Heterogeneous Gender Effects under Loss Aversion in the Economics Classroom: A Field Experiment[1]

Maria Apostolova

William Cooper[2]

Gail Hoyt

Emily Marshall

Gatton College of Business and Economics Suite 335
University of Kentucky
Lexington, KY 40506

DRAFT
Cite only with authors' permission

October 17, 2013

Abstract

This paper evaluates the impact of loss aversion as a behavioral motivator on students' class performance, merging the behavioral economics and the educational incentives literature. The authors conducted an experiment with undergraduate students at the University of Kentucky, where student grades were framed in two different ways. In the treatment sections, the final course grade was framed as a loss, so that students begin the semester with full marks and as the course progresses lose points for less than perfect exam, quiz, and project scores. In contrast, in the control sections a traditional grading scheme was implemented where students begin the course with zero points and earn points throughout the semester as assignments are completed. We find that, at conventional significance levels, an individual in the treatment class did not have a statistically different final grade than an individual in the control class. However, we uncover a heterogeneous gender effect. Males in the treatment class score between 2.88 and 4.19 percentage points higher on the final grade than males in the control class, ceteris paribus. Conversely, females in the treatment class score between 3.30 and 4.33 percentage points lower on the final grade than females in the control class.

[1] We would like to extend our thanks to William Hoyt for administrative assistance in setting up our study, and John List, Eric Chiang, Chris Bollinger and Carlos Lamarche for comments and suggestions. We also wish to thank Bob Gillette for advice and encouragement as well as attendees at the University of Kentucky Economics Workshop who provided valuable comments and feedback.

[2] Corresponding author: wecoop3@uky.edu @uky.edu

1 Introduction

For over one hundred years, economists have operated under the assumption that agents are rational. However, recent advancements in psychology and behavioral economics provide evidence that people do not always make rational decisions. Behavioral economists study psychological heuristics, or cognitive shortcuts, frequently observed as influencing human behavior and causing individuals to occasionally make irrational choices. Humans think using two different systems– the Automatic system, which operates rapidly and instinctively while the other, the Reflective system, requires critical thinking and analysis. When individuals make decisions using the Automatic system, this can result in irrational behavior. Some of the biases that can emerge from using the Automatic system include anchoring, availability, representativeness, overconfidence, loss aversion, and status quo bias.[3]

In this paper, we study the impact of loss aversion on individual behavior and outcomes in the economics classroom. Loss aversion is a pattern that has been observed in human decision-making processes. The general consensus is that, typically, people are about twice as unhappy about losing an asset, as they are happy about acquiring the same possession. Economists would define this as irrational behavior. According to expected utility theory, the same good or service should result in an equal utility increase or decrease whether it is acquired or lost; however, it has been documented that people hate losses more than they like gains. Kahneman and Tversky (1979) provide an alternate theory that describes decision-making under risk, reassessing the conventional model of expected utility theory. Based on empirical evidence, they present several

[3] These terms were originally coined by Sustein and Thaler (2009).

cases where agents violate the axioms of expected utility theory. For instance, individuals respond asymmetrically when they are faced with the prospect of gains or losses. Kahneman and Tversky (1979) develop what is referred to as "prospect theory," in which the value function is concave for gains and convex for losses based on deviations from a reference point. In general, the value function is steeper for losses than for gains, representing the behavioral phenomenon that people are more upset about losing something they once had, as they are happy about gaining the same thing.

Our investigation will expand the current literature on loss aversion by examining individual behavior in an educational setting. We hypothesize that if people dislike losses more than they enjoy gains, we can induce higher student performance by framing grades as a point reduction, as opposed to earning points throughout the semester. Our empirical analysis finds that, on average, an individual in the treatment class did not have a statistically different final course grade than an individual in the control group, controlling for other variables that have been shown to affect scholastic performance.[4] However, we find that males in the treatment group score about 2.88 and 4.19 percentage points higher on the final course grade than males in the control class, holding all other factors constant, which is consistent with the hypothesis derived from the behavioral literature. In addition, our results show that females in the treatment class score about 3.30 and 4.33 percentage points lower on the final course grade than females in the control class, ceteris paribus.

[4] With a slight abuse of notation, we use group and class interchangeably.

The paper proceeds as follows. Section 2 gives an account of the relevant literature. Section 3 discusses the design of our field experiment. Section 4 gives a selection of summary statistics and variable definitions. Section 5 discusses our empirical methodology, and Section 6 provides the results of our analysis. Section 7 concludes.

2 Review of the Literature

The field of behavioral economics has gained momentum over the last decade. A plethora of experimental evidence shows the significance of psychological heuristics that differentiate humans from the rational actors that most economists study. In line with the center theme of this analysis our review of the literature surveys empirical pieces that study the behavioral bias of loss aversion and student achievement.

The following recent works explicitly incorporate loss aversion in the experimental design. Among them, two studies evaluate the effects of loss aversion on student outcomes. Fryer Jr. et al. (2012) explore the impact of teacher incentive program framing on student outcomes and teacher performance. In several Chicago schools, they implement a loss aversion component in the teacher incentive scheme, and in others, employ the traditional gain structure. Under the traditional reward system, teachers receive an outcome-related bonus at the end of the school year based on student performance. Alternatively, the novel loss scheme awards teachers the same bonus at the beginning of the school year, but they are obligated to return a portion if their students do not meet performance targets. The results show a statistically and economically significant positive relationship between math scores and teacher compensation in the loss group.

Specifically, students' math scores increase between 0.20 and 0.40 standard deviations, which is consistent with prospect theory and loss aversion.

Similarly, Levitt et al. (2012) analyze the effects of cognitive shortcuts on degree of student success, rather than teacher motivation and performance. The authors assess the effectiveness and magnitude of monetary and non-monetary incentives on student test performance. To this end, cash and non-cash rewards of different sizes are introduced in the experiment design. In addition, the authors test for loss aversion-induced behavior by framing the rewards as gains or losses, which is closely related to the present analysis. Contrary to Fryer Jr. et al (2012) study Levitt et al. (2012) do not find evidence for improved student outcomes under a loss framework.

Many field experiments examine framing effects outside of the education environment. Hossain and List (2012) investigate the influence of behavioral factors, specifically loss aversion, on work performance. The authors use two primary types of reward systems for group and individual performance to test whether framing bonuses as gains or losses significantly alters worker productivity at a Chinese electronics manufacturing company. Unlike the above loss aversion experiments conducted in an educational environment where the effects of loss framing are ambiguous, in a business setting both the gain and loss structure exert a positive and statistically significant influence on productivity. Consistent with prospect theory, productivity is more favorably affected by a bonus loss than by a bonus gain.

There is an extensive body of literature in the area of economic education investigating the impact of various factors that influence student learning and performance in the college economics classroom. We use these works primarily as justification for the inclusion of various explanatory variables to ensure that we control for factors that could be affecting student performance other than the treatment.[5]

To date, there appear to be no studies examining the direct impact of grading structure on learning and course performance in economics classes and more specifically, no studies considering the potential behavioral response to loss aversion resulting from different framing of the course grade. In addition, while there is an extensive body of literature examining gender differences and incentive effects,[6] our review of the literature finds no attempts to measure heterogeneous gender effects in response to a loss aversion framework. To our knowledge this is the first study that examines student outcomes and gender heterogeneity in the context of a loss framing of the course grade.

This brief survey of the existing literature suggests that loss aversion can represent a strong incentive for behavioral adjustments in various settings. Since the majority of the current research finds that framing of incentives as losses is associated with better productivity and learning outcomes, the existing empirical evidence provides a sufficient foundation to proceed

[5] Miller and Rebelein (2012) summarize the literature that analyzes the impact of various teaching techniques such as cooperative learning, classroom experiments, case studies, experiential learning, and undergraduate research on student learning in economics courses. Benedict and Hoag (2012) review studies that consider the role of varying usage of graphs and quantitative skills on student performance. Owen (2012) surveys the existing literature regarding the impact of student characteristics (gender, major, attendance, prior course experience, etc.) on behavior and performance in economics classes. Finally, Grove and Wu (2012) summarize the body of work on the impact of class and instructor characteristics on student performance in undergraduate economics courses.

[6] A notable example, Kremer, Miguel, and Thornton (2009) assess the impact of merit scholarships on school performance of girls in Kenya. The authors find that offering a merit scholarship to a cohort of Kenyan girls not only increases their educational attainment through increased study effort but also uncovers the presence of positive externalities associated with increased effort of boys.

with an analysis of the impact of loss aversion on student grades and performance in economics courses.

3 Field Experiment Design

The goal of this experiment is to discover whether a simple change in the grading scheme inspired by the behavioral economics principle of loss aversion can motivate students to perform better in a course.[7] Specifically, we implement a grading system in which students are “given” points when the course commences and progressively lose points throughout the semester for incorrect answers (a penalty contract-aversion), as opposed to earning points on assignments and exams as they are completed (a reward contract-reciprocity). For the remainder of the paper, the sections in which the loss aversion-grading scheme was implemented will be referred to as the treatment and the sections in which the canonical grading system was employed will be identified as the control.

The syllabus of the treated class contained the following paragraph:

> You begin the class with 500 points out of 500 points. Points will be deducted from 500 based on your performance. For example, if you score an 85 on an exam that is worth a maximum of 100 points, you lose 15 points. To obtain your final grade, take your remaining points and divide by 500. Your final grade for the course is based on the distribution in Table 1.

(Table 1: Class Grade Components)

[7] Protocol number 12-0953-P4S of the University of Kentucky Institutional Review Board.

The only difference between this grading scheme and the control is that the students see full points for all exams and assignments that have not yet been completed. In the treatment section, all of the students start the semester with a grade of 100% (A) in the class and can view this online via Blackboard.[8] As each exam or assignment is completed, the students' grades are updated, and they can see their overall grade drop as a result. (Unless, of course, they receive full marks for an exam or assignment.)

We implemented this experiment in a total of four classes, two for each instructor. Each teacher had a control section and a treatment section. Instructor X taught two sections of Principles of Macroeconomics, and instructor Y taught two sections of Economics and Business Statistics.[9] The treatment class was the second section for instructor X, and the first section for instructor Y.[10] Class sizes were comparable both across control and treatment groups, as well as between the two different courses.[11] The intent of the experiment design is to hold constant any difference in student success that may arise due to peer effects or direct effects from class sizes.

On any given day, the same lecture and/or quiz were given in each section of the respective courses. With the exception of minor differences in discourse, the two sections for each instructor were taught identically. The evaluation criterion for each course consisted of

[8] In the traditional grading system, students begin the semester with zero points and can earn up to 500 points throughout the semester. The control group sees this type of grading structure on Blackboard.

[9] The two sections of Principles of Macroeconomics were taught on Monday, Wednesday, and Friday from 11:00 a.m. to 11:50 a.m. and 12:00 p.m. to 12:50 p.m. The two sections of Economics and Business Statistics were taught on Monday, Wednesday, and Friday from 12:00 p.m. to 12:50 p.m. and 1:00 p.m. to 1:50 p.m. respectively. This was the most uniform class schedule across instructors our department could accommodate.

[10] We randomly selected which section was the treatment for each instructor. In addition, we performed tests for mean differences on pre-treatment characteristics between the treatment and control sections for each course. The tests showed that the sections were not statistically different (with the exception of Asian race and age), indicating that our randomization technique is valid. See Section 4 for additional details.

[11] At the end of the semester, both sections of Principles of Macroeconomics had 55 students enrolled. Final course enrollment for the first section of Economics and Business Statistics was 43 and for the second section was 47. Demographic descriptive statistics of each section, as well as formal tests of statistical differences, can be found in Table 3.

twenty quizzes worth five points each, a group project worth 100 points, and three exams worth 100 points each, as shown in Table 1 above. Daily quizzes and assignments were graded for completion—not correctness. The twenty quizzes were a combination of Turning Point clicker quizzes and problem sets. Students were notified at the beginning of the semester that their lowest four quiz grades would be dropped. Therefore, the points for the daily quizzes and assignments category were taken out of 80 instead of 100 and converted to a percent (ratio of total quiz points to 100). However, a student cannot score over 100 in this category, even if he or she completed all 20 quizzes.

At the beginning of the semester, students were assigned to groups of approximately four to five members. Over the term, each group turned in five parts of a project that summed to a total of 100 points. A student's project grade was the simple average of two grades: the group project grade and an individual contribution grade. The last week of classes, each student was asked to evaluate his peers based on several different criteria. The average of a student's peer ratings out of a possible 36 points is her individual contribution grade.[12] However, if a student received below a B- on the individual-contribution grade, her final project grade was only her individual-contribution grade.[13] The peer evaluation portion of the project grade is removed from our evaluation criterion due to its high subjectivity. Only the portion of the project grade scored by the instructor is included for analysis.

[12] The peer evaluation includes nine questions regarding various aspects of effort and contribution toward the group project with responses scored on a four point Likert scale.

[13] For example, if a group's project grade is 94% and an individual-contribution grade is 76%, then the individual's grade will be 76%, not $\frac{94+76}{2}$=85%.

The final examination was only cumulative in the sense that it required knowledge from the other exams, but students were not explicitly tested on their ability to solve problems from earlier in the semester. Exams and major project components were due on the same day in both sections and both courses. Final grades were calculated by summing the modified quiz score, project grade, and the three exam grades.[14] The total course points divided by 500 and converted to a percentage was a student's final course grade.

4 Descriptive Statistics

Data for our study come from four sources: a pre-course survey, a post-course survey, the registrar's office, and instructor records. On the third day of class, students completed surveys to gather information on basic demographics and educational background. On the day of the final exam, another survey was administered to solicit student opinion on the course experience and self-reported values for attendance and work effort. Data regarding grade point average (GPA), SAT scores, ACT scores, and major were also collected from the campus registrar's office. Additionally, course performance variables were collected from instructor grade sheets. Table 2 below provides descriptions for variables used in the analysis.[15]

(Table 2: Variable Descriptions)

Our full sample consists of 176 students. Approximately 30.86% of the students are

[14] In addition, as outlined in the Institutional Review Board (IRB) documentation, each student received 20 points extra credit regardless of whether or not she participated in the study.
[15] Please note this is not a comprehensive list of all variables for which data was collected.

female and 77.91% are white. Despite the fact that our sample consists predominantly of sophomores and juniors, the mean age at the beginning of the semester is 21.95 years. Sophomores and juniors comprise 83.50% of the sample, but we control for age in our analysis. Average cumulative GPA prior to the beginning of the Spring 2013 semester is 3.00 for the full sample. Approximately 14.04% of all students are economics majors.[16] Table 3 below provides a more comprehensive summary of the sample descriptive statistics. Our experiment was designed to ensure identical treatment and control groups based on pretreatment characteristics. Table 3 also highlights statistically significant mean differences between various subsamples in an effort to determine whether the treatment and control sections were statistically different.

(Table 3: Descriptive Statistics)

Columns 2 and 3 show statistics on means and number of observations for the full sample treatment and control groups. Columns 4 and 5 display the same data for each course. Shaded cells and bold text indicate statistically significant mean differences between the treatment and control group (columns 2 and 3) and the two courses (columns 4 and 5).

There are some statistically significant differences between students in the macroeconomics course and the business statistics course, which is a higher-level course. Students in business statistics have a statistically higher average college GPA than students in the

[16] For the most part, our sample appears to be representative of the undergraduate Gatton College of Business and Economics population. The University of Kentucky undergraduate business school is approximately 80.77% white (non-Hispanic only), 35.92% female, and the average cumulative GPA is approximately 3.14.

principles course. The students enrolled in the higher-level course are also most likely a year further along in college than the students in the introductory course; therefore, the two courses also differ in terms of age. The college major composition of the two courses in our sample also reflects differences in their nature—there are more economics majors in business statistics. In our regression analysis, we include an indicator variable for course that accounts for any intra-course differences.

At the end of the semester, we had student consent to use complete data for a total of 85 individuals in the two treatment sections and 91 in the control sections.[17] As anticipated, our treatment and control sections do not differ significantly in terms of demographic characteristics. We note two exceptions: age and race. In the control group, 17.78% of students identified themselves as Asian versus 4.88% in the treatment group. The second exception is mean age. Even though our treatment and control groups are not perfectly isomorphic, we are able to correct for the differences by controlling for the statistically significant variables in our empirical estimation.

5 Empirical Methodology

Our main evaluation criterion is the final course grade (Final Grade) calculated without any curve designated by the instructor (un-curved), the individual component of the project grade, and bonus points related to the field experiment. In addition, the final course grade for the

[17] Over the course of the semester, 3 students withdrew from the treatment classes and 2 withdrew from the control sections. Students that dropped from the control group did not enroll in the treatment class or vice versa. In addition, there were 3 students that did not complete the course. Of these 8 students, 2 of them did not consent to participate in the study at the beginning of the semester. Therefore, a total of 6 students from the original dataset were excluded for analysis.

assignments category is calculated as stated on the syllabus, with the four lowest quiz grades dropped. The total course grade is scored out of 500 points, but for the purpose of empirical investigation (and letter grade assignment), it is converted to a percentage. Thus, the dependent variable is measured on a 0 to 100 scale.

We follow the econometric literature that studies treatment effects. The baseline ordinary least squares (OLS) regression specification is

$$Final\ Grade_{ij} = \beta_0 + \beta_1 treat_{ij} + \beta_2 macro_j + \beta_3 \underline{X_i} + \varepsilon_{ij} \qquad (1)$$

where $treat_{ij}$ is a dummy variable equal to 1 if the student was in a treatment section, $macro_j$ is a dummy variable equal to 1 if the individual was in principles of macroeconomics, $\underline{X_i}$ is a set of individual specific control variables, and ε_{ij} is the stochastic error term.[18] Subscripts indicate student i in course j. The $macro_j$ variable captures any omitted instructor fixed effects that could potentially affect student performance and inclusion of both $macro_j$ and $treat_{ij}$ identifies the course section.

The parameter of interest is β_1, which represents the change in final course grade resulting from the treatment (the loss aversion grading scheme). For our choice of control variables, we relied primarily on the educational outcomes literature. In addition, we explicitly controlled for differences in student composition between the treatment and control classes. The set of control variables includes demographic characteristics (age, race, gender, year in college), cumulative college GPA, ACT math score, a dummy variable for whether or not the individual

[18] We explore clustering the standard errors at the level of the treatment as a robustness check. See Section 6 for a complete discussion.

attended a private high school, a dummy variable indicating if the student is an economics major, the number of times the student accessed blackboard, the number of assignments the individual did not turn in, the number of hours worked per week, and total number of high-school and college math courses the student has taken.

The second OLS regression specification augments the baseline model with an interaction term between the treatment variable and gender in order to explore the possibility of a heterogeneous gender-specific effect.

$$Final\ Grade_{ij} = \beta_0 + \beta_1 treat_{ij} + \beta_2 female_i + \beta_3 treatXfemale_{ij} + \beta_4 macro_j + \beta_5 \underline{X_i} + \varepsilon_{ij} \quad (2)$$

We focus on this hypothesis for two reasons. First, as show in Figure 1 below, the differences in mean final grades for males and females in the treatment and control groups merits further analysis.

(Figure 1)

For both males and females, the means are statistically different at the 5% level. Second, as discussed previously, the existing literature on student incentives and outcomes commonly finds a gender effect, albeit, in the absence of loss framing.[19]

6 Results

[19] See, for example, Kremer, Miguel, and Thronton (2009).

Table 4 shows the baseline results from regression specification (1). The dependent variable is the final course grade using un-curved exam scores and excluding project peer evaluation. At conventional significance levels (0.01, 0.05, and 0.10), there is no statistically significant impact of the loss aversion-grading scheme on a student's final course grade. Therefore, we cannot claim that framing matters for overall student outcomes. As anticipated, we note a few demographic characteristics that have a significant impact on course grade. These results reference regression (1.2), as it is the most accurate eliminating the majority of omitted variable bias by controlling for many factors that contribute to student course performance. See Table 4 below.

(Table 4)

An Asian student has a final course grade that is 6.60 percentage points higher than a black student, holding all other factors constant. A white student has a final course grade that is 4.64 percentage points higher than a black student, ceteris paribus. Holding all other variables constant, a female has a final course grade that is 3.82 percentage points lower than a male student. We will return to discussing gender effects in Table 5.

Variables traditionally used to measure various aspects of academic ability such as GPA and ACT Math score have favorable and statistically significant effects on the final course grade. Class standing is also a statistically significant predictor of final grade. A student who is further along in his or her studies tends to do better than an otherwise comparable student who has not been in college as long. Also, a relatively obvious result, the number of assignments not turned in has a negative marginal impact on the final course grade. We interpret the number of assignments with a score of 0 to be an indicator of effort level. Drawing from the school performance and outcomes literature where it is standard to include a measure of family wealth, we add a dummy variable for whether a student attended a private high school; however, this indicator variable is not statistically significant.[20]

Next, we explore the existence of a differential treatment effect based on gender. Table 5 displays the regression output for model (2).

(Table 5)

If our experiment design resulted in a successful randomization there should be no need to control for any variables other than the treatment itself because the treatment and control groups should be identical. However, as discussed previously, the groups differ significantly on a few dimensions. It is encouraging that regardless of these differences the coefficients for both male and female students are statistically significant in the basic regression (2.1.) with no additional controls.[21] Our baseline results (regression 2.1) indicate that the final course grade of a male student in the loss framing class is 2.88 percentage points higher, on average, than a male student in the traditional grading scheme class. Interestingly, the loss grade framing has an opposite

[20] The most commonly used measure of household wealth is eligibility for free or reduced-price lunch; however, this is not applicable in a college setting.
[21] The coefficients on Treatment and TreatXFemale are jointly statistically significant at the 5% level. See Table 5 for details.

effect on outcomes for female students. On average, the final course grade of a female student in the treatment group is 4.30 percentage points lower than that of a female student in the control group.

Based on results from regression 2.4, which includes additional control variables, males in the treatment class earn a final course grade that is 3.49 percentage points higher than a male student in the control class, on average. The negative effect of the loss framing for female students is slightly reduced. The final grade for a female student in the treatment group is 3.49 percentage points lower compared to a female in the control class. The additional explanatory variables have magnitudes and significance levels that are similar to those already discussed and presented in Table 4.

A robustness check using clustered standard errors at the treatment level (section) confirmed our results. We used the wild bootstrap procedure for recovering standard errors suggested by Cameron, Gelbach, and Miller (2008) because of the need to correct for the small number of clusters.

7 Conclusion

The results of our experiment testing for loss aversion behavior among undergraduate students suggest that framing has a differentiated gender impact on student outcomes in the economics classroom. Framing the course grade as a loss rather than a gain of points increases a male student's grade between 2.88 and 4.19 percentage points. This result implies that using a loss aversion-grading scheme can have positive effects on male final course grades. At the same time, under the loss aversion-grading scheme, a female student's grade decreases between 3.30 and 4.33 percentage points compared to the traditional grading scheme. Therefore, the educator faces a tradeoff between the costs and benefits of such an innovative grading structure.

Reference List

Benedict, Mary Ellen, and John Hoag. (2012). Factors influencing performance in economics: graphs and quantitative usage. In *The International Handbook on Teaching and Learning Economics*, edited by Gail Hoyt and KimMarie McGoldrick. Edward Elgar Press, pp. 334-40.

Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller. (2008). "Bootstrap-based improvements for inference with clustered errors." *The Review of Economics and Statistics* 90, no. 3: 414-427.

Fryer Jr, Roland G., Steven D. Levitt, John List, and Sally Sadoff. (2012). *Enhancing the Efficacy of Teacher Incentives through Loss Aversion: A Field Experiment*. No. w18237. National Bureau of Economic Research.

Grove, Wayne A., and Stephen Wu. (2012). Factors influencing student performance in economics: class and instructor characteristics. In *The International Handbook on Teaching and Learning Economics*, edited by Gail Hoyt and KimMarie McGoldrick. Edward Elgar Press, pp. 351-59.

Hossain, Tanjim, and John A. List. (2012). "The behavioralist visits the factory: Increasing productivity using simple framing manipulations." *Management Science* 58, no. 12: 2151-2167.

Kahneman, Daniel, and Amos Tversky. (1979). "Prospect Theory: An Analysis of Decision under Risk." *Econometrica* 47, no. 2: 263-292.

Kremer, Michael, Edward Miguel, and Rebecca Thornton. (2009). "Incentives to learn." *The Review of Economics and Statistics* 91, no. 3: 437-456.

Levitt, Steven D., John A. List, Susanne Neckermann, and Sally Sadoff. (2012). *The behavioralist goes to school: Leveraging behavioral economics to improve educational performance*. No. w18165. National Bureau of Economic Research.

Miller, Joshua D., and Robert P. Rebelein. (2012). Research on the effectiveness of non-traditional pedagogies. In *The International Handbook on Teaching and Learning Economics*, edited by Gail Hoyt and KimMarie McGoldrick. Edward Elgar Press, pp. 323-33.

Owen, Ann L. (2012). Student Characteristics, behavior, and performance in economics classes. In *The International Handbook on Teaching and Learning Economics*, edited by Gail Hoyt and KimMarie McGoldrick. Edward Elgar Press, pp. 341-50

Sustein, C. R. and Thaler, R. H. (2009). *Nudge: Improving decisions about health, wealth, and happiness*. Penguin Books.

Table 1: Class Grade Components

Task	Points	Percentage of Grade
Daily Quizzes and Assignments	100	20%
Group Project	100	20%
Exam I	100	20%
Exam II	100	20%
Final Exam	100	20%
Total	500	100%

Table 2: Variable Descriptions

Variable	Description	Source
Final course grade	Final course grade (excludes peer evaluation, bonus points, and exam curves)	Blackboard
Macro	Course (1 for Principles of Macroeconomics)	
Female	Gender (1 for female)	Beginning of semester survey
White	Race, White	Beginning of semester survey
Asian	Race, Asian	Beginning of semester survey
Black	Race, Black	Beginning of semester survey
Age	Age (calculated from birth year and month)	Beginning of semester survey
Private HS	Private high school (1 if the student attended a private high school)	Beginning of semester survey
College GPA	College Cumulative GPA including the previous semester, but excluding the semester of the study	Registrar's office
ACT math	Highest ACT math	Registrar's office
Math all	Number of high school and college math courses taken	Beginning of semester survey
Econ major	Primary major economics (1 if the student's primary major is economics)	Registrar's office
Access total	Number of times accessed Blackboard	Blackboard
Assignments zero	Number of assignments not turned in	Blackboard
Class	Year in college (Freshman=1, Sophomore=2, Junior=3, Senior=4, Professional, second year=5)	Registrar's office
Hours work	Hours per week of work	End of semester survey

Table 3: Descriptive Statistics

Variable	Full Sample (1)		All Treat (2)		All Control (3)		All Macro (4)		All Statistics (5)	
	n	Mean	n	Mean	N	mean	n	mean	n	mean
Final course grade	176	75.6	85	76.0	91	75.2	**95**	**77.2***	**81**	**73.8***
Female	175	0.31	84	0.29	91	0.33	94	0.31	81	0.31
White	172	0.78	82	0.83	90	0.73	92	0.79	80	0.76
Asian	172	0.12	**82**	**0.05***	**90**	**0.18***	92	0.09	80	0.15
Black	172	0.10	82	0.12	90	0.09	92	0.12	80	0.09
Age	166	21.95	**80**	**22.32***	**86**	**21.60***	**88**	**21.62***	**78**	**22.32***
Private HS	171	0.29	81	0.30	90	0.29	**90**	**0.21**	**81**	**0.38**
GPA, Cumulative Total	171	3.00	81	3.02	90	2.98	**90**	**2.90**	**81**	**3.11**
ACT math	171	21.81	81	22.43	90	21.24	90	22.17	81	21.41
Math all	176	6.74	86	6.68	91	6.79	**95**	**6.00***	**81**	**7.609***
Econ major	171	0.14	**81**	**0.20**	**90**	**0.09**	**90**	**0.02***	**81**	**0.27***
Access total	176	66.83	85	65.52	91	68.06	**95**	**99.52***	**81**	**28.49***
Assignments zero	176	3.64	85	4.16	91	3.15	95	3.55	81	3.75
Class	171	2.78	**81**	**2.96***	**90**	**2.62***	**90**	**2.48***	**81**	**3.12***
Hours work	176	10.79	85	11.69	91	9.95	95	10.73	81	10.87

*Note: Shaded cells and bold text indicate statistically different means between the indicated samples. ** and *** indicate significance at the 5% and 1% levels respectively.*

Table 4
Dependent variable: Final course grade, full sample

	(1.1)	(1.2)
Treatment	1.946	1.377
	(1.292)	(0.969)
Course	3.593***	7.023***
	(1.262)	(1.487)
Female	-4.149***	-3.816***
	(1.267)	(1.086)
Asian	7.259**	6.599**
	(3.069)	(2.868)
White	7.323***	4.641**
	(2.468)	(2.054)
Age	-0.766	-0.756*
	(0.521)	(0.434)
Class		1.921**
		(0.870)
College GPA		3.520***
		(1.226)
ACT math		0.229***
		(0.071)
Private HS		1.436
		(1.094)
Econ major		3.658**
		(1.766)
Access total		-0.012
		(0.013)
Assignments zero		-0.964***
		(0.243)
Hours work		-0.079*
		(0.044)
Math all		0.202
		(0.205)
Observations	163	163
Adjusted R-squared	0.141	0.491

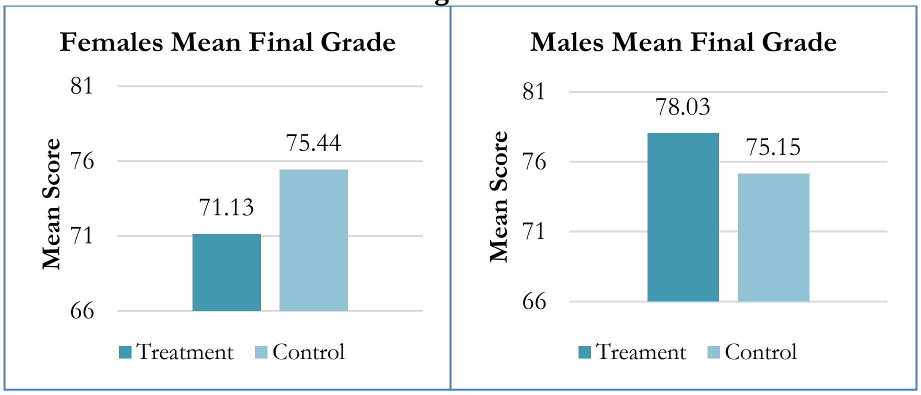
*Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The omitted race category is black.*

Table 5
Dependent variable: Final course grade, full sample

	(2.1)	(2.2)	(2.3)	(2.4)
Treatment	2.882*	2.796*	4.186***	3.486***
	(1.484)	(1.534)	(1.515)	(1.120)
Female	0.283	0.253	-0.632	-0.482
	(1.820)	(1.631)	(1.589)	(1.525)
TreatXFemale	-7.182***	-7.125***	-7.481***	-6.979***
	(2.683)	(2.451)	(2.529)	(1.936)
Course		3.380***	3.449***	6.665***
		(1.233)	(1.230)	(1.462)
Asian			6.816**	6.395**
			(3.031)	(2.735)
White			7.555***	4.954**
			(2.427)	(1.980)
Age			-0.817*	-0.945**
			(0.490)	(0.394)
Class				2.274***
				(0.794)
College GPA				2.897**
				(1.184)
ACT math				0.237***
				(0.071)
Private HS				1.488
				(1.037)
Econ major				3.978**
				(1.640)
Access total				-0.009
				(0.013)
Assignments zero				-0.976***
				(0.220)
Hours work				-0.084**
				(0.040)
Math all				0.189
				(0.203)
Prob > F (Treatment + TreatXFemale)	0.026	0.016	0.006	0.001
Observations	175	175	163	163
Adjusted R-squared	0.053	0.089	0.179	0.526

*Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The omitted race category is black.*

Figure 1


Females Mean Final Grade
81
76
71
66
Mean Score
71.13
75.44
Treatment
Control
Males Mean Final Grade
81
76
71
66
Mean Score
78.03
75.15
Treament
Control