

# The Impact Of Students' Expectations Of Grades And Perceptions Of Course Difficulty, Workload, And Pace On Faculty Evaluations

Barry Thornton, Jacksonville University, USA  
Michael Adams, Jacksonville University, USA  
Mohamad Sepehri, Jacksonville University, USA

## ABSTRACT

*Decisions concerning tenure, promotion, and merit raises are of crucial importance to college and university faculty. These decisions are greatly affected by the evaluation of faculty by their students. It is often argued that student evaluations of faculty are influenced by a number of factors that do not reflect the important elements of university level instruction, such as subject knowledge and clarity of exposition. Rather, some faculty believe that if a professor is an easy grader, has a low workload, or if the class itself is considered easy, he or she is more likely to receive a favorable student evaluation. This paper utilizes a sample of faculty evaluations from the College of Business of a small southeastern university to investigate these hypotheses.*

**Keywords:** faculty evaluations; teaching effectiveness; student evaluations of teaching; Dr. Fox effect

## INTRODUCTION

Decisions regarding tenure, promotion, and compensation are of tremendous importance to both college faculty and administrators. Professors are typically evaluated on their performance in three areas for these decisions: teaching effectiveness, professional development/scholarly activity, and service to the institution and community. The weights assigned to these three areas differ across institutions, with greater emphasis on scholarly activity at large research universities and more importance attached to teaching effectiveness at smaller institutions.

There are issues of concern for faculty in all three areas of evaluation. For example, in the area of scholarly activity, there is the question of how to measure and adjust for journal quality. Professors also have reservations about the measurement of teaching effectiveness. The student evaluation of teaching (SET) is used as the primary tool for measuring teaching effectiveness at most institutions of higher education. A large body of research shows that some professors contend that, rather than measuring learning, these evaluations may be influenced by factors such as the professor's charisma, charm, ability to entertain, easy grading policies, and low workload (Marsh, 1987; Mukherji and Rustagi, 2008; Simpson, 1995; Simpson and Siguaw, 2000; Yunker and Sterner, 1988).

Consequently, much research has been conducted on the topic of SETs and what they measure. A search for the phrase "student evaluation of teaching" in Google Scholar on 13 July 2010 showed approximately 1,500,000 articles and citations that deal with this topic. In spite of the large amount of research done in this area, the evidence on how well SETs measure student learning is mixed. Some studies provide empirical evidence that SETs do measure teaching effectiveness (Centra, 2003; Marsh, 1987). Other research suggests that SETs are significantly affected by instructor charisma and ability to entertain (Marsh and Overall, 1981; Naftulin, Ware, and Donnelly, 1973; Ware and Williams, 1975), easy grading (Greenwald and Gillmore, 1997; Weinberg, Fleisher, and Hashimoto, 2007) and light workload (Greenwald and Gillmore, 1997).

This study investigates the factors that influence student evaluations of the instruction provided by professors in the business school of a small southeastern university. Like many universities, this institution utilizes the Student Instruction Report (SIR) II to measure teaching effectiveness. SIR II is a product of the Educational Testing Service (ETS), and it has been in use for more than 30 years. While research available on the ETS web page suggests that this instrument is a valid and reliable tool for measuring student learning (Centra, 2006), this view is not universally accepted by college faculty, as noted above.

## **DATA**

A sample of 80 SIR II reports for full-time faculty teaching in the College of Business of a small southeastern university during the fall semester of 2008 was used for this analysis. Each SIR II report measures students' evaluations of the instruction provided by their professor for an individual class.

SIR II consists of 45 questions in total. The first 40 questions have Likert scale responses with values ranging between 1 and 5, with 5 being the best response. The last five questions collect student information, such as gender, class level, and expected grade. Students are provided with the questionnaire in the final weeks of the semester and after administration, they are processed by ETS. The SIR II report is then generated and made available to both faculty members and the appropriate administrators after grades have been assigned for the classes.

Question 40 asks students to "Rate the quality of instruction in this course as it contributed to your learning." This is the overall evaluation and it is normally considered to be the most important item in the SIR II report. When SIR II was being developed, factor analysis was used to group the first 39 questions into eight dimensions of college instruction (Centra, 2006). Presumably the responses to these other 39 questions and eight categories are systematically related to the overall evaluation. The eight dimensions of college instruction are listed as follows:

1. Course Organization and Planning (Questions 1 through 5)
2. Communication (Questions 6 through 10)
3. Faculty/Student Interaction (Questions 11 through 15)
4. Assignments, Exams, and Grading (Questions 16 through 21)
5. Supplementary Instructional Methods (Questions 22 through 28)
6. Course Outcomes (Questions 29 through 33)
7. Student Effort and Involvement (Questions 34 through 36)
8. Course Difficulty, Workload, and Pace (Questions 37 through 39)

Several variables were extracted from the SIR II reports for this study, including the overall evaluation and the mean values of six of the eight dimensions of college instruction (dimensions A, B, C, D, F, and G). The mean response was not calculated for Course Difficulty, Workload, and Pace (dimension H). Instead, the fraction of the class giving a particular response to each question was used instead. For example, the analysis looks at the proportion of students who thought the pace of the class was very fast or somewhat fast. Other variables used in this study are the instructor's gender, the number of students who filled out the evaluation, class level (graduate or undergraduate), and student perceptions of workload, effort, pace, as well as their grade expectations. Responses to Supplementary Instruction Methods (dimension E) were not used in this research, since many students chose the response "Not Used" in a large fraction of the classes in the sample.

## **METHODOLOGY**

Multiple regression was utilized for the analysis of the sample data. The mean overall evaluation was regressed on 13 independent variables which are listed in Table 1.

**Table 1: Independent Variables in the Regression Model**

Independent Variable	Description
Gender	Binary variable set equal to 0 if female instructor, 1 if male instructor
Number	Number of students responding to the questionnaire
Organization and Planning	Mean response to course organization and planning
Communication	Mean response to communication
Interaction	Mean response to faculty/student interaction
Grading	Mean response to assignments, exams, and grading
Outcomes	Mean response to course outcomes
Effort	Mean response to student effort and involvement
Difficulty	Fraction of respondents who thought class was very or somewhat difficult
Workload	Fraction of respondents who thought workload was much heavier or heavier than other classes
Pace	Fraction of respondents who thought pace of instruction was very fast or somewhat fast
Expect A	Fraction of respondents who expected to earn an A in the class
Graduate	Binary variable set equal to zero if undergraduate class, 1 if graduate class

**RESULTS**

Table 2 shows the regression results from the “kitchen sink” or full model. While the statistics associated with the evaluation of the overall model (the F statistic and adjusted r-squared) indicate strong explanatory power, only three variables were significantly related to the dependent variable - number, organization and planning, and outcomes.

**Table 2: Multiple Regression Results – Full Model – Dependent Variable is Overall Evaluation**

Independent Variable	Estimated Coefficient	T-Ratio	P-Value
Constant	-0.199	-0.53	0.600
Gender	0.006	0.14	0.886
Number	-0.006	-1.97	0.053
Organization and Planning	0.608	3.95	0.000
Communication	0.117	0.183	0.525
Interaction	0.690	0.87	0.386
Grading	0.013	0.13	0.899
Outcomes	0.383	3.93	0.000
Effort	-0.155	-1.43	0.159
Difficulty	0.080	0.52	0.608
Workload	-0.097	-0.78	0.439
Pace	0.074	0.49	0.628
Expect A	-0.053	-0.51	0.611
Graduate	0.022	0.46	0.650
Adjusted R Squared = 89.6%		F = 53.32	P-Value = 0.000

An iterative procedure was employed to arrive at a reduced model. The variable with the smallest (in absolute value) t-ratio was removed from the model and the regression equation was estimated again. This process was continued until all the variables included in the model had coefficients that differed significantly from zero. The variables removed from the full model (in order from first removed to last) were Grading, Gender, Pace, Graduate, Expect A, Communication, Workload, Difficulty, and Interaction. Table 3 shows the regression results for the reduced model.

**Table 3: Multiple Regression Results – Reduced Model – Dependent Variable is Overall Evaluation**

Independent Variable	Estimated Coefficient	T-Ratio	P-Value
Constant	0.068	0.25	0.803
Number	-0.006	-1.94	0.056
Organization and Planning	0.720	8.91	0.000
Outcomes	0.392	5.73	0.000
Effort	-0.133	-2.28	0.026
Adjusted R-Squared = 90.3%		F = 184.87	P-Value = 0.000

## ANALYSIS

After controlling for the influence of the other variables in the reduced model, there doesn't appear to be any evidence to support the hypotheses that the overall evaluation is affected by grading, workload, or pace. However, the sample evidence does provide support for the hypothesis that the amount of effort that a class requires is negatively related to the overall evaluation. The three individual questions that make up the category Student Effort and Involvement (Effort) are:

1. 34. I studied and put effort into this course ...
2. 35. I was prepared for each class (writing and reading assignments) ...
3. 36 I was challenged by this course ...

The possible responses to these questions are:

- 1 Much Less Than Most Courses
- 2 Less Than Most Courses
- 3 About the Same as Others
- 4 More Than Most Courses
- 5 Much More Than Most Courses

Since the estimated coefficient on the variable Effort is negative, it can be concluded that professors teaching courses that are more challenging for students or courses that require more effort and preparation by students are likely to have lower overall evaluations, after controlling for the effect of the other variables in the model.

The other factors that play a role in determining overall evaluations from the SIR II report are Organization and Planning, Outcomes, and Number. The negative coefficient on Number indicates that, after controlling for the effects of other variables, professors teaching smaller classes can be expected to have higher overall evaluations. It should be noted that the size of the coefficient is small - for each additional student, it is expected that the overall evaluation will be lower by 0.006 points.

In contrast, both Organization and Planning and Outcomes have a relatively large positive impact on a professor's overall evaluation. In particular, for every one unit increase in a faculty member's mean rating on Organization and Planning, it can be expected that the overall evaluation will be larger by 0.72. For the variable Outcomes, which measures (among other things) students' perceptions of their learning in the course, for each one unit increase in the mean value, it is expected that a professor's overall evaluation will be higher by 0.392, *ceteris paribus*.

## CONCLUSIONS

One of the most basic lessons of inferential statistics is that inferences from a sample are only applicable to the population from which the sample was drawn. Strictly speaking, the inferences from this study are only relevant for the business school of the small southeastern university from which the sample data were taken. However, it is likely that this business school is similar to other business schools in small universities, so the applicability of the results may be wider than just one small university.

Perhaps the most surprising result of this analysis is the variables that do not play a role in determining the value of a faculty member's overall evaluation on the SIR II report. Communication includes such items as clarity of exposition and enthusiasm for course material, and Interaction (Faculty/Student Interaction) includes concern for student progress and helpfulness. Yet, these variables were not found to be significantly related to the overall evaluation.

For professors who contend that the overall evaluation from SIR II reports is "contaminated" by the effects of easy grading, low workloads and low effort, the results of this research are mixed. There is no evidence that easy grading or low workloads translate into higher evaluations. However, there is support for the belief that teaching a challenging class requiring higher levels of student effort and preparation can result in lower overall evaluations.

#### **AUTHOR INFORMATION**

**Dr. Thornton** is a professor of economics at Jacksonville University. His research interests are firm valuation, IPO pricing, sustainability, and student evaluations of teaching.

**Dr. Adams** is an associate professor of finance at Jacksonville University. His research interests include firm valuation, IPO pricing, sustainability, and a variety of topics in higher education.

**Dr. Sepehri** is the Associate Dean and professor of management & international business at Jacksonville University. His research interests are international trade, global business strategies, regional economic integration, and organizational behavior and development.

#### **REFERENCES**

1. Centra, J. A. (2003). Will students receive higher student evaluations by giving higher grades and less coursework? *Research in Higher Education*, Volume 44, Number 5, October 2003, 495 – 518.
2. Centra, J.A. (2006). The Student Instructional Report II™: Its development, uses, and supporting research (An issue paper from ETS). Retrieved from Educational Testing Service SIR II Website: [http://www.ets.org/Media/Products/SIR\\_II/pdf/53228\\_sirII\\_white\\_paper.pdf](http://www.ets.org/Media/Products/SIR_II/pdf/53228_sirII_white_paper.pdf)
3. Greenwald, A.G., and Gillmore, G.M. (1997). No pain, no gain? The importance of measuring course workload in student ratings of instruction. *Journal of Educational Psychology*, Volume 89, Number 4, 743 – 751.
4. Marsh, H.W. (1987). Students' evaluation of university teaching: Research findings, methodological issues, and directions for future research. *International Journal of Educational Research*, Volume 11, Number 3, 253-388.
5. Marsh, H.W., and Overall, J.U. (1981). The relative influence of course level, course type, and instructor on students' evaluation of college teaching. *American Educational Research Journal*, Volume 18, Number 1, March 1981, 103 – 112.
6. Mukherji, S., & Rustagi, N. (2008). Teaching evaluations: Perceptions of students and faculty. *Journal of College Teaching & Learning*, Volume 5, Number 9, September 2008, 45 – 54.
7. Naftulin, D.H., Ware, J.E., & Donnell, Y. (1973). The Dr. Fox lecture: A paradigm of educational seduction. *Journal of Medical Education*, Volume 48, September 1973, 630 – 635.
8. Simpson, P.M., and Siguaw, J.A. (2000). Student evaluations of teaching: An exploratory study of the faculty response. *Journal of Marketing Education*, Volume 22, Number 3, December 2000, 199 – 213.
9. Simpson, R.D. (1995). Uses and misuses of student evaluations of teaching effectiveness. *Innovative Higher Education*, Volume 20, Number 1, Fall 1995, 3 – 5.
10. Ware, J.E., and Williams, R.G. (1975). The Dr. Fox effect: A study of lecturer effectiveness and ratings of instruction. *Journal of Medical Education*, Volume 50, February 1975, 149 – 156.
11. Weinberg, B.A., Fleisher, B.M., and Hashimoto, M. (2007). Evaluating methods for evaluating instruction: The case of higher education (A working paper from the NBER). Retrieved from the National Bureau of Economic Research Website: <http://www.nber.org/papers/w12844>
12. Yunker, P.J., and Sterner, J. (1988). A survey of faculty performance evaluation in accounting. *Accounting Educators' Journal*, Volume 2, 63 – 71.

NOTES