Using Cooperative Learning to Improve Student Understanding of Exam Evaluation

Jill L. Caviglia-Harris
Economics and Finance Department
Salisbury University
Salisbury, MD 21804-6860
Phone: (410) 548-5591
Fax: (410) 546-6208
jlcaviglia-harris@salisbury.edu

January 2007

PRELIMINARY DRAFT: PLEASE DO NOT QUOTE

Using Cooperative Learning to Improve Student Understanding of Exam Evaluation

Abstract

This paper describes the design and administration of a cooperative learning exercise with the multiple goals of assisting students with exam preparation, providing a better understanding of the grading process and enabling students to critically evaluate essay questions and answers. Lessons from the administration in the first semester are described and used to improve the use and design of the exercise in the following semester. Overall students reported benefits of the exercise that include a better understanding of the grading process and recommended its use in following semesters.

Using Cooperative Learning to Improve the Understanding of Exam Evaluation

Introduction

Most students do not have a comprehensive understanding of the processes and criteria used to evaluate their work, in particular for the evaluation of essay questions and other writing assignments. This generally results from an expected lack of experience grading as well as a limited knowledge of the subject matter. Although this level of understanding may not be included in the course goals for many classes, if students were given the opportunity to better understand the development of criteria used to evaluate their writing, this may enable them to improve their writing skills, guide their studying, and at a minimum help them to better understand the criteria used to evaluate their work.

It is well known that students generally have inflated views of their expected grades (Grimes 2002, Hacker et al. 2000, Balch 1992). This may stem from overstated views of their potential, but may also stem from a lack of understanding of exam and assignment evaluation. A greater understanding of the grading process may therefore assist students in better predicting their expected exam and course grades. Even though student expectations of course grades have been found to be inflated, the literature is inundated with studies that use expected grades, rather than actual course grades, to estimate student performance (Ballard and Johnson 2005, Cebula and Toma 2002), empirically evaluate different teaching techniques (Lage et al. 2000), assess student interest in economics (Jensen and Owen 2001) investigate instructor effects (Finegan and Siegfried 1999), and to study the evaluation of instructors by students (McPherson 2006, Bosshardt and Watts 2001, Krautmann and Sander 1999), among others. The reliance on this measurement of student performance is due to the availability of these data, since most teaching

evaluations generally include several questions about the student and the expected grade in the course. These data are not necessarily a poor substitute for actual grades in the analysis if they do not bias estimations. However, an additional related issue is that student misperceptions may influence or hinder studying habits and approaches.

This paper explains the development and use of a cooperative learning exercise to assist students with writing answers to essay questions and provide them with a better understanding of the evaluation criteria used in grading. The goals of the exercise are two-fold: 1) for students to better understand how to answer essay questions and identify the important theories to draw from when formulating an answer and 2) for students to better understand their potential exam grade and prepare accordingly. The paper describes the development and administration of the exercise that was administered in class two different semesters. Lessons from the administration in the first semester are described and used to improve the use and design of the exercise in the following semester.

Overconfidence

One form of discontent in the classroom that many instructors face results from unmet expectations of students on exams. Several studies have cited empirical evidence of inflated grade expectations by students (Hacker et al. 2000, Grimes 2002). This form of overconfidence is hypothesized here to stem from some combination of three sources: 1) students have received similar grades in previous courses; 2) students believe (or hope) that their performance is better than average; 3) students lack an awareness of their own knowledge deficits and therefore do not have the ability to understand the grading and evaluation process.

The first explanation for overconfidence listed above may stem from experiences in previous courses. For example, a study of grade distribution by department at Salisbury University for the years 1998-2002 suggests that this may be the case, at least at the local level.[1] Of the 45 major disciplines offered at the university, students taking economics courses received the lowest percentage of As and Bs for all but one year, with percentages ranging from 34-49% of enrolled students. In 1998, only geology students received a lower percentage of As and Bs.[2] It is therefore likely that students may be accustomed to receiving higher grades on exams in fields outside of economics. Students may therefore employ a strategy known as the "availability heuristic" in which they make judgments based on imperfect information or the information that is available at the time of the decision (McKelvie 1997). According to Tversky and Kahneman (1973) people base their judgments on the ease with which events can be recalled. If students can recall recent above average grades, they are more likely to predict future occurrences. This hypothesis is tested by estimating predicted grades with data on student characteristics including GPA prior to taking the course.

Also likely to play a role in the overestimation of exam performance is the general condition of overconfidence of own ability. There is well documented evidence that most individuals believe that they perform better than average (Alicke et al. 1995, Chambers and Windschitl 2004, Schaefer et al. 2004, Smith 2005). This form of overconfidence, cited as the "above average effect" (Fang and Moscarini 2005) has several sources of evidence. In a review of research on performance in the work place, Meyer (1975) finds consistent evidence that between 58 and 83 percent of workers rated themselves within the top 10 percent and less than 5 percent rated themselves below average. In Svenson (1981) 92.8% of a sample of American

[1] Salisbury University internal report, Office of Institutional Research.
[2] Interestingly, over the study period the remaining disciplines that also distributed the lowest percentage of As and Bs were all in the sciences: chemistry, computer science, geography, geology, physics and mathematics.

students reported themselves as safer than the average driver (c.f. Fang and Moscarini 2005). Soll and Klayman (2004) find judges to overestimate their knowledge on a common set of questions. Finally, Chambers and Windschitl (2004) review studies finding a majority of individuals to report being less susceptible than average to heart attack, attempted suicide, being unemployed, becoming the victim of a crime, and having an unwanted pregnancy. Taking this further, the "self-serving bias" is the tendency of individuals to attribute their successes to positive personal traits, but their failure to situations out of their control. In a classroom setting this thought process can lead to students to attribute good grades to skill, intelligence and study habits, but poor grades to external factors such as bad instructors or "tricky' exam questions (Hacker et al. 2000).

The third reason cited to partially explain overconfidence is the inability of students (or non-experts) to evaluate their own performance. According to several studies this applies to a greater degree to lower achieving students (Krueger and Mueller 2002, Balch 1992). Many of these studies find better performing students to predict their exam grades with greater accuracy and conclude that higher achieving students understand the material and their own limits, better direct their studying, and therefore exhibit greater ability to predict their grades. However, it is also just as likely that these studies are capturing the "better than average effect". Independent of academic ability, students in general believe (or hope) that they will perform better than average.

This paper identifies student overconfidence on exam and quiz grades with a comparison of data on expected grades before and after each of the exams to actual grades. Several issues are investigated including changes in expected grades pre and post exam and changes in

expected grades occurring over the course of the semester. In addition, any impacts on expected grades of the cooperative learning exercise are presented.

Cooperative Learning

Cooperative learning is the instructional use of structured group learning in the classroom requiring purposeful implementation. Cooperative learning requires students to use acquired knowledge in a group setting with a structured learning task (Cameron 1998). Often cited as PIES, the key elements of this instructional method are positive interdependence (individual and group success are positively correlated), individual accountability (individual performances are evaluated), equal participation (all participants have an opportunity to contribute), and simultaneous interaction (occurs when more than one participant is involved at the same time) (Bartlett 1998, McGoldrick 2005). Much of the research on the psychology of learning suggests that students learn better when they are involved and have the ability to respond in the classroom, and that the redundancy introduced with the visual aspects of interactive teaching can assist in recollection of specific ideas and concepts (Saunders 1998).

Cooperative learning exercises can be divided into five broad categories including discussion, reciprocal teaching, graphic organizers, writing, and problem solving (McGoldrick 2005). Within each of these categories there are endless methods for which cooperative learning can take place. This exercise involves elements of discussion, writing, and problem solving. Specifically, the exercise draws from the Think/Pair/Share method of discussion, the Peer Editing method of writing and the Structured Problem Solving methods of problem solving. Think/Pair/Share exercises require students to think independently about a problem then pair up with another student to share answers. The Peer Editing method instructs students to write an

essay, exchange drafts, edit their partner's work and then discuss differences and similarities as a group. Structured Problem Solving involves the development of a series of problem solving steps to solve a particular problem (see McGoldrick 2005 for further details).

The three basic types of cooperative learning groups include formal, informal, and base (Bartlett 1998). Formal groups are those that are selected by the instructor to complete an assignment of a task involving meetings inside and/or outside of the classroom. Informal groups are those are that assigned or chosen by students to complete a short in class assignment. Finally base groups are those that are chosen either by the instructor or students for the duration of the course. Several studies have found cooperative learning to work best with heterogeneous groups of students (Maier and Keenan 1994), however more recent studies have found no statistical difference in class performance when comparing heterogeneous and homogenous groups (Maloof and White 2005). Therefore, for this exercise various approaches for pairing were employed. In the first exercise, students were paired with neighbors, in the next two exercises students were paired randomly by numbers called out by the instructor prior to beginning the exercise; finally in the last exercise (the only one used in the second semester) students were paired by ability. Students that performed well were paired with students that did poorly (i.e. the best performing student was paired with the worst, the second best with the second worst, etc).

The Essay Evaluation Exercise

This exercise addresses the key elements of cooperative learning through design, and can be broken down into the following steps:

1) Students take an essay quiz.
2) This quiz is photocopied, the original is graded.
3) The copied quiz is returned to students (ungraded).
4) Students evaluate the quiz answers in pairs using quiz answers, notes, and books.

5) The answers are discussed as a class.
6) Student determine weights they would use to evaluate the quiz.
7) The instructor discusses the weights used in grading and returns the graded quizzes with a key.

Positive interdependence can be facilitated by creating a common goal, requiring students to share resources and assigning complementary roles (Bartlett 1998). In this exercise the development of the weighting factors used in the student grading process creates a group goal requiring students to pool their ideas. Individual accountability is an important aspect of collaborative learning; the intent of which is to create incentives for individual participation by all students. This aspect is accomplished in this exercise through instructor grading of the quizzes. Equal participation is encouraged by allocating time to the development of individual quiz answers. Finally, simultaneous interaction occurs as student groups develop the correct quiz answers and weighting factors for grading. Specifically, the exercise requires many students to work at the same time instead of one student answering a pointed question as is more typical in larger class discussions. Also important to cooperative learning is reporting out and closure. Reporting out occurs when students discuss the problem solution orally or in writing. The purpose is to recap the exercise and reinforce positive interdependence. In this exercise reporting out occurs with the class discussion of the essay answer and weighting factors while closure is provided with the return of the quizzes and any discussion prompted by the grading comparison.

In more detail, the essay evaluation exercise involves the distribution of an in class essay quiz to be evaluated by a class peer and the instructor. The quiz questions require students to apply knowledge and economic theory discussed in lecture (and covered in the course textbook) in a situation that they have not yet addressed. This exercise was first administered three times

throughout the semester in a micro economics principles course of approximately 35 students. The second time this exercise was administered only once in a micro economics principles class of similar size by the same instructor in the following semester.

There are two types of grading methods appropriate to essay questions: analytical and holistic. The analytical method requires a key (or rubric) that is developed detailing points for separate components of a question. This method of evaluation can be used to provide detailed feedback to students. The holistic method is the evaluation of an essay from an overall perspective and therefore depends on expert judgment (Walstad 2006). The analytical method is applied in the cooperative learning exercise since the holistic approach requires expertise not available to students. Even so, it is likely that students use the holistic approach to evaluate their own work, but since they are not experts in field or in the evaluation process, are more likely to misestimate grades. Thus, one goal of this exercise is to move students from thinking in a holistic manner to being more analytical in their assessment; essentially providing them with tools that can help improve and guide responses to open ended and other essay type questions.

An endless number of essay questions can be used for this exercise. For explanatory purposes the questions prepared for the first such class exercise will be discussed briefly. Students were asked to answer the following questions:

> 1) Explain what occurs in the market for fish if supply increases ***and*** demand decreases.
> a) Show these changes in a graph below and explain them in words.
> b) Which change, the change in equilibrium quantity or equilibrium price is ambiguous, which is not? Explain your answer.
>
> 2) Consider a public policy aimed at smoking. Studies indicate that the price elasticity of demand for cigarettes is about 0.2. If the government would like to reduce smoking by 10 percent, how much should it increase the price?

Two classes prior to exam day, a quiz containing the two-part essay question above (or similar) was distributed to students. Once completed, the essay question answers were

photocopied by the instructor. One copy was graded while the second was distributed to students the following class period.

To begin the exercise, students were asked to work in pairs (neighbor, random, and skill paring methods were experimented with over the four exercises administered). The photocopied quizzes were distributed to the class. Students were instructed to prepare an answer key on what they believed to be the correct answers using their own knowledge, quiz answers, notes, and books. Students were given approximately 10 minutes to complete this task. Once this step was completed, the correct answers were discussed as a class for approximately 5 minutes. After this review, the groups were given approximately 5 minutes to assign weights for the 2 questions adding to a total of 10. The points assigned to each of the two questions were then further divided according to student choice. Throughout the course of the exercise, the instructor monitored the progress of student groups and kept time. Any groups that completed the assignment early were asked to join other groups that completed the task and evaluate similarities and differences between the group essay answers and weights. Lastly, students were asked to grade their partner's quiz based on these weights. (Students were instructed that this would not be the grade that would be recorded). The instructor-graded quizzes were then returned to students along with a key. Class discussion was used to evaluate and compare the group determined weights to the weights as listed on the answer key.

One of the more important aspects of the exercise is the provision of an answer key with specific answer weights. Explaining the development of these weights could serve as an important guide to students. Once the quizzes were returned, it was explained that students would not necessarily receive points for correct answers (such as for question 1 above) but rather for providing evidence of understanding. It was also explained that the more complex topics

were given higher weights. For the questions above, the weighting was determined as follows. Question 1 was worth 8 points, while question 2 was worth 2. In question 1, the graph of the shifts in supply and demand was worth 2 points. Full credit was given for graph that included both these shifts correctly with labeled curves and changes in price and quantity that were labeled on the axes. The explanation of these shifts, including the impact on price and quantity was worth 3 points (1.5 points for the explanation of the price change, 1.5 points for the explanation of the quantity change). Lastly, the explanation of the ambiguous and unambiguous changes was worth 3 points (1.5 points for the explanation of the price change, 1.5 points for the explanation of the quantity change). For question 2 the correct answer required students to solve the percentage change in price. A discussion of the elasticity of demand formula was worth 1 point. This calculation was worth 1 point.

Finally, the skill pairing method (other than the random or neighbors pairing methods) was found to be the most conducive to the exercise. There was a greater degree of participation and less conversation. The higher skilled students assisted the lower scoring students with the development of the correct answers. Several groups finished early and were able to assist with groups that did not answer the questions as quickly.

Evaluation of the Exercise

The first time this exercise was used in class in the first semester, student participation was enthusiastic. A majority of students did not answer the essay questions accurately even with their notes, books, and quiz answers. The review of these answers therefore served as an applied exercise, and the following instructor lead discussion of these answers served as a review of one of the more complex topics covered to date. Student enthusiasm declined each of the three

times the exercise was completed within the first semester. Evaluations were used to quantify student responses and eventually to revise the exercise for the following semester.

Each time the exercise was administered the student determined weights were relatively close to those of the instructor. This outcome is likely related to the class discussion of the correct answers. Even so, the objective of the exercise is not for student to replicate the instructor's key but rather to help them develop the skills and methodology to evaluate essay answers. More specifically, the major benefits of the exercise include learning the grading rubric method in addition to the review of complex topics. An evaluation of the exercise was administered twice in the first semester (once after the first exercise and once at the end of the semester, after all three exercises were completed) and once in the second semester (at the end of the semester). Students were asked about the value of the exercise in helping them to understand the grading process, to write essay questions, to prepare for upcoming exams and to predict their grades. They were also asked if they would recommend doing such exercises in the future.

When asked how helpful the first exercise was in preparation of the upcoming exam and in writing essay questions, on a scale of 1 to 5 where 1 represented "not helpful" and 5 represented "extremely helpful", the average response was a "3.0" representing "helpful". This response was not correlated to the learning type, gender or GPA. Students that rated themselves as active learners (with a reliance on class exercises and practice questions) were not likely to respond that the exercise was more helpful than students that reported themselves as being more passive learners (with a preference for reading the text and studying outside of class).

The first semester the exercise was administered, student responses dropped on the evaluation administered at the end of the semester. Students reported that the exercises were "somewhat helpful" (average=2.4) rather than "helpful" in their preparation for future exams and

in answering essay questions (average=2.2). And, a majority did not feel that the exercise helped them to predict their grades. While a majority of students (64%) felt that the overall benefit of the exercise was helping them to understand the grading process, only 36 percent of the class recommended that it be used again the following semester. Students reported discontent with the exercise after multiple repetitions due to inconsistencies between the instructor's determined weights and their own, along with differences in the grading by the instructor and their partner. Although the student derived weights were similar to the instructors on average, slight differences in the weights that continued throughout the semester become frustrating. Students focused more on determining the weights and accurately grading the quiz rather the developing a methodology and applying it.

These responses prompted a change in the administration of the exercise in the following semester. Instead of administering the exercise multiple times, the exercise was administered once in the semester. In addition, students were no longer asked to grade their partners quiz. As a result of these changes, student ratings of the exercise were considerably improved. In this case, students reported that the exercise was "helpful" (average=3.1) in their preparation for future exams and in answering essay questions (average=3.2). Again, more than a majority of students (88%) felt that the overall benefit of the exercise was helping them to understand the grading process, with a large majority (88%) recommending that it be used again the following semester.

Expected Grades

Expected grades were recorded before and after taking exams to gain insight into whether a better understanding of exam evaluation had any impact on grade prediction. From the data

collected, a few observations can be made (see Table 1): 1) grade expectations generally fall after the completion of the exam and 2) grade expectations become more precise over the course of the semester, in particular for predictions after taking the exam, 3) expected grades for the class exceed the average or "C" range before taking the exam. This suggests, at least anecdotally, that the reason students over estimate their performance may not due to receiving higher grades in other courses (otherwise expected grades prior to taking the exam would fall over the course of the semester as students became more familiar with receiving lower grades in this economics course), nor is it do to a lack of understanding about the evaluation process. Instead the more important determinant appears to be the "above average effect" or general well wishing.

In the second semester the cooperative learning exercise was implemented after the first exam (and before the second) to determine if any significant differences were evident between expected grades before and after the administration of the cooperative learning exercise. There is no evidence that this exercise improved expected grades. Similar to the prior semester, expected grades fall for the second exam and improve after taking it.

To further investigate these conclusions, estimations are made on expected grades using student characteristics for prior GPA, SAT score, year (or cumulative hours) and gender (Table 2). Results suggest that expected grades prior and after taking the exam are not significantly influenced by gender, previous GPA, SAT, year, or performance on the exam. Also estimated is the difference between the predicted grade and the actual grade on the exam (i.e. accuracy of the prediction) before and after taking the exams. In these estimations, better students (as measured by prior GPA) are better predictors before taking the second exam, although these significant effects are not evident after taking the exam or for exam 1. These estimations provide further

support that independent of student ability, gender and other characteristics, students can make significant improvements to the accuracy of the prediction after taking the exam.

Finally, expected grades are presented in Table 3 for students that performed below and above average in the course overall. These data suggest that accuracy of prediction is similar between the two groups on average. While the predicted grade is lower for the students that performed below average the accuracy is similar to the other group for each exam both before and after administration. Together the results suggest that students are relatively good predictors of their grades, that grade inflation is not statically correlated with the variables tested, and therefore the overestimations would not bias regressions that substitute predicted grades for actual grades.

Lessons Learned

This exercise will provide the greatest benefits if administered once, early in the semester. Students would be given an opportunity to understand the grading process of the instructor prior to taking any exams and apply any lessons learned throughout the semester. The exercise is suited for classes of any size. The only prohibitive issue would be grading the essay questions of student in very large lectures. However, if teaching assistants (TAs) are used for grading, this exercise could also serve the dual role of training TAs that may be inexperienced graders. TAs could be used to monitor the progress of student throughout the exercise as well.

This exercise also has the additional benefit of assisting instructors with question design. According to Walstad (2006), poor question design can reduce the reliability of the test format as a measurement of student achievement. Vague open-ended questions may encourage student to include material not connected to course content or encourage guessing. The careful design of

test outcomes, development of precise questions and the establishment of consistent standards and processes for grading can improve the reliability of essay scores in the evaluation of economic understanding (Walstad 2006). Furthermore, it is believed that higher order thinking can be better evaluated using well designed, but less pointed questions that allow the student to develop a thesis, argument, and provide supporting evident. The appropriate grading method for such questions is more often the holistic approach giving the instructor greater ability to evaluate total content, theme, and the interconnections of ideas. This exercise does not apply to such questions or grading methods. Instead it is better suited for questions that can be broken into parts and assigned weights. Even so, teaching students to break questions into parts can serve to assist them with the greater task of developing more complex answers to open ended questions.

Conclusions

This paper describes the design and administration of a cooperative learning exercise with the multiple goals of assisting students with exam preparation, providing a better understanding of the grading process, and enabling students to critically evaluate essay questions and answers. Overall the cooperative learning exercise was shown to be helpful for student studying and in student understanding of the grading process. However, the exercise did not improve the precision of predicted exam or quiz grades for students. It is hypothesized that at least some component of expected grades represents the hopeful wishes of students that are unlikely to be influenced through such means.

The administration of the exercise multiple times enabled a better understanding of the weaknesses and possibilities for improvement. In retrospect, three applications in a single course are likely too many applications. Students were less receptive with the final exercise than the

previous two. According to students responses to the end of the semester evaluation this was because with each quiz the weights were determined differently, depending on the context and complexity of the question. Students were not able to master the weighting scheme; however they found the knowledge about the development of the weighting scheme important. The administration of this exercise once in a single semester provided greater benefits. In addition, the benefits of students grading a colleague's paper are likely minimal. It would likely be more beneficial for students to grade their own paper, to apply the developed weighting scheme, or to skip this step entirely.

It is often the case that active learning approaches are not appropriate or very difficult to administer in large lectures. The exercise would be applicable in such a setting, with the additional benefit of training TAs and evaluate their grading. Finally, the exercise takes approximately 20-25 minutes, and is therefore better suited for concepts that require review or additional attention.

Acknowledgements

This paper is the result of participation in the Teaching Innovations Program (TIP) in Economics, co-sponsored by the Committee on Economic Education (CEE) of the American Economic Association (AEA), and funded by the National Science Foundation. I would like to thank the coordinators of this program. In particular I would like to thank KimMarie McGoldrick, the coordinator of the cooperative leaning module, from which this paper evolved.

Table 1: Student Reported Expected and Actual Grades

Quiz or Exam	Expected grade before	Points From the Actual Grade	Expected grade after	Points From the Actual Grade	Actual Grade
First Semester					
Exam 1	79.22	-6.22	76.00	-3.00	70.03
Exam 2	76.12	-5.63	71.90	-1.41	70.49
Second Semester					
Exam 1	85.40	-13.24	76.64	-4.96	72.16
Exam 2	80.12	-8.7	74.89	-3.59	71.41

Table 2 – Estimation of Expected Grades

	Prior Expected Exam 1 Grade	Post Expected Exam 1 Grade	Prior Expected Exam 2 Grade	Post Expected Exam 2 Grade	Prior Difference in Expected Exam 1 Grade	Post Difference in Expected Exam 1 Grade	Prior Difference in Expected Exam 2 Grade	Post Expected Difference in Exam 2 Grade
Constant	62.62***	49.76***	62.50***	45.05***	19.08***	13.77***	13.22***	5.88
	(7.07)	(9.66)	(6.19)	(11.30)	(4.86)	(4.26)	(4.43)	(5.58)
Hours completed Prior to Taking the Course	-0.06	0.00	-0.05	-0.08	0.02	0.12	0.12	0.10
	(0.07)	(0.10)	(0.07)	(0.13)	(0.09)	(0.08)	(0.08)	(0.10)
Gender; =1 for Female; 0 Otherwise	0.94	-2.24	-2.69	-2.16	3.38	0.08	-1.50	2.60
	(2.50)	(3.44)	(2.36)	(4.02)	(3.09)	(2.70)	(2.72)	(3.23)
GPA Prior to Taking the Course	1.30	1.46	0.63	-0.80	-1.44	-1.67	-4.12**	-1.40
	(1.82)	(2.50)	(1.88)	(3.13)	(2.24)	(1.96)	(1.90)	(2.17)
SAT Score	0.00	0.00	0.00	0.00	-0.01	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)
Exam Score	0.27***	0.34***	0.24***	0.41****	NA	NA	NA	NA
	(0.09)	(0.13)	(0.09)	(0.15)				
Semester Dummy=1 for Second Semester; 0 otherwise	5.79**	1.91	3.73	4.61	2.77	-0.22	3.17	-0.13
	(2.50)	(3.43)	(2.37)	(3.89)	(3.20)	(2.76)	(2.70)	(3.10)
R-squared	0.24	0.19	0.28	0.27	0.17	0.08	0.13	0.07
Adj. R-Squared	0.15	0.09	0.18	0.16	0.08	0.01	0.03	-0.04
n	56	56	52	52	56	56	52	44

Notes: standard errors in parenthesis; *, **, *** indicate significance at the 10, 5, and 1 percent levels, respectively

Table 3: Student Reported Expected and Actual Grades by Student Type

	First Semester		Second Semester	
	Above Average	**Below Average**	**Above Average**	**Below Average**
Exam 1				
Prior to Exam 1	83.81	74.57	88.07	83.26
After Exam 1	81.56	70.20	78.89	74.65
Exam 1 Grade	83.98	66.75	83.07	62.98
Exam 2				
Prior to Exam 2	78.63	73.89	81.96	78.50
After Exam 2	75.16	69.00	76.90	73.21
Exam 2Grade	77.26	64.46	82.11	60.29

References

Alicke Mark D., M. L. Kloz, David L. Breitenbecher, Tricia J Yurak, and Debbie S. Vredenburg (1995) "Personal Contact, Individualism, and the Better-Than-Average Effect," *Journal of Personality and Social Psychology* 68 (5): 804-825.

Balch, W. (1992) "Effect of Class Standing on Student's Prediction of their Final Exam Score," *Teaching of Psychology* 19 (3): 136-141.

Ballard, Charles and Marianne Johnson (2005) "Gender Expectations and Grades in Introductory Microeconomics at a US University," *Feminist Economics* 11(1): 85-122

Bartlett, Robin L. (1998) "Making Cooperative Learning Work in Economics Classes," In *Teaching Economics to Undergraduates: Alternatives to Chalk and Talk* (1998) William E. Becker and Michael Watts, editors, Cheltenham, U.K. and Northampton, Mass.: Edward Elgar.

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

Cameron, Beverly (1998) "Active and Cooperative Learning Strategies for the Economics Classroom," in *Teaching Undergraduate Economics: A Handbook for Instructors* William B. Walstad, Phillip Saunders, editors, Boston; London and Toronto: Irwin/McGraw-Hill.

Cebula, Richard and Michael Toma (2002) "The Effect of Classroom Games on Student Learning and Instructor Evaluations," *Journal of Economics and Finance Education* 1 (2): 1-10.

Chambers, John R. and Paul D. Windschitl (2004) "Biases in Social Comparative Judgments the Role of Nonmotivated Factors in Above-Average and Comparative-Optimism Effects," *Psychological Bulletin* 130 (5): 813-838.

Fang, Hanming and Giuseppe Moscarini (2005) "Morale Hazard," *Journal of Monetary Economics 52* (4): 749-777.

Finegan, T. Aldrich and John J. Siegfried (1999) "Do Introductory Economics Students Learn More if Their Instructor Has a Ph.D.?" *The American Economist* 42 (2) 34-46.

Grimes, Paul W. (2002) "The Overconfident Principles of Economics Student: An Examination of a Metacognitive Skill," *Journal of Economics Education* Winter: 15-30.

Hacker, Douglas J., Linda Bol, and Dianne D. Horgan (2000) "Test Prediction and Performance in a Classroom Context," *Journal of Educational Psychology* 92 (1): 160-170.

Isley. P and H. Singh (2005) "Do Higher Grades Lead to Favorable Student Evaluations," Journal of Economic Education 36 (Winter) 29-42.

Jensen, Elizabeth J and Ann L. Owen (2001) "Pedagogy, Gender and Interest in Economics," Journal of Economic Education (Fall): 323-343.

Johnson, D. W, and R. T. Johnson and K.A. Smith (1991) *Active Learning: Cooperation in the College Classroom* Edina, MN: Interaction Book Company.

Krautmann, Anthony C. and William Sander (1999) "Grades and Student Evaluations of Teachers" *Economics of Education Review* 18 (1): 59-63.

Krueger, Joachim and Ross A. Mueller (2002) "Unskilled, Unaware or Both? The Better-Than-Average Heuristic and Statistical Regression Prediction Errors in Estimates of Own Performance," *Journal of Personality and Social Psychology* 82 (2): 180-188.

Lage, Maureen J., Glen J. Platt, and Michael Treglia (2000) "Inverting the Classroom: A Gateway to Creating an Inclusive Learning Environment," *Journal of Economic Education* (Winter) 30-43.

Lage, Maureen J. and Michael Treglia (1998) "Gender and Active Learning," In *Teaching Economics to Undergraduates: Alternatives to Chalk and Talk* (1998) William E. Becker and Michael Watts, editors, Cheltenham, U.K. and Northampton, Mass.: Elgar.

Maloof, Joan and Vanessa K. B. White (2005) "Team Study Training in the College Biology Laboratory," *Journal of Biological Education* 39(3): 120-124

Maier, Mark H. and Diane Keenan (1994) "Teaching Tools: Cooperative Learning in Economics," *Economic Inquiry* 32 (2): 358-361.

McGoldrick, KimMarie (2005) "Teaching Innovations Program: Cooperative Learning Module," TIP Workshop May 20-22, 2005, University of North Carolina-Chapel Hill.

McPherson, Michael A. 2006 "Determinants of How Students Evaluate Teachers," *Journal of Economic Education* 37 (Winter) 3-20.

McKelvie, Stuart J. (1997) "The Availability Heuristic: Effects of Fame and Gender on the Estimated Frequency of Male and Female Names," *Journal of Social Psychology* 137(1): 63-78.

Meyer, Herbert H. (1975) "The Pay-for-Performance Dilemma," *Organizational Dynamics* 3: 39-40.

Saunders, Phillip (1998) "Learning Theory and Instructional Objectives," in *Teaching Undergraduate Economics: A Handbook for Instructors*. William B. Walstad, Phillip Saunders, editors, Boston; London and Toronto: Irwin/McGraw-Hill.

Schaefer, Peter S., Cristina C. Williams, Adam S. Goodie, and W. Keith Campbell (2004) "Overconfidence and the Big Five," *Journal of Research on Personality* 38: 473-480.

Smith, Randolph A. (2005) "The Classroom as a Social Psychology Laboratory," Journal of Social and Clinical Psychology," 24 (1): 62-71.

Soll, Jack B. and Joshua Klayman (2004) "Overconfidence in Interval Estimates," *Journal of Experimental Psychology: Learning, Memory, and Cognition*," 30 (2): 299-314.

Svenson, O., (1981) "Are we all less risky and more skillful than our fellow drivers?" *Acta Psychologica* 47: 143–148.

Tversky, A. and Kahneman, D. (1973) "Availability: A Heuristic for Judging Frequency and Probability," *Cognitive Psychology* 5: 207-302.

Walstad, William B. (2006) "Testing for Depth of Understanding in Economics Using Essay Questions," *Journal of Economic Education* (Winter): 38-47