Munich Personal RePEc Archive

The Effect of Classroom Games on Student Learning and Instructor Evaluations

Cebula, Richard and Toma, Michael

Jacksonville University, Armstrong Atlantic State University

30 November 2000

Online at https://mpra.ub.uni-muenchen.de/53125/
MPRA Paper No. 53125, posted 23 Jan 2014 03:07 UTC

The Effect of Classroom Games on Student Learning and Instructor Evaluations

Richard Cebula and Michael Toma*

Abstract

Assuming that instructors of economics are utility maximizers, they may find it useful to engage in classroom behavior that is likely to generate favorable outcomes with respect to student course evaluations. This is especially true if student course evaluations are used in assessing teaching effectiveness for tenure, promotion, and salary decisions. In this paper, we present evidence that the use of a classroom gaming exercise can raise instructor evaluations and enhance student learning outcomes. The tests are conducted in a framework that indirectly controls for grade inflation and considers student attendance and grade expectations as other sources of influence on instructor evaluation ratings.

Introduction

Recent work regarding instructor evaluation ratings has re-opened the debate on whether the statistical correlation of grades and instructor evaluations can be explained by instructor manipulation of evaluations through the adoption of lenient grading policies (Greenwald and Gillmore 1997, Wilson 1998). Although earlier research found that instructors could 'buy' higher evaluations with easier grading polices, more recent evidence suggests that this influence may be trivial (Greenwald 1997). However, Greenwald and Gillmore (1997) argue that the higher course grades and ratings relationship is most consistent with the theory of lenient grading policies. This prompts Wilson (1998, p. A12) to comment that "professors who want high ratings have learned that they must dumb down material, inflate grades, and keep students entertained."

In this study, we provide evidence that instructor evaluation ratings are improved when a supplemental gaming exercise is introduced into an upper level economics course that neither "dumbs" down the material nor inflates grades but rather does keep students more entertained.

Brief Literature Review

Given the widespread use of instructor evaluations as a tool for assessing teaching effectiveness for tenure, promotion, and salary decisions (White 1995, Becker and Watts 1999), a utility maximizing instructor will engage in behavior that raises evaluation scores. Indeed, a large literature on the subject has developed that explores instructor response to the use of evaluations (for example, see Becker 1975, McKenzie 1975, and Kibbs 1975). The controversial aspect of this literature is that some empirical evidence suggests that instructors attempt to "buy" higher evaluation scores by adopting easier grading policies (Mirus 1973, Nelson and Lynch 1984, Zangenehzadeh 1988). A potential problem of the tests is noted by Seiver (1983) and Stratton, Myers, and King (1994), who point out that these tests cannot differentiate between rising evaluation scores induced by better teaching (and thus, higher grades) on the one hand and or reduced grading standards on the other hand.

Although the use of classroom games in teaching economics was pioneered over fifty years ago by Chamberlin (1948), their use in the modern classroom has only recently grown in frequency (See Holt (1999) and Holt's website at http://www.people.virginia.edu/~cah2k/classy2k.htm for an index of 91 related manuscripts). Fels (1993) observes that this may be a result of high start-up costs and the lack of evidence suggesting that student comprehension of course material is enhanced by using games to illustrate theory. Holt's web-site collection of classroom experiments and games addresses the former issue, our work addresses the latter issue and also considers the effect of using classroom games on student ratings of

* Richard Cebula, Shirley and Philip Solomons Eminent Scholar, Armstrong Atlantic State University, Savannah, GA 31419; Michael Toma, Department of Economics, Armstrong Atlantic State University, Savannah, GA 31419.

instructors.

Given the reported relationship between students' expected grades and evaluation scores, techniques that increase teaching efficiency, as reflected in student comprehension of material, should also result in higher evaluation scores for the instructor (Stratton, et. al. 1984). The empirical evidence on this is limited. Shmanske (1988) finds a weak positive relationship between instructor evaluation scores and student performance, whereas Aigner and Thum (1986) report that facets of teaching style significantly influence evaluation scores but are unable to directly measure teaching effectiveness.

A brief overview of the classroom method is provided in the next section. The data set is described in the following section along with two elementary statistical tests regarding the effect of the portfolio project on student performance and instructor ratings. Next, regression results are reported for whether the use of the game as a teaching technique is associated with higher instructor evaluation ratings. A sampling of open-ended student comments drawn from the evaluation forms is presented as well. The comments qualitatively highlight the difference in student response to the version of the course with or without the portfolio project. The last section is the conclusion.

Classroom Method

Bringing course material to life (entertainment value?) for the student can be accomplished easily in a financial economics course by incorporating an investment portfolio project into course material. The setup is a straightforward version of the Stock Market Game discussed by Wood, O'Hare, and Andrews (1992) and Bell (1993). In the present study, the technique is applied in an upper level economics course, Money and Banking.

Students are required to manage a portfolio of assets and submit monthly updates of portfolio performance, along with an explanation for their investment decisions. The explanation is the most important component of the update, for this is where students are required to apply course material to decision making in financial markets. Grades are not affected by portfolio performance, but by the quality of the rationale behind the investment decisions. As an added incentive, students compete against each other for recognition as the outstanding investment analyst in the class.

Data and Tests of Difference in Means

The data set is drawn from eight Money and Banking courses instructed by one of the authors at two universities populated by similar students. Both universities are state supported commuter campuses set in a suburban location providing residence halls for only a small fraction of total enrollment. Of the eight sections, one section is from each of 1993 and 2001, while two each are from 1995, 1996, and 1998. Total enrollment in the eight sections was 246, from which 172 student evaluation forms were available. The portfolio project sample (189 students, 121 forms available) consists of both 1995 and 1996 sections as well as one section each from 1998 and 2001. The non-portfolio project sample (57 students, 52 forms available) consists of one section each from 1993 and 1998.

The student evaluation forms from both universities included comparable information regarding whether the course was well taught, organized, motivating, challenging, and whether the instructor was prepared and unbiased. Although one form used a four point scale (strongly agree, agree, disagree, strongly disagree) while the other included a fifth 'undecided' category between agree and disagree, this would tend to work against the results presented below. First, the problem is partially mitigated by having one section each from the university using the four point scale form in the portfolio and non-portfolio samples. Second, given that a majority of the ratings from the five point scale form were' agree' or better, it can be inferred that the lack of a middle 'undecided' category would tend to bias the ratings from the four point scale forms in the upward (better) direction toward the 'agree' response. Given the smaller size of the non-portfolio sample, the upward bias in the sample average ratings would be larger relative to effect of the bias in the larger portfolio sample. Finally, given the structure of the test, the upward bias in the non-portfolio sample would tend to decrease the likelihood of observing a statistically significant difference in mean ratings across the two samples.

In the remaining portion of this section, we present the results of two elementary tests designed to evaluate whether the supplemental game resulted in enhanced student learning and increased ratings of the course and instructor. The first test focuses on student comprehension as measured by the percentage of total points possible that each student earned. The only difference in the way students earned grade credit across the samples is that students in the portfolio sections also earned points associated with the project while students in the non-portfolio sections only earned points by taking periodic examinations. The examinations consisted of a multiple choice, short answer, and essay questions. The examinations were not standardized across the different sections, but question topics were similar in depth, breadth, and difficulty.

The grades from each student population (all portfolio and non-portfolio sections) are assumed to have a normal distribution as is the standard error of the differences in means. The test of difference in means across the two samples is conducted using the z-statistic. The data and test result are found in Panel A of Table 1.

The average score in the two samples is significantly different. In sections with the portfolio project, the average score was 69.3 percent, while in the sections without the supplemental exercise the average was 53.3 percent. The t-statistic is 7.49, indicating that the difference is significant with 99 percent confidence. This result suggests that student learning outcomes were significantly enhanced by including a supplemental exercise in the course and is consistent with the findings of Fels (1984) and Mehdizadeh (1990).

Table 1. Effect of Portfolio Project

Panel A. Test of Change in Average Student Scores

	Sections without Project	Sections with Project	Test Z-Statistic
Average Student Score (percent)	53.3	69.3	7.49
Standard Deviation (percent)	14.0	14.4	
Number of Students	57	189	

Panel B. Test of Change in Average Evaluation Rating

	Average Evaluation Rating		
Evaluation Question Topic	Sections without Project	Sections with Project	Test Z-Statistic
Class is Well Taught	3.82	4.14	1.96
Organization of Material	3.92	4.31	2.29
Motivated Students	3.75	4.06	1.96
Instructor Preparation	4.46	4.65	1.81
Challenged Students	3.94	4.19	1.97
Unbiased Treatment of Students	4.51	4.58	0.61
Number of Students	52	121	

Notes: Instructor ratings range from one to five; five is the highest score; The critical value for the test statistic is 2.58 at the 99 percent confidence level and 1.96 at the 95 percent confidence level.

Next, we consider the effect on student evaluation ratings of instructor performance when the investment project is included in the course. Prior to commenting on these tests, however, a caveat should be noted. Given the data set, a potential problem is that the evaluation scores could be expected to rise through time as instructor experience increases. Although we are unable to directly control for this phenomenon in these tests, the combination of 1998 data with 1993 data in the non-portfolio sample partially mitigates this problem. In addition, the evaluation question topics may shed some light on this issue. It seems reasonable that greater experience would most likely be reflected in the categories for "instructor preparation" and "course organization" and, to a lesser extent, in the "well taught" ratings.

We hypothesize that the evaluation components most clearly reflecting active student learning would be those regarding student "Motivation" and whether the course was "Challenging." Increased ratings on whether the instructor motivated students could arise from the relevance of the material with respect to financial markets. The aspect of whether the course was challenging would arise as students recognize the difficulty in application of theory to financial market activity. It also seems reasonable to expect an increase in "Well Taught" ratings as student interest and appreciation of the material is heightened through completion of the portfolio project. Whether an instructor is "Unbiased" would least likely be effected by increased instructor experience or the addition of the portfolio project.

In Table 1, average instructor evaluation ratings for six different topic areas are presented. The scores range from one to five, with a rating of five reflecting the highest level of student satisfaction with the instructor. The results from the difference in means test for the ratings across the portfolio and non-portfolio samples are found in Panel B of Table 1.

The statistical results are consistent with expectations. The portfolio project was designed to stimulate interest in the course material. Thus, we expected to observe increases in the "motivation" and "challenging" ratings since students had the opportunity to apply what they learned in the classroom to decision-making in actual financial markets. The z-statistics for these categories indicate a significant increase in the average rating at the 95 percent confidence level. It was also hypothesized that an increase in the "well taught" rating could result from the addition of the active student learning exercise (the investment project). The z-statistic for this question topic indicates (with 95 percent confidence) an increase in average rating score in sections that included the portfolio project.

The question topics that would most likely reflect increased experience are "preparation" and "organization," although "organization" could also be influenced by student appreciation of the way the portfolio project is integrated into the course. The rating for "preparation" is not significantly different. The "organization" rating displays a significant increase (95 percent confidence) when the portfolio project is part of the course. Zangenehzadeh (1988) finds that that instructor preparation and organization significantly influence student assessment of course quality. Thus, the ratings in these two categories could be influenced by the dual influence of increasing experience and the introduction of the portfolio project.

Also confirming prior expectations are the evaluation ratings for "unbiased" which show no significant increase. It seems unlikely that a hands-on investment project would increase students' perception of whether or not the students believed their work was impartially graded.

Regression Results

In this section we present regression results from a sub-sample of the data set for which additional student demographic information and grade expectations information are available. The data are available from 151 evaluation forms from six sections of the course (one from 1993, two each from 1995 and 1996, and one from 2001). A data summary is presented in Table 2.

In the model, the instructor rating is regressed on a set of binary variables characterizing the effects of grade expectations, number of absences, and whether the portfolio project was included in the course. The general specification is as follows:

$$\text{Rating} = \beta_0 + \beta_1(\text{Bexp}) + \beta_2(\text{Cdexp}) + \beta_3(\text{Absent4}) + \beta_4(\text{Port}) + \varepsilon \quad (1)$$

Rating is the student rating of the instructor for each of several characteristics about quality of instruction and course management. More specifically, students rated whether the course material was presented in an organized manner and was challenging. In addition, student assessed instructor preparation, motivation, bias or lack thereof toward students, and a general assessment of whether the course was taught well. The error term in the regression, ε, is assumed to adhere to the standard assumptions of the classical linear regression model. It is a white noise error term.

The student demographic variables considered first are student expectations about their grade in the course and self-reported student absences. Course evaluations were distributed during the last two weeks of the course before the final exam and before the portfolio project term paper was submitted. Student grade expectations are modeled as dummy variables where, for example, if a student expected an 'B' grade, *Bexp* = 1; *Bexp* = 0 otherwise. Since only two students in the sample were expecting a 'D' grade, these students are modeled jointly with students expecting a 'C' grade by the variable *Cdexp*. *Absent4* is a dummy variable that equals one when a student self-reported four or more class absences, and zero otherwise.

The constant in the regression, β_0, jointly characterizes students who reported three or less absences with students who expected to receive an 'A' grade in the course. This specification for the constant is chosen in light of the collinearity between high grade expectations and few absences, and Durden and Ellis' (1995) result that relatively low levels of class absence do not significantly affect student learning.

Table 2. Summary Statistics for Regression Sample

Variable Name	Variable Definition	Mean	Standard Deviation	N
Taught	Taught well rating	4.05	1.03	150
Org	Organized rating	4.23	0.85	151
Mot	Motivation rating	3.89	0.96	150
Prep	Preparation rating	4.50	0.76	151
Chal	Challenging rating	4.01	0.88	150
Unb	Unbiased rating	4.48	0.77	146
Port	Dummy for portfolio project	0.80	0.40	151
Pab	Interactive, Port x Abexp	0.58	0.50	150
Pcd	Interactive, Port x Cdexp	0.23	0.42	150
Aexp	'A' grade expected	0.28	0.45	150
Bexp	'B' grade expected	0.43	0.50	150
Cdexp	'C' and 'D' grade expected	0.29	0.45	150
Absent4	Absent more than 4 times	0.13	0.34	149

Notes: N equals number of observations.

Port is a dummy variable equaling one if the student was in a section of the course that included the portfolio project, and zero if not. Also considered in an auxiliary regression is an interactive effect between *Port* and students grouped by grade expectations. One group of students considered is those who expected 'A' and 'B' grades. This group is designated by the variable *Abexp*, which equals *Aexp* + *Bexp*. The second group is students who expected 'C' and 'D' grades (*Cdexp*). The interactive variables are defined as *Pab*, which equals (*Port* x *Abexp*), and *Pcd*, which equals (*Port* x *Cdexp*). These interactive variables allow for the identification of different effects of the portfolio on ratings by grade expectations depending on whether the student believed he or she was in the top or bottom half of the grade distribution.[1]

Since the constant of the regression characterizes students who regularly attended classroom lectures and expected an 'A' grade, the partial derivatives of the *Rating* variables with respect to *Bexp* and *Cdexp* are expected to be negative. Students expecting lower grades have been shown to penalize their instructors with lower ratings (Stumpf and Freedman 1979). In addition, attribution theory (Greenwald 1980, Gigliotti and Buchtel 1990, and Theall, Franklin, and Ludlow 1990) suggests that while students reward themselves with the credit for performing well in a course, they blame the instructor when their performance is sub-par.

If it is reasonably assumed that students who do not attend class regularly tend to perform poorly (Park and Kerr 1990, Romer 1993, and Durden and Ellis 1995), then the partial of the *Rating* variables with respect to *Absent4* is expected to be negative. The lack of attendance results in sub-par performance, is attributed to the instructor, and reflected in lower ratings.

The use of supplemental exercises has been found to increase instructor evaluations ratings (Fels 1984, Mehdizadeh 1990), and perhaps explains the growing interest in, and compilation of, resources about classroom games (Holt 1999 and Holt's website cited above). Thus, the portfolio dummy (*Port*) and its interactive forms are expected to be positively related to *Rating*, although the effect may vary by grade expectation.

The ordinary least squares results for Equation 1 are reported in Table 3. The results for each rating are reported in the columns of the table. The entries are the estimated coefficients with the t-statistics reported in parentheses below the coefficient. White's (1980) test for heterscedasticity in the residuals indicates that the null of homoscedastic errors cannot be rejected.

The presence of the portfolio project in the course is associated with significantly higher evaluation

[1] No students self-reported as expecting to fail the class.

ratings across nearly all evaluation categories. With respect to the 'taught well' and 'motivates' ratings, *Port* generates an increase in *Taught* by two-thirds of a point (17 percent increase) while *Mot* increases by one-half of a point (13% increase). In addition, *Chal* increased by 11 percent when the portfolio project was included in the course. There was no significant effect on whether students perceived their work was evaluated in an unbiased manner. These results, based on data from six sections of the course, are generally consistent with the z-statistics reported for the full sample (eight sections) in the previous section.

Table 3. Influences on Instructor Ratings I

Independent Variable	Dependent Variable					
	Taught	Org	Mot	Prep	Chal	Unb
Constant	3.90**	4.25**	3.90**	4.43**	3.94**	4.50**
	(17.29)	(23.05)	(18.50)	(27.33)	(19.79)	(25.24)
Bexp	-0.29	-0.22	-0.42*	-0.18	-0.34*	-0.01
	(-1.56)	(-1.41)	(2.31)	(-1.30)	(-2.01)	(-0.07)
Cdexp	-0.60**	-0.28	-0.51*	-0.12	-0.32	-0.43*
	(-2.82)	(-1.57)	(-2.52)	(-0.79)	(-1.66)	(-2.48)
Absent4	-0.66**	-0.82**	-0.56*	-0.51**	-0.33	-0.03
	(-2.71)	(-4.03)	(-2.39)	(-2.86)	(-1.51)	(-0.14)
Port	0.68**	0.33*	0.50**	0.34*	0.45*	0.14
	(3.32)	(1.96)	(2.60)	(2.31)	(2.47)	(0.83)
R-square	0.17	0.15	0.13	0.09	0.08	0.07
Adj. R-Square	0.15	0.12	0.10	0.06	0.06	0.04
F-stat (5,n-5)	7.34	6.09	5.31	3.48	3.21	2.54
White's test	9.94	10.10	9.29	4.40	8.61	6.86
N	148	149	148	149	148	144

Notes: N equals number of observations; ** indicates a rejection of the null hypothesis at 99 percent confidence level; * indicates a rejection of the null hypothesis at 95 percent confidence level. The critical value for the f-statistic with 6 and n-6 degrees of freedom is 2.21 at the 95 percent confidence level and 3.02 at the 99 percent confidence level. The critical χ^2 value for White's test with 8 degrees of freedom at the 90 percent confidence level is 13.36.

Students' expectations for a high grade also significantly influenced whether students rated the instructor highly. Recall that the constant in the regression characterizes the rating assigned by students who expected an 'A' in the course and who missed three or less classes. These "better" students, who serve as the baseline of comparison in the model, rated the instructor at approximately 3.9 to 4.5 across the *Rating* categories. Although the variables characterizing lower grade expectations (*Bexp* and *Cdexp*) are not significant across all specifications at conventional levels, they do have the expected sign. Moreover, *Cdexp* is significant at the 99 percent confidence level in the regression for the most general assessment of the course, whether it was taught well. Students expecting 'C' or 'D' grades offered an overall assessment that was 17 percent lower (six-tenths of a point) than their classroom peers expecting better grades.

In addition, the results of this analysis suggest that relatively high levels of student absence were associated with a significant reduction in instructor rating. As reported in Table 3, the coefficient for *Absent4* indicates that the average rating for *Taught, Org, Mot,* and *Prep* was significantly reduced by between 12 and 19 percent. To the extent that greater absence is associated with lower levels of student performance, the result is consistent with attribution theory discussed above. These students may have attributed their attendance-related poor performance to a 'poor' instructor and reduced their rating accordingly.

The results of a supplemental regression exercise are reported in Table 4. In these models, the effect of the portfolio project is allowed to vary by the students' grade expectations rather than being constrained to have an equal effect across all students regardless of their grade expectations. The results are generally consistent with the constrained model, although the portfolio project loses its effectiveness in raising instructor ratings among students who expected to receive a grade in the lower half of the grade distribution.

The portfolio's effect on instructor rating is amplified among students expecting better grades. On average, the overall assessment of the course (*Taught*) rises by 24 percent among students expecting an 'A' grade in sections with the portfolio project as compared to 17 percent among all students expecting a top grade. An f-test restricting the estimated coefficients of *Pab* and *Pcd* in this specification for 'well taught' rejects the null of equal effects at the 89 percent confidence level. Although this does not meet conventional criteria for significance, it is suggestive that the effect of the portfolio may vary by whether students expect a grade in the upper or lower half of the grade distribution.

Table 4. Influences on Instructor Ratings II

Independent Variable	Dependent Variable					
	Taught	Org	Mot	Prep	Chal	Unb
Constant	3.71**	4.42**	3.77**	4.41**	3.82**	4.53**
	(14.78)	(20.58)	(16.09)	(24.40)	(17.25)	(22.70)
Bexp	-0.28	-0.22	-0.41*	-0.18	-0.33*	-0.01
	(-1.49)	(-1.40)	(-2.28)	(-1.29)	(-1.97)	(-0.08)
Cdexp	-0.05	-0.24	-0.10	-0.08	0.07	-0.53
	(-0.13)	(-0.73)	(-0.26)	(-0.28)	(0.19)	(-1.65)
Absent4	-0.59**	-0.82**	-0.51*	-0.51**	-0.29	-0.04
	(-2.44)	(-3.95)	(-2.17)	(-2.79)	(-1.29)	(-0.20)
Pab	0.88**	0.34	0.65**	0.36*	0.59**	0.10
	(3.68)	(1.72)	(2.88)	(2.05)	(2.78)	(0.49)
Pcd	0.19	0.30	0.13	0.31	0.09	0.23
	(0.51)	(0.96)	(0.36)	(1.12)	(0.28)	(0.79)
R-square	0.19	0.14	0.14	0.08	0.09	0.07
Adj. R-Square	0.16	0.12	0.11	0.06	0.06	0.04
F-stat (6,N-6)	6.46	4.85	4.59	2.77	2.90	2.05
White's test	10.58	10.10	8.84	4.44	8.71	6.95
N	148	149	148	149	148	144

Notes: N equals number of observations; ** indicates a rejection of the null hypothesis at 99 percent confidence level; * indicates a rejection of the null hypothesis at 95 percent confidence level. The critical value for the f-statistic with 6 and n-6 degrees of freedom is 2.10 at the 95 percent confidence level and 2.80 at the 99 percent confidence level. The critical χ^2 value for White's test with 8 degrees of freedom at the 90 percent confidence level is 13.36.

Qualitative Evidence: Testimonials

In addition to coding a rating for the categories on the instructor evaluation form, students from all sections had the opportunity to provide open-ended written comments about the course. These comments may serve as additional qualitative evidence supporting the statistical results. A sampling of typical comments obtained from the students' instructor evaluation sheets from courses not including the portfolio project (Without Portfolio) and from courses including the project (With Portfolio) are presented next.

Typical Quotes: Without Portfolio

"Course was strictly lecture. Not sure what other materials are available, but may want to consider using them."

"I found [that] this course is very important for economic knowledge, but felt lost in the process."

These and other quotes indicate that students were less interested in the material and may have been overwhelmed by the course. The difference in tone and content is clear when these comments are contrasted with those made by students in the sections which completed the investment project.

Typical Quotes: With Portfolio

"The mock portfolio was a great idea. It was a good way to get students interested in the class and demonstrated the theories we learned."

"I enjoyed doing the portfolio project because I was able to apply what we were learning in this class to actual life."

"I think that the stock market game is a great idea to make the class more interesting. There is no way of getting around the fact that some of the information learned in this class is boring [?!] but the game makes it all worth it."

"I have enjoyed the fact that we get to practice what we learn in class through our continuous portfolio management project. I have definitely enjoyed learning to read the *Wall Street Journal*. It's important that you have encouraged us to relate class material to the real world."

"This is the most useful block of college instructing I have ever received. If I had access to this information thirty years ago as a college undergraduate, I would not now be starting a second career after retirement from the military."

Conclusion

The results obtained in this study shed light on the issue of whether the use of classroom games has an impact on student learning. This provides some important evidence with respect to Fels' (1993, p. 365) concern that the cost of classroom game implementation could exceed the "meager evidence of [their] educational value." A test is conducted in which, ceteris paribus, student examination scores were observed to increase in the sample that included a financial market portfolio project. Holding lecture style and grading policy fixed across the two samples and examining instructor ratings sheds light on the issue of whether student learning, enhanced by the portfolio project, influences evaluation scores.

The results in the present analysis suggest that students comprehended more of the material, and assigned higher scores to their evaluation of the instructor. Furthermore, the analysis controlled for student grade expectations, thus permitting the identification of the effect of a supplemental gaming exercise on instructor ratings. This method avoids the issue raised by Seiver (1983), Aigner and Thum (1986) and Zangenehzadeh (1988) who demonstrate that students with higher grade expectations rate instructors more favorably. Finally, these results were obtained using a test method that indirectly controlled for the effect of grade inflation and grading policy. Thus, the higher evaluation scores do not simply reflect a "selling out" to students in an effort to raise evaluation ratings.

It is worthwhile to note that very little other course material was altered or "crowded out" as a result of including the investment project. While the course material was little changed, student reaction was more favorable and student performance was enhanced in sections including the portfolio project. In addition, instructor evaluation ratings significantly increased in the presence of the supplemental exercise. Thus, we provide evidence that classroom gaming exercises may increase student performance and raise instructor evaluation ratings. These findings are tentatively attributed to the fact that it became readily apparent to the students how useful the classroom material could be in the "real world."

References

Aigner, Dennis, and Frederick Thum. 1986. "On Student Evaluation of Teaching Ability." *Journal of Economic Education* 17(4): 243-265.

Becker, William. 1975. "The University Professor as a Utility Maximizer and Producer of Learning, Research and Income." *Journal of Human Resources* 10(1): 107-15.

Becker, William, and Michael Watts. 1999. "How Departments of Economics Evaluate Teaching." *American Economic Review Papers and Proceedings* 89(2): 344-49.

Bell, Christopher. 1993. "A Noncomputerized Version of the William and Walker Stock Market Experiment in a Finance Course." *Journal of Economic Education* 24(4): 317-323.

Chamberlin, Edward. 1948. "An Experimental Imperfect Market." *Journal of Political Economy* 56(2): 95-108.

Durden, Gary, and Larry Ellis. 1995. "The Effects of Attendance on Student Learning in Principles of Economics." *American Economic Review Papers and Proceedings* 85: 343 –46.

Fels, Rendigs. 1984. "Student Papers on Macroeconomic Policy." *Journal of Economic Education* 15(3): 237-38.

______. 1993. "This is What I Do and I Like It." *Journal of Economic Education* 24(4): 365-70.

Gigliotti, Richard, and Foster Buchtel. 1990. "Attribution Bias and Course Evaluations." *Journal of Educational Psychology* 82(2): 341-351.

Greenwald, Anthony. 1980. "The Totalitarian Ego: Fabrication and Revision of Personal History." *American Psychologist* 35: 603-618.

______. 1997. "Validity Concerns and Usefulness of Student Ratings of Instruction." *American Psychologist* 52: 1182-1186.

______, and Gerald Gillmore. 1997. "Grading Leniency is a Removable Contaminant of Student Ratings." *American Psychologist* 52: 1209-1217.

Holt, Charles. 1999. "Teaching Economics with Classroom Experiments." *Southern Economic Journal* 65(3): 603-10.

Kibbs, Paul. 1975. "The Use of Course Evaluation Scores to Influence Teaching and Research Activities." *Journal of Economic Education* 6(2): 93-98.

McKenzie, Richard. 1975. "The Economic Effects of Grade Inflation on Instructor Evaluations." *Journal of Economic Education* 6(2): 99-106.

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

Mirus, Rolf. 1973. "Some Implications of Student Evaluation of Teachers." *Journal of Economic Education* 5(1): 35-46.

Nelson, Jon, and Kathleen Lynch. 1984. "Grade Inflation, Real Income, Simultaneity, and Teaching Evaluations." *Journal of Economic Education* 15(1): 21-37.

Park, Kang, and Peter Kerr. 1990. "Determinants of Academic Performance: A Multinomial Logit Approach." *Journal of Economic Education* 21(Spring): 101-11.

Romer, David 1993. "Do Students Go to Class? Should They?" *Journal of Economic Perspectives* 7: 167-74

Seiver, Daniel. 1983. "Evaluation and Grades: A Simultaneous Framework." *Journal of Economic Education* 14(3): 32-38.

Shmanske, Stephen. 1988. "On the Measurement of Teacher Effectiveness." *Journal of Economic Education* 19(4): 307- 314.

Stratton, Richard, Steven Myers, and Randall King. 1994. "Faculty Behavior, Grades and Student Evaluations." *Journal of Economic Education* 25(1): 5-15.

Stumpf, Steven, and Richard Freedman. 1979. "Expected Grade Covariation with Student Ratings of Instruction: Individual and Class Effects." *Journal of Educational Psychology* 71: 293-302.

Theall, Michael, Jennifer Franklin, and Larry Ludlow. 1990. "Attributions and Retributions: Student Ratings and Perceived Causes of Performance." *Instructional Evaluation* 11: 12-17.

White, Halbert 1980. "A Heteroscedasticity-Consistent Covariance Matrix and a Direct Test for Heteroscedasticity." *Econometrica* 50: 1-26.

White, Lawrence. 1995. "Efforts by Departments of Economics to Assess Teaching Effectiveness: Results of an Informal Survey." *Journal of Economic Education* 26(1): 81-85.

Wilson, Robin. 1998. "New Research Casts Doubt on Value of Student Evaluations of Professors." *Chronicle of Higher Education* 44(19), 16 January: A12.

Wood, William, Sharon O'Hare, and Robert Andrews. 1992. The stock market game: Classroom use and strategy. *Journal of Economic Education* 23(3): 234-46.

Zangenehzadeh, Hamid. 1988. Grade inflation: a way out. *Journal of Economic Education* 19(3): 217-26.