality and enthusiasm and not on what I have learned.
5) The higher the grade that I expect to receive in a class, the more positive my evaluation.
6) I don't write many comments on the evaluation form for fear of being identified.
7) Overall, I think the professor and course evaluation process is important.

Student Responses about Professors

8) Professors take my evaluation comments seriously.
9) My evaluations are used in professor tenure and salary raise decisions.

10) Professors use their evaluations to improve their courses.
11) When students give low evaluations, professors adjust to improve their teaching.
12) Professors adjust their behavior at the end of the semester to get better evaluations.

Student Responses about the Evaluation Process
13) Completing the evaluation form in the beginning of a class is better than later in the class.
14) The questions asked on the form are clear to me.
15) The questions asked on the form are relevant to evaluating a course/professor.
16) Overall, I think the professor and course evaluation process is effective,

School _____

Gender: ___ M ___ F

Class Ranking: ___ Freshman ___ Sophomore ___ Junior ___ Senior

TABLE 2
COMPARISON OF SAMPLE TO POPULATION DEMOGRAPNICS

	Sample (N=316)	Population (N=2228)
First Year	13.6%	32.1%
Second Year	24.0	24.0
Third Year	26.3	22.0
Fourth Year	36.1	21.9
Male	48.3	41.4
Female	51.7	58.6

ADMINISTRATION AND ANALYSIS

The FCEQ was administered to approximately 320 randomly selected students in a cross-section of business classes in the School of Business Administration at a small, AACSB-accredited liberal arts university. A comparison of the sample demographics to the actual population indicates some skewing of the sample. Fewer first-year students were represented in the sample and more senior-year students were represented than as reflected in the university population. The second and third-year population was almost exactly represented in the sample. Male and female students were fairly closely represented in the sample as compared to the university gender make-up. Demographic comparisons are illustrated in Table 2 above. Given the exploratory nature of this study, adjustments in sampling procedures will be addressed in future replications.

Data analysis was straightforward given that this was an exploratory study in student perceptions of the SET process. After descriptive statistics were assessed, ANOVA was used to assess gender and year-in-school differences for the entire sample on the 16 individual items. Additionally, a multiple regression analysis was performed using Item 16 as the dependent variable for the other 15 scale items. Table 3 indicates the item number, mean and standard deviation for each item.

TABLE 3
ITEM NUMBERS, MEANS AND STANDARD DEVIATIONS

Item #	Mean	Std. Dev.
1.	3.03	.80
2.	1.84	.75
3.	3.23	.84
4.	2.46	.71
5.	2.45	.80
6.	1.95*	.90
7.	2.91	.83
8.	2.54	.76
9.	2.32	.78
10.	2.56	.77
11.	2.36	.76
12.	2.37	.78
13.	2.16	.96
14.	3.29	.62
15.	2.98	.68
16.	2.64	.78

* Reverse scored item

RESULTS

In considering the means for the individual items, the authors decided that items with averages of 2.70 or above represented positive perceptual sets on the part of the students in this sample. Hence, within this interpretive framework for the "Student Responses about Themselves (Items 1 – 7), dimension students indicated that they took the process seriously (Item 1, Mean = 3.03), felt comfortable with giving negative feedback (Item 3, Mean = 3.23) and believed that the SET process is important (Item 7, Mean = 2.91).

Currents suggest that faculty member gender does not matter greatly to the students in terms of rating (Item 2, Mean = 1.84) and that students have limited reservations about sharing negative qualitative comments about faculty (Item 6, Mean = 1.95). Additionally, it would appear that within this sample, expected grade in a class (Item 5, Mean = 2.45) and the personality of a faculty member (Item 4, Mean = 2.46) are not of major importance in their ratings of faculty.

In looking at the results for Items 8 – 12 (Student Responses about Professors), all item means are in the 2.32 – 2.56 range. Given the four point scaling for the FCEQ, the five items related to that dimension remain in the "neutral" range. It appears that students in the current sample are somewhat unsure about how evaluations are used or whether faculty behavior actually changes or is influenced constructively by SET feedback.

Lastly, in reviewing the means for the third section of the questionnaire, Student Responses about the Evaluation Process (Items 13 – 16), while the students appear to find the SET questions clear (Item14, Mean = 3.29) and relevant to faculty evaluation (Item 15, Mean = 2.98), they are less than wildly enthusiastic about the overall process (Item 16, Mean = 2.64).

The authors recognize that Item 13 was probably confusing as presented so the results are likely without much importance in the current study.

Using ANOVA to assess gender and class rank differences, results suggest a number of significant statistical differences in responses related to gender and class ranking. For gender differences, five significant differences emerged. Female students took the evaluation process more seriously than their male counterparts (Item # 1, Males = 2.87, Females = 3.17, F = 11.41, p = .001). Additionally, female students perceived the evaluation process as more important than did their male counterparts in the sample (Item #7. Males = 2.79, Females = 3.01, F = 5.72, p = .017). Gender difference results also suggest that male students may experience more cynicism related to the SET process than do female students. Males were significantly different than female students in terms of their perception that the higher the grade projected the higher their evaluation of a professor (Item # 5, Males = 2.54, Females = 2.37, F = 3.62, p = .058) and their belief that professors adjusted their in-class behavior at the end of the semester to achieve higher evaluations (Item # 12, Males = 2.49, Females = 2.27, F = 6.031, p = .015). Overall, these findings suggest that females have a more positive and less cynical view of the SET process and dynamics than do males, although only four significant gender differences emerged.

In regards to year-in-school (class ranking), three significant, yet relatively consistent differences were identified. First-year students (more so than sophomores, juniors and seniors) tended not to rate female professors higher than their male counterparts (Item # 2, First Year = 1.58, Second Year = 1.88, Third Year = 1.98, Fourth Year = 1.82, F = 3.098, p = ,027). Upperclassmen in the sample, therefore, reported rating their female professors higher than their male professors. Results on other items suggest that first-year students believe that faculty took the evaluation process comments more seriously than sophomores, juniors and seniors. (Item # 8, First Year = 2.93, Second Year = 2.43, Third Year = 2.52, Fourth Year = 2.49, p = .003) Lastly, in addressing the relevance of evaluation questions to actual professor evaluation, the first-year students reported a higher perception of question relevance than did any of the three upper-class groups. (Item # 15, First Year = 3.31, Second Year = 2.91, Third Year = 2.94, Fourth Year = 2.96, F = 3.809, p = .01).

The regression analysis in which the first 15 items were regressed on the evaluation effectiveness omnibus item (Item 16) rendered a significant ANOVA level ($F = 23.46$, $p < .000$) with approximately 51% of the variance explained by the model. Five items emerged as statistically significant in their contribution to the understanding of students' overall perceptions of the effectiveness of the faculty class evaluation process.

The five items that loaded significantly on their relationship to perceived SET process effectiveness were (in order of significance): 1) Evaluation Question Relevance in Reference to Teaching Evaluation (Item 15, t-value = 6.99, p > .000); 2) Importance of the Process, Student Perspective (Item 7, t-value = 3.93, p > .000); 3) Use of Evaluations by Faculty to Improve Their Courses (Item 10, t-value = 3.50, p = .001); 4) Use of Evaluations to Improve Their Teaching (Item 11, t-value = 3.14, p = .002): 5) Comfort with Giving a Negative Evaluation of a Professor (Item 3, t-value = 2.47, p = .014). Item 8 (Professors Take Evaluation Comments Seriously was marginally significant when using the .05 convention of significance (t-value = 1.88. p = .06).

DISCUSSION

The pattern of means as positive, negative or neutral should be addressed initially. This study suggests that any item with a mean at or above 2.7 is a positive indicator related to student

perceptions of the particular items and item whose mean was below 2.3 would be a negative indicator worthy of commentary. Those items with means 2.3 and 2.7 probably aren't very meaningful in terms of understanding student perceptions at this point. Overall, it would appear that the students have a fairly ambivalent collective perceptual set relative to the process and to the individual items related to the process given that eight of the sixteen items fall within the 2.3 to 2.7 range.

Five items at or above the 2.7 threshold on the scale suggest that students generally take the process seriously, feel comfortable giving "bad" professors negative evaluations and feel the questions asked on the form are both clear and relevant. Given the mean of 2.64 for the effectiveness item, further study and a rescaling of the instrument may provide a clearer perspective on the perceptions of process effectiveness that students retain.

Some specific item responses may warrant further discussion. One of the lowest rated items is # 9: perceived use of evaluation data in tenure and salary decisions. Many students simply do not understand the importance and use of SET data in the overall framework of faculty performance appraisal. The focus group which produced the instrument was also less aware of the use of the data than faculty generally perceive. In fact, some individual students actually disagreed that evaluative data was even used. Also, related to previous studies, students may be less susceptible to cognitive dissonance (giving a professor a poor evaluation in anticipation of making a bad grade) if they knew of the importance of the process. Just as supervisors are trained in performance appraisal, perhaps student should be given a statement of the importance and use of their evaluations. Finally, faculty should feel secure in the fact that, in general, students do take the process seriously (despite the misperceptions of some), and feel comfortable with giving a professor a poor evaluation. With all of the professional rancor and inconclusive research on exactly what is measured with SETs, at least our data suggests that students bring a fair amount of authenticity to the process.

The gender differences are quite limited within this sample. This finding may suggest fairly consistent perceptions of the SET process by male and female students. While there are only four significant differences between female and male students, there was no significant difference in the perceptions that the process is both important and overall an effective process. As noted above, it would appear that the female members of this sample have somewhat a more positive view of the process than do the males.

When considering year-in-school (class) as related to overall perceptions of the SET process and dynamics, results suggest surprisingly consistent perceptions across classes represented in the sample. Thirteen of the sixteen items demonstrated no significant differences based on year-in-school. As was consistent across these three significant findings, upper-class students did not differ significantly from one another on any of the three variables, while departing significantly from the views of the freshmen in the sample. It is noteworthy that no significant differences emerged for upper-class students on any of the 16 items in the FCEQ. Perhaps once indoctrinated into the SET process, students come to see the process as a routine end-of-semester exercise rather than as a meaningful exercise in feedback that may have a positive effect on educational quality and continuous faculty improvement.

The regression analysis explained a significant level of variance (51%). The first ranked item is Question Relevance to students. The SET form used at our institution is relatively simple and straightforward with only six questions, using Likert-type scales, related to course objectives, learning environment, communication style, etc. and two general questions asking for a global rating of, first, the course and finally the instructor. The remainder of the evaluation

form lists typical open-ended questions, such as, what students liked best/least and suggestions for improvement. This data indicates that simplicity may be the best approach in designing a course evaluation form, one that is generic enough to be used across disciplines. Students using the same form for all courses also may tend to take the process more seriously; faculty would use the feedback; and would be able to more easily compare professors and, therefore, feel more comfortable giving "bad" professors a negative evaluation. Other data and studies provide some support for the above discussion.

Overall then, while the current FCEQ is hardly a perfect instrument, it has helped to understand better how students see the evaluation process. With further instrument development and expanded replications of this original study, the authors hope to learn more about the perceptions of students in an effort to make the evaluation process more meaningful and relevant to them and to the faculty who engage in the process.

DIRECTIONS AND SUGGESTIONS FOR FUTURE RESEARCH

Given the exploratory nature of this research and the relatively scant attention student perceptions of the SET process has received, our intention is to do a more advanced study in the near future. At present the FCEQ is being revised based on feedback provided by colleagues at a recent SoBA research forum. A number of the items will be rewritten to make them clearer and to eliminate compound sentences that may be confusing to students who complete the instrument. It is anticipated that the revised FCEQ will include approximately 25 items. The scaling of the instrument will also be changed from its current 1- 4 scale to a 1 – 10 scale.

Our goal will be to increase sample size and to acquire a large sample of liberal arts students from within the College of Arts and Sciences. This will allow us to compare business students and liberal arts students' perceptions of their respective SET processes. While their evaluation process is different from that in our SoBA, given the generic nature of the revised FCEQ will lend itself to usage in other learning contexts where SETs are used to evaluate teaching and teachers. Lastly, a sample the perceptions of graduate students enrolled in our MBA program at several different campuses of the university will also be taken.

In conclusion, the research described in this paper represents just the third recent study of students' perceptions of SET processes and dynamics and the first to consider gender and class differences as related to perceptions of SET. While the current FCEQ is an imperfect instrument, with further instrument development and replication, a greater contribution to this area of research in the future is anticipated.

REFERENCES

Darby, J. (2006a) The effects of elective or required status of courses on student evaluations. *Journal of Vocational Education & Training*. 58(1), 19-29.

Darby, J. (2006b) Evaluating courses: An examination of the impact of student gender. *Educational Studies*, 32 (2), 187-199.

Helterbran, V.R. (2008). The ideal professor: Student perceptions of effective instructor practices, attitudes and skills. *Education,* 129(1), 125-138.

Giles, A., Martin, S., Bryce, D. and Hendry, G. (2004) Students as partners in evaluation: Student and teacher Perspectives. *Assessment and Evaluation in Higher Education*. 29(6), 681-685.

Heckert, T. (2006). Relations among student effort, perceived class difficulty appropriateness, and student evaluations of teaching: Is it possible to "buy" better evaluations through lenient grading? *College Student Journal*, 40(3), 588-596.

Maurer, T. (2006). Cognitive dissonance or revenge? Student grades and course evaluations. *Teaching of Psychology*. 33(3), 176-179.

Moore, M.J. (2008). Student perceptions of teaching evaluations. *Journal of Instructional Psychology*, 35(2), 177-181.

Mintu-Wimatt, A., Ingram, D., Milward, M.A., and Russ, C. (2006) On different teaching delivery methods: What happened to instructor course evaluations? *Marketing Education Review*. 16(3), 49-57.

McPherson, Michael (2006). Determinants of how Students evaluate teachers. *The Journal of Economic Education*. 37(1), 3-20.

Sedlmeier, Peter (2006). The role of scales on student ratings. *Learning and Instruction*. 16(5), 401-415.

Sojka, J., Gupta, A.K. and Deeter-Schmelz, D.R. (2002). Student and faculty perceptions of student evaluations of teaching. *College Teaching*, 50(2), 44-49.

Wachtel, H.K. (1998). Student evaluation of college teaching effectiveness: A brief review. *Assessment & Evaluation in Higher Education*, 23(2), 191-211.