A Chair's Guide to Student Evaluations of Teaching

William P. O'Dea*

ABSTRACT

This paper attempts to analyze the factors that influence student evaluations of teaching effectiveness. The student evaluation instrument at SUNY-Oneonta contains thirteen items. The thirteenth item is summative and asks the students to provide an "overall evaluation of teaching effectiveness." The first twelve items represent the different qualities that a good teacher should possess. Using ordinary least squares analysis, we find that 96 percent of the variation in the summative "overall effectiveness" item can be explained by the twelve qualities of good teaching. Adding additional variables such as the instructor's gender, years of experience and grade distribution does not improve the model's explanatory power and none of these variables are statistically significant. The results indicate that our students value instructors who deliver well-organized lectures, answer questions clearly and are available for consultation. Importantly, the results show that our students do not penalize faculty members who are rigorous and maintain demanding grading standards.

I. INTRODUCTION

As department chair, one of my major responsibilities is writing evaluation letters for my faculty members who are requesting contract renewal, tenure or promotion. In these letters, I am expected to assess the faculty member's performance in the standard areas of teaching, research and service. However, since the primary mission of the College at Oneonta is undergraduate education, the section of my letter dealing with teaching is especially important.

In assembling the folders that accompany their requests, we require our faculty to provide a portfolio of materials that can be used to evaluate their teaching. They are expected to provide copies of syllabi and examinations, their grade distributions, a self-evaluation of their teaching and two peer reviews based on classroom visits. Last, and by no means least, they must submit summaries of the student perception of instruction (SPI) questionnaires administered in all courses taught during the preceding four semesters.

By examining syllabi and reading examinations and peer reviews, it is possible to form a reasonably accurate impression of the level at which courses are being taught and the demands being placed on students. However, reading through this material takes a considerable amount of time and the impression formed is unavoidably subjective. The SPI data, by contrast, come in a readily digestible form that gives the illusion, if not the reality, of being scientifically precise. In addition, our SPI form has a bottom line question that asks the student to provide an "overall evaluation of teaching effectiveness." In my letter, I am expected to address the SPI data. Not to do so would place my faculty members at risk.

* Department of Economics, Finance and Accounting, State University of New York College at Oneonta, Oneonta, NY 13820. (607) 436-2127 odeawp@oneonta.edu

I am also a member of the College Promotion and Tenure Committee. Attempts to parse the meaning of the SPI data are at the heart of our discussions of teaching performance. Indeed, it has struck me that even members of the Committee who profess to place little credence in the SPI results still use them to support their arguments for or against a candidate's teaching effectiveness.

Having looked at the SPI results from thousands of course sections, I assume that in filling out the questionnaires our students are trying to tell us something. Whether that something is what we as an institution want to know is what is at issue. We want to evaluate the quality of teaching. Can the faculty member effectively convey a body of knowledge? Can the students apply the concepts developed in a course to relatively unstructured problems? Of course, we would also like our students to retain what they learn. Rephrased, do the SPI data provide insight into the quality of an instructor's teaching rather than into the instructor's popularity or entertainment value?

The outline for the reminder of the paper is as follows. The next section is devoted to a brief literature review. The literature on student evaluations of instruction is enormous. Unfortunately, for a chair looking for guidance, this enormous effort has produced relatively few hard answers. The third section discusses how the SPIs are administered at the College at Oneonta and presents summary SPI data along with data on instructor and course characteristics. The results of preliminary data analysis are also presented in the third section. The fourth section concludes.

II. AN IMPRESSION OF THE LITERATURE

The literature on student evaluations of instruction is huge. As of 1999, Becker and Watts (1999) report that "nearly 2000 studies have been conducted (p. 344)." Work in this area did not cease in 1999. Having neither the time nor the inclination to wade through 2000 academic papers, I decided to limit my attention to papers appearing in the *Journal of Economic Education* during the last 25 years. My premise is that, since the JEE is the leading journal in the field of economic education, reading articles appearing there should provide an accurate impression of the state of the art. By limiting my attention to the JEE I reduced my reading list to approximately 30 articles. Of these 30, I have read about 15. Consequently, this section represents my impression of the literature.

The literature appears to have three major concerns. The first is to identify the important explanatory variables that drive the student evaluations. The second is to establish the relation between grading standards and the student evaluations. Siegfried and Fels (1979) put the matter very bluntly: "By far the most controversial issue is whether instructors can 'buy' higher evaluations by lowering the (effort) price to students of achieving a given grade (p. 931)." SPIs became common in the middle to the late 1970s. Before this time, few colleges asked students to evaluate the quality of teaching. Now, very few colleges don't ask their students to rate their instructors (Becker and Watts, 1999). The period during which SPIs became increasingly common coincided with an increase in student grades. The temptation to infer a causal relation is obvious (Zaganzadeh, 1998). The third focus of the literature is to identify proper estimation techniques. The least cost solution to any estimated problem is to resort to ordinary least

squares (OLS). The problem is that OLS may not be the best technique to employ in the statistical analysis of students evaluations.

In addressing the first two concerns, Needham (1978) provides a simple theoretical framework that is useful. Students have utility functions whose arguments are leisure time and some academic result. For Needham, the academic result is the course grade. But there is no reason why it could not be the amount learned. We can visualize classes as being populated by a mixed group of students some of whom view the purpose of a college education as being the acquisition of a body of knowledge and skills and some of whom view the point as being the acquisition of a marketable credential. The utility function has the conventional convex shape. The student also has a concave production possibility frontier (PPF) that describes the rate at which leisure time can be converted into a grade (learning). The optimal point is where the PPF is tangent to the highest reachable indifference curve. Instructors are able to influence this process by rotating the PPF to the right (assuming that leisure is placed on the vertical axis). In other words, for a given sacrifice of leisure, the student can receive a higher grade (learn more). A rightward shift of the PPF enables the student to reach a higher indifference curve. An increase in student welfare would then lead to better evaluations.

Instructors can shift the PPF by fair means and foul. Instructors can improve the quality of their teaching. For example, they can deliver better organized and clearer lectures, provide more vivid examples, give more precise and direct answers to student questions, hold more frequent and convenient office hours, and tell better jokes to keep the students engaged and awake. These strategies would work for both types of student. Higher SPI scores in this scenario would indicate an improvement in the quality of teaching. The problem is that it is also possible to enable students to obtain higher grades with a smaller time commitment by relaxing standards. For example, instructors can give easier examinations or grade them more leniently. This strategy would be appreciated by the “education as a credential” group. However, the “body of knowledge” group would not appreciate the decrease in the quality of their education. Depending on which group of students is larger, the SPI ratings could either go up, down or remain constant.

In a setting where faculty members make up and grade their own examinations, which is the norm at the College at Oneonta, grades cannot be interpreted as an unambiguous measure of what the students have learned. High grades can be the product of effective instruction or lax standards. Low grades could reflect rigorous instruction or ineptitude.

With regard to the question of what instructor qualities are valued by students, the literature is reassuring. Boex (2000) reports that “from the students’ point of view, organization and clarity was the single most important attribute of effective economics instruction (p. 213).” Bosshardt and Watts (2001) find that students care about enthusiasm and preparation. DeCanio’s (1986) results highlight the importance of organization and structure. Nelson and Lynch (1984) find “clarity of communication and instructor enthusiasm” to be the most important factors explaining favorable student evaluations (p. 21).

With regard to the impact of grades on evaluations, the literature is somewhat reassuring. Giving higher grades does not appear to be a foolproof strategy for obtaining higher SPI ratings. Nelson and Lynch (1984) report that a one point increase in the average grade in a course (on a four point scale) only raises the instructor rating by .15 (on a five point scale). DeCanio (1986) finds "no evidence that the expected grade influences the instructor's teaching effectiveness score (p. 172)." On the other hand, Mirrus found that a one point increase in the average grade would lift the SPI rating by .85 of a point (cited in Siegfried and Fels, 1979). In Boex's (2000) study, the average grade has a statistically significant impact on the student ratings. However, other factors such as clarity of instruction and organization tend to be more important.

Of course, even if grades and SPI scores are positively related, this is not necessarily a problem if high grades are due to effective instruction that results in more learning. Again, the problem is that grades do not provide an unambiguous measure of what students have learned, since examination design and grading are under the control of the individual instructor. Of particular interest, then, are the papers by Carrell and West (2010) and Gramlich and Greenlee (1993). Both papers consider settings (required math and science courses at the United States Air Force Academy and the introductory and intermediate economic theory courses at the University of Michigan, respectively) where course content and examination design and grading are under department control rather than under the control of the individual instructor. In addition, at both institutions, students are randomly assigned into course sections. These are settings where the "teacher as coach" metaphor might be accurate. Carrel and West found that instructors whose students performed better on the common examinations tended to receive better student evaluation scores. Gramlich and Greenlee found a "statistically significant but small" relation between student grades and faculty ratings (p.4). Interestingly, Carrel and West found that instructors whose students performed well in introductory level courses did not tend to perform as well in upper-level follow on courses. They found that "student evaluations are positively correlated with contemporaneous professor value added and negatively correlated with follow-on achievement (p 412)." "Deep learning", which is the object of the educational enterprise, was not recognized and rewarded by the student evaluations.

The nature of the data sets used in the statistical analysis of student evaluations raises some interesting econometric issues. For studies where the data point is an individual student evaluation the discrete nature of the responses suggests that some sort of logit model rather than OLS might be the proper estimation techniques (DeCanio, 1986). If the data point is a course section, the average of student responses to the items on the SPI questionnaires can be continuous between the floor and the ceiling response. However, the existence of the floor and the ceiling still argues in favor of a logit approach. In addition, the existence of upper and lower limits might result in heteroskedasticity because the error distribution is truncated (Siegfried and Fels, 1979).

Moreover, a single equation estimation approach might not be appropriate (Mehdizadeh, 1990). For example, student grades or expected grades are commonly used as an explanatory variable in the

equation accounting for student ratings. However, given the control that instructors have over the grading process, it would be inappropriate to treat grades as an exogenous. Since the grades are endogenous, an equation characterizing the process by which grades are determined would be needed. These points having been made, I am going to use a simple OLS model for my initial data analysis.

III. BACKGROUND AND DATA ANALYSIS

At Oneonta, it is College policy that the SPI questionnaires be administered in every course every semester. Individual departments can use a College questionnaire or substitute an instrument of their own design. The Division of Economics and Business, which houses the Departments of Economics, Finance and Accounting (EFA) and Management, Marketing and Information Systems, uses its own questionnaire.

The timing and mechanics of administering the SPI instrument are up to the individual faculty member. Typically, the SPIs are administered during the last two weeks of the semester. Some brave souls administer the evaluations during the final examination period. Instructors are not supposed to administer the questionnaires. Some instructors recruit other members of the faculty to hand out and collect the forms. Other instructors explain the process and hand out the forms designating a student in the class to collect the forms and return them to the department office. The department secretary then types up the written comments. The forms are then sent to the Office of Institutional Research which generates summary statistics and provides analysis of the individual questions.

The questionnaire used by the Division of Economics and Business is relatively compact. It asks the students to answer thirteen questions and provide their GPAs. At other institutions, the students are asked to answer many more questions. For example, the Penn State form that provided the data for the Nelson and Lynch (1984) study contained 33 questions. The Georgia State University questionnaire asked the students to answer 35 questions (Boex, 2000). The questions we ask are fairly standard and deal with issues of classroom management, course organization and testing. We don't have a question dealing with instructor enthusiasm but we do ask about poise. Unlike the instruments at other institutions, we don't ask the students to provide their expected grade for the course or to rate the value of the course. Our bottom line question focuses on instructor effectiveness. On our form, one is excellent (or the equivalent) and five is poor. On the back of the form, the students can provide written comments. The student comments tend to be terse. If our students write more than a sentence or two, it is generally a sign that they are not pleased.

Table 1 provides data on the average scores for the thirteen questions for the course sections taught in EFA from the fall of 2008 to the spring of 2010. There were a total of 147 course sections. In general, the scores are slightly over two which means that on the" overall evaluation of teaching effectiveness" and most of the individual components of instructor performance our students think their instructors are (almost) very good.

One way to approach the form is to regard the first twelve questions as describing the qualities that we think that an excellent instructor should possess. To determine which of these qualities have the greatest impact on the bottom line "overall effectiveness" rating, I regressed "overall effectiveness" on the twelve qualities using OLS. The Koenker-Bassett test (Gujarati, 2003) indicates that heteroskedasticity is not present. Some of the correlations between the questionnaire items are high enough to suggest that multicollinearity could be a problem, but at this stage I have made no effort to address the issue.

Table 1: Summary Statistics Teacher Evaluation Questionnaire

Question	Mean	Standard deviation
Rigorousness of the course	2.28	.52
Organization of the Course	2.23	.61
Teaching skill is evidenced by classroom presentation	2.24	.69
How well instructional materials were coordinated with lecture	2.2	.64
Poise and self confidence	1.86	.54
Planning and clarity of examination questions	2.34	.66
Ability to answer questions and clearly explain concepts	2.14	.68
Tolerance of attitudes and opinions of others	1.88	.5
Ability to maintain control of class discussion	2.07	.53
Availability for consultation	1.96	.47
Maintains demanding grading practices	2.06	.43
Fairness in grading	2.11	.43
Overall evaluation of teaching effectiveness	2.25	.66

The results are shown in Table 2. Overall, our twelve qualities of effective teaching account for 96 percent of the variation in the "overall evaluation of teaching effectiveness "ratings, which suggests that our questionnaire includes the major qualities that our students believe that an effective instructor should possess. The single most important explanatory variable is "teaching skill as evidenced by classroom presentations." My interpretation is that our students value an instructor's ability to deliver a coherent presentation. The ability to answer questions, the planning and clarity of examination questions, the coordination of materials with the lecture, and availability for consultation were also very important. The results indicate that delivering a rigorous and demanding course and maintaining demanding grading practices tend to result in higher (worse) ratings on the "overall evaluation" item. However, the results are small and statistically insignificant. The results indicate that our students do not overly penalize instructors who hold them to high standards. It could be that our students expect economics, finance and accounting courses to be technical and difficult and are not shocked when this expectation is met.

To determine whether the existence of a floor and a ceiling for the ratings affected the results, I reran the regression dropping the observations at the very top (ratings of 1.5 and below) and the very bottom (ratings of 4 and above). The coefficient estimates did not change very much nor did the standard errors.

It is also possible that course and instructor characteristics affect the student evaluations. (Although the results in Table 2 indicate that there is not a lot of variation left to account for.) Table 3 presents summary data for course and instructor characteristics that have been employed as explanatory variables in other studies. The number of students enrolled in the course captures the possibility that students might prefer smaller classes because they allow more interaction between the instructor and the students. The percentage of surviving students who complete the SPI form accounts for the possibility that the sample of students completing the SPI forms might not be representative of the entire population of students in the course. Thus, the SPI data are subject to sample selection bias. Boex (2000) raises the possibility that students who are disappointed with a course are likely to stop attending class and thus will not fill out the SPI questionnaire. The result would be a higher overall SPI rating. Becker and Watts (1999) suggest that instructors might manipulate the response rate by deciding when to administer the form. For example, at our institution, an instructor might administer the evaluation on a Friday, which is a day with a relatively high absenteeism rate. Of course, if the group of students completing the questionnaires includes the most talented and hardworking students in the course, these would be the students who are best able to evaluate an instructor's effectiveness.

TABLE 2: OLS Regression Results

Variable	Estimated Coefficient	Standard Error	t-statistic
Constant	-.0742	.0814	-.9115
Rigor	-.0117	.04	-.293
Organization	.0758	.0652	1.1631
Teaching skill	.2982	.0846	3.5232**
Coordination of materials	.1528	.0628	2.4324*
Poise	.01984	.0505	.393
Planning and clarity of exam questions	.1635	.0515	3.1767**
Ability to answer questions	.2	.0561	3.5664**
Tolerance	.0133	.0419	.3185
Maintain Control	-.0104	.0352	-.2965
Availability for consultation	.1283	.0434	2.9539**
Demanding grading	-.0583	.053	-1.0994
Fairness of grading	.0938	.0441	2.12*

Number of observations=147 R^2 = .9657 adjusted R^2 = .9627 F = 315.0679
** significant at the 99 percent confidence level * significant at the 95 percent confidence level

Table Three also contains data on the level of the course. The introductory 100 level courses have very heavy enrollments of students who come from outside the Division of Economics and Business and are taking the course to satisfy general education requirements or a related work requirement for their majors. These students might be more interested in an easy grade than in what they learn. The 200 and 300 level courses are taken mainly by Business Economics, Economics, and Professional Accounting majors. The 200 level required courses are the intermediate economic theory courses, corporate finance, and business law, which are taken by Business Economics and Professional Accounting students to satisfy major requirements. Presumably, some of these students would not be very disappointed if, say, intermediate microeconomic theory were dropped from the list of major requirements. The 200 and 300 level major courses are taken by economics and accounting majors and students pursuing a concentration in finance, i.e. these course are taken by volunteers who should know what they are in for.

TABLE 3: Summary Statistics Class and Instructor Characteristics

Variable	Mean(Proportion)	Standard deviation
Enrollment	34.63	14.12
Withdrawals	2.32	3.05
Number of SPI forms completed	25.46	9.91
Percentage of survivors completing SPI form	80.45	13.4
100 level	42.86 (63/147)	
200 level required	31.29 (46/147)	
200 level major	8.84 (13/147)	
300 level major	17.12 (25/147)	
Median grade	2.82	.45
Male	65.36 (96/147)	
Non-native	15.65 (23/147)	
Rank (Assoc., full)	24.5 (36/147)	
Semesters of teaching experience at Oneonta	11.52	11.56

The average of the student rating for the overall effectiveness item is 2.54 (halfway between very good and good) for all 100 course sections. For 200 level required courses, 200 level major courses and 300 level major course, the overall rating is much lower (better); the ratings are 2.17, 1.9 and 1.81 respectively. The difference between the overall rating in the 100 level sections and the upper division sections is statistically significant. The median grade for all course sections taught in EFA is a 2.82 which is a little over a B-. The median grade for all course sections taught in the College is in the A-/B+ range. It is important to note that the median grade is based on the entire population of students who complete a course. This population is larger than the population completing the SPI forms. Again, the issue is whether the group of students completing the form is representative. Table 3 also provides data on the

gender and the nationality of the course instructor. The percentage of courses taught by senior faculty members is also shown. In order to be promoted to the ranks of associate or full professor at our College a faculty member must demonstrate outstanding proficiency in the area of teaching. (Of course, there is also the possibility that once promoted the faculty member can rest on his or her laurels.) Finally, to account for the possibility that instructors with more experience are more effective, the table provides data on the number of semesters of prior teaching experience at Oneonta. In the next step of my research, I intend to collect data on the time of day at which courses are taught, the average GPA reported by the students in a course section, and instructor research productivity.

Table 4: OLS Regression Results for Course and Instructor Variables

Variable	Estimated coefficient	Standard Error	t-statistic
Constant	2.34	.6281	3.729**
Enrollment	.0067	.006	1.1169
Response rate	.0062	.0041	1.514
200 Required	-.3635	-.3635	2.6951**
200 Major	-.1231	.2437	.5052
300 major	-.03546	.2302	1.54
Median grade	-.266	.1462	1.932+
Male	.0397	.1158	.3426
For	.495	.1745	2.8371**
Rank	-.4575	.1292	3.5406**
Semesters of Exper.	.0108	.0056	1.9318+

Number of observations=147 R^2 = .3075 adjusted R^2 = .2566 F = 6.0388
** Significant at the 99 percent confidence level + Significant at the 90 percent confidence level

To gauge the impact of these variables on the student evaluation, I regressed the "overall evaluation of teaching effectiveness" on the course and instructor characteristics. The results are shown in Table 4. The coefficients for the dummy variables for the upper division courses (the dummy took on a value of 0 for 100 level course sections) indicate that the "overall evaluations of teaching effectiveness" ratings are lower (better) in the upper division courses than they are in the 100 level courses. The difference is statistically significant for the 200 level required courses. The rank dummy variable, which takes on a value of 1 for courses taught by faculty at the associate and full professor ranks, shows that student ratings of faculty effectiveness are significantly better in sections taught by senior faculty. On the other hand, the dummy variable for course taught by instructors born outside of the United States (FOR) shows that student evaluations of faculty effectiveness are significantly higher (worse) in sections taught by non-native instructors. The results indicate that the more semesters of teaching experience an instructor has the higher (worse) his or her ratings trend to be. The estimate is significant at the ten percent level. The median grade coefficient indicates that one point increase in the median grade in a course section lowers

(improves) the "overall evaluation of teaching effectiveness" by approximately .25 of a point. The coefficient is significant at the ten percent confidence level. Gender of the instructor, the number of students enrolled in the course and the response rate on the SPI forms do not appear to matter much. As a final exercise, I then added four of the course and instructor variables to the list of instructor qualities in Table two. The variables I selected are: whether the course section was upper division (a dummy variable which takes on a value of one for a course taught at the 200 or 300 level), the median grade, whether the course was taught by a non-native instructor, and whether the course was taught by a senior faculty member. The results are presented in Table 5. Adding the course and instructor variables contributes no explanatory power to the model. Indeed, adjusted R^2 declines slightly. None of the course or instructor variables are statistically significant. Changes in the median grade now have a negligible and insignificant impact on the "overall evaluation of teaching effectiveness" rating. The implication is that the main factors driving a faculty member's rating on the "overall effectiveness" summative item are the twelve characteristics of effective teaching. For example, the higher (lower) evaluation of teaching effectiveness in 100 level course sections can be traced back to the fact that the students in the sections gave their instructors higher (worse) ratings in each of the twelve characteristics.

Table 5: OLS Regression Results

Variable	Estimated coefficient	Standard Error	t-statistic
Constant	.0785	.1304	.6024
Rigor	-.0081	.0422	.1923
Organization	.0795	.0705	1.128
Teaching skill	.3168	.095	3.3362**
Coordination of materials	.1343	.0714	1.8811+
Poise	.0227	.0521	.4359
Planning and clarity of exam questions	.1621	.0529	3.0679**
Ability to answer questions	.1953	.0602	3.2457**
Tolerance	.0024	.0438	.0548
Maintain Control	-.0199	.0391	.511
Availability for consultation	.1348	.0452	2.9834**
Demanding grading	-.0562	.0554	1.0137
Fairness of grading	.1053	.0468	2.2532*
Upper Division	-.0209	.0314	.6638
Median grade	-.0015	.0346	.0424
For	.0146	.0457	.319
Rank	.0298	.0355	.8383

Number of observations=147 R^2 = .9661 adjusted R^2 = .9619 F =231.7745
** Significant at the 99 percent level * Significant at 90 percent level + Significant at the 90 percent level

IV. CONCLUSION

Overall, my results suggest that the component items in our evaluation form do a good job of accounting for the variation in the summative "overall evaluation" item. The qualities of good instruction that matter most to the students are qualities that should also matter to the institution. Importantly, the results indicate that there is no need for instructors to attempt to improve their ratings by lowering the rigor of their courses or by relaxing their grading standards. The results indicate that the basic "blocking and tackling" of teaching, delivering clear lectures, responding effectively to student questions and being available to the students outside of class are the best way to ensure good evaluations. To paraphrase Tip O'Neil, it is probable that all evaluations are local. I am not certain that my results would apply to other institutions or even to other departments at my own College.

The data on which my estimates are based cannot be viewed as the outcome of a controlled experiment. Indeed, I don't want to have to conduct such an experiment on a regular basis. Many studies in the literature have access to information on individual students. I don't have access to individual student data nor do I want access to such data. Like most department chairs in my position, I have access to an imperfect data set. The question is whether I have extracted the maximum amount of meaning out of it. This paper represents a beginning but there is much work yet to be done.

REFERENCES

Becker, William E., and Michael Watts. 1999. "How Departments of Economics Evaluate Teaching." *The American Economic Review,* 89(2): 344-349.

Boex, L. F. Jameson. 2000. "Attributes of Effective Economics Instructors: An Analysis of Student Evaluation." *The Journal of Economic Education,* 31(3): 211-227.

Bosshardt, William, and Michael Watts. 2001. "Comparing Student and Instructor Evaluations of Teaching." *The Journal of Economic Education,* 32(1): 3-17.

Carrell, Scott E., and James E. West. 2010. "Does Professor Quality matter? Evidence from Random Assignment of Students to Professors." *Journal of Political Economy,* 118(3):409-432.

Decanio, Stephen J. 1986. "Student Evaluation of teaching: A Multinomial Logit Approach." *The Journal of Economic Education,* 17(3): 165-176.

Gramlich, Edward, and Glen A. Greenlee. 1993. "Measuring Teaching Performance." *The Journal of Economic Education,* 24(1): 3-13.

Gujarati, Damodar N. 2003. *Basic Econometrics*, 4th edition. New York, NY:McGraw-Hill.

Mehdizadeh, Mostafa. 1990. "Loglinear Models and Student Course Evaluations." *The Journal of Economic Education,* 21(1): 7-21.

Needham, Douglas. 1978. "Student Effort, Learning and Course Evaluations." *The Journal of Economic Education,* 10(1): 35-43.

Nelson, Jon P., and Kathleen A. Lynch. "Grade Inflation, Real Income, Simultaneity, and Teaching Evaluations." *The Journal of Economic Education,* 15(1): 21-37.

Siegfried,Jjohn J., and Rendig Fels. 1979. “Teaching College Economics: A Survey.” *Journal of Economic Literature*, 17(3): 923-969.

Zagaenzadeh, Hamid. 1988. “Grade Inflation: A Way Out.” *The Journal of Economic Education*, 19(3): 217-230.