2019

Improving Student Performance Through Loss Aversion

Ben O. Smith
University of Nebraska at Omaha, bosmith@unomaha.edu

Rebekah Shrader
California State University, Stanislaus

Dustin R. White
University of Nebraska at Omaha, drwhite@unomaha.edu

Jadrian J. Wooten
The Pennsylvania State University

Steve Nath
University of Nebraska at Omaha

See next page for additional authors

Recommended Citation

Smith, B. O., Shrader, R., White, D. R., Wooten, J., Dogbey, J., Nath, S., . . . Rosenman, R. (2019). Improving student performance through loss aversion. Scholarship of Teaching and Learning in Psychology. Advance online publication. http://dx.doi.org/10.1037/stl0000149

Authors
Ben O. Smith, Rebekah Shrader, Dustin R. White, Jadrian J. Wooten, Steve Nath, Michael O'Hara, Nan Xu, and Robert Rosenman

Improving Student Performance Through Loss Aversion

Ben O. Smith, Rebekah Shrader, Dustin R. White, Jadrian Wooten, John Dogbey, Steve Nath, Michael O'Hara, Nan Xu, and Robert Rosenman

University of Nebraska at Omaha, CSU Stanislaus, University of Nebraska at Omaha, The Pennsylvania State University, Louisiana Tech University, University of Nebraska at Omaha, University of Nebraska at Omaha, Amazon.com, and Washington State University

Author Note

Ben O. Smith, Department of Economics, University of Nebraska at Omaha

Rebekah Shrader, Department of Economics, CSU Stanislaus

Dustin R. White, Department of Economics, University of Nebraska at Omaha

Jadrian Wooten, Department of Economics, The Pennsylvania State University

John Dogbey, Department of Economics & Finance, Louisiana Tech University

Steve Nath, Department of Accounting, University of Nebraska at Omaha

Michael O'Hara, Department of Finance, Banking & Real Estate, University of Nebraska at Omaha

Nan Xu, Amazon.com

Robert Rosenman, School of Economic Sciences, Washington State University

Correspondence concerning this article should be addressed to Ben O. Smith, Department of Economics, University of Nebraska at Omaha, 6708 Pine Street Omaha, NE 68182.

Contact: bosmith@unomaha.edu

We thank Emily Marshall, Gail Hoyt, Regan Gurung, and the anonymous referees for helpful comments. We further thank Halli Tripe and Jackie Lynch for editorial suggestions.

Abstract

Framing an outcome as a loss causes individuals to expend extra effort to avoid that outcome (Tversky & Kahneman, 1991). Since classroom performance is a function of student effort in search of a higher grade, we seek to use loss aversion to encourage student effort. This field quasi-experiment endows students with all of the points in the course upfront, then deducts points for each error throughout the semester. Exploiting two course sequences in the business school of a Midwestern university, a control for domain-specific knowledge, this study examines the impact of loss aversion when controlling for the student's knowledge in a specific subject. This quasi-experiment indicates that students perform three to four percentage points better when controlling for student ability and domain knowledge (148 subjects). This result is significant at the 1% level in our most robust specification ($p = 0.0020$). This result is confirmed by a specification including four courses and controlling for student characteristics (217 subjects, $p = 0.0190$).

Keywords: prospect theory, loss aversion, student grades, incentives

Improving Student Performance Through Loss Aversion

Policymakers have had tremendous success implementing principles from behavioral sciences to induce improved decision-making among their constituents. The work of the Behavioural Insights Team, in particular, has been successful in using methods from the behavioral sciences to improve outcomes in the United Kingdom.[1] Simple insights have resulted in reductions in litter, improved unemployment processes, and reductions in crime.

Recent research suggests that many of the practices that improve social outcomes can also be used to improve outcomes in the classroom. Notifying parents of K-12 students of their child's absenteeism in comparison to their peers results in a reduction in absenteeism (Robinson, Lee, Dearing, & Rogers, 2017). A study at the collegiate level showed that informing students how an upcoming assignment could impact their overall grade resulted in increased performance on the assignment among treated students (Smith, White, Kuzyk, & Tierney, 2018).

Researchers in psychology learning have used similar techniques to explore ways of improving classroom performance. Warnings of the cognitive downsides of too much lecture (Cerbin, 2018), or the impact of goal orientation on graduation rates (Hoyert, O'dell, & Hendrickson, 2012) have been explored. Some studies focus on mechanical issues such as randomizing exam item order (Cathey, Sly, & Barry, 2018).

Grades in the classroom act as both an incentive and a signal of quality. Many studies have examined the importance of this incentive in multiple contexts. Student satisfaction with a course increased with both perceived learning and grade outcomes (Bean & Bradley, 1986; Lo, 2010). Even when controlling for ability measures, happiness and course satisfaction were correlated with grade (Quinn & Duckworth, 2007); overconfident students, in particular, are more likely to be unhappy with their grade outcome and dissatisfied with the course as a result (Grimes, 2002). When it comes time in the semester for students to evaluate their instructor, students often value their grade over what they might have learned (Clayson, Frost, & Sheffet, 2006).

The Present Study

Seminal works in behavioral economics emphasized the importance of an individual's reference point on the formation of preferences (Kahneman & Tversky, 1979; Tversky & Kahneman, 1991). These papers found that framing decisions in the context of a loss as opposed to a gain had a strong tendency to reveal inconsistent preferences, thereby changing the behavior of an individual. When re-framing a choice as a loss, individuals chose an outcome that differs from the outcome that they would have preferred prior to the re-framing of the choice. This pattern of behavior is referred to as loss aversion.

Given the importance students place on grades, it is only natural to explore whether this framing can be exploited to increase learning in academic coursework. This paper explores whether instructors are able to manipulate student preferences utilizing the concept of loss aversion by placing all assignments on a negative scale, with 0 denoting a perfect score, and all other scores framed as missed (negative) points. Students start the semester with the full number of points available throughout the semester and lose points on each subsequent assignment where errors were made (rather than traditionally gaining points on each assignment). Since students should be more resistant to poor grades when those grades are framed as a loss, loss aversion predicts that students should increase their effort in order to offset the potential loss of points in a particular class.

Two previous studies have attempted to ascertain if loss aversion could be used to improve student performance. In the first such study, students in the loss frame under performed those in the control state (Bies-Hernandez, 2012). However, this study made no attempt to control for student characteristics/abilities; the results were derived by comparing the raw performance indicator of the three treatment class sections to the three control sections.

Some effort has been made to determine the effect of loss aversion on classroom performance when controlling for some student characteristics (Apostolova-Mihaylova, Cooper, Hoyt, & Marshall, 2015). This study did not find a statistically significant effect from presenting grades in a loss frame. However, the study suggested that female students were negatively impacted by the treatment while male students were positively impacted by the treatment.

Apostolova-Mihaylova et al. (2015) suffered from a number of empirical shortcomings that this study attempts to rectify. The previous studies had no mechanism to control for student domain knowledge and relied on unusual controls for mathematics ability; no attempt was made to control for selection due to the consent process. This study uses a path-dependant course structure to control for student ability in a given subject domain; when using the larger dataset that lacks a path-dependant course, the study accounts for mathematics ability with ACT scores. Moreover, this study utilizes selection-correction models that account for bias that could be introduced by the consent process itself. This leads to our first hypothesis:

Hypothesis 1 *When properly controlling for student ability, students who experience grades in a loss frame (counting down) will outperform those who are in a gain frame (counting up).*

In the course of this investigation we will also determine if our dataset shows evidence of a heterogeneous-sex effect from the loss aversion treatment when controlling for student ability as was found in Apostolova-Mihaylova et al. (2015). This can be stated as our second hypothesis:

Hypothesis 2 *When properly controlling for student ability, the two sexes will be differentially impacted by the loss frame, resulting in statistically significant difference in the average effect size by sex.*

The authors suspect that the heterogeneous impact on the two sexes found by Apostolova-Mihaylova et al. (2015) was the result of methodological issues: Most notably, a very small number of female students in the study and less accurate controls for student ability. Therefore, we do not expect to confirm hypothesis 2.

Table 1
List of Study Courses

	Term	Enrolled	Identified as Consented	In Admin. Records	Has ACT
Accounting II (ACCT 2020)					
- Treatment	Spring 2016	45	35	35	25
- Control	Spring 2016	44	39	38	33
Macroeconomics (ECON 2220)					
- Treatment	Fall 2015	49	41	33	30
- Control	Fall 2015	50	47	42	40
Microeconomics (ECON 2200)					
- Treatment	Spring 2015	45	38	32	26
- Control	Spring 2015	48	45	40	37
Business Law (LAWS 3930)					
- Treatment	Spring 2016	21	19	19	13
- Control	Spring 2016	26	23	21	13

"Enrolled" is the number of students enrolled in the course at the end of the semester. "Identified as Consented" is the number of students who filled out the IRB consent form which matched an entry in the gradebook. "In Admin. Records" is the number of students who entered a student ID on the IRB consent form that the Office of Institutional Effectiveness could match in administrative records. "Has ACT" is the number of students with ACT scores of those in the administrative records.

Method

Participants

The courses included in our study were two sections each of Principles of Microeconomics (ECON 2200), Principles of Macroeconomics (ECON 2220), Principles of Accounting II (ACCT 2020), and Business Law Fundamentals (LAWS 3930). Each course was taught in the College of Business Administration (CBA) at a regional metropolitan university in the Midwest (University X). In total, there were 217 participants who (a) consented, (b) existed in the administrative records, and (c) the registrar had a recorded ACT score. See Table 1 for a breakdown.

The quasi-experiment had minimal variation in the student population. All four courses in this quasi-experiment were part of CBA's core curriculum, such that each course was required of students in all concentrations within the College (e.g. Accounting, Finance, Economics). Further, the courses were not generally required outside of Business and Economics. Some students might take one of the courses as an elective, but these students are rare, particularly in Macroeconomics and Accounting II because of the prerequisite requirements.

At the time of the quasi-experiment, Macroeconomics and Accounting II were part of

Table 2

Means of Variable Observations by Course and Section – Sequenced Courses

	Obs.	Final % Score	Prior Sequence Grade	GPA	Age	Earned Credits	% Male
Accounting II							
- Treatment	35	0.8591	2.8569	3.3050	20.9429	48.4571	0.5429
- Control	38	0.8650	2.9918	3.4093	20.1053	49.4737	0.5000
Macroeconomics							
- Treatment	33	0.8844	2.9191	3.1048	20.2121	36.6061	0.6970
- Control	42	0.8237	2.9131	3.1625	20.1667	35.1905	0.5714

Mean values of consenting students in Macroeconomics (ECON 2220) and Accounting II (ACCT 2020). Both classes are part of a sequence where all Macroeconomics students must have already taken Microeconomics and all Accounting II students must have already taken Accounting I. "Final % Score" is the final percentage earned in the course. "Prior Sequence Grade" is the grade earned in the earlier Economics or Accounting course on a four point scale. "GPA" is the student's overall college GPA prior to entering the course, and "Earned Credits" is the number of college credits earned prior to entering the course. Overall, we have 85 males and 63 females in this dataset. Race information was not collected as the university is overwhelmingly white and a race variable could uniquely identify a student.

a class sequence where all Macroeconomics students must have already taken Microeconomics and all Accounting II students must have already taken Accounting I. This results in two beneficial effects for our study. First, it further reduced the heterogeneity of the student population: any student taking either course as an elective must have also taken the prerequisite course (i.e., Microeconomics or Accounting I) as an elective prior to enrolling in the course. Such a student is very rare. Second, because all students in the two sequenced courses must take a class in the subject matter before entering the observed course, we observed a grade for each student in the same field (or domain) as the quasi-experimental course. We refer to Macroeconomics and Accounting II as the "sequenced courses" in our quasi-experiment. A total of 148 participants were in this sub-sample.

Table 2 shows the mean value for each variable across the 148 observations in the sequenced courses (68 treatment, 80 control). We used these observations in our sequenced class specifications. While mean values differed substantively from course to course, they did not vary greatly from treatment to control within a given course. "Final % Score" represents the final percentage earned, from 0 to 1, in the given course. "Prior Sequence Grade" is the grade in Accounting I or Microeconomics on a four point scale (analogous to the typical GPA scale). "GPA" is the student's overall college GPA prior to entering the course, and "Earned Credits" is the number of college credits earned before the beginning

Table 3

Means of Variable Observations by Course and Section – All Courses

	Obs.	Final % Score	GPA	Age	Earned Credits	ACT Math	ACT Comp	% Male
Accounting II								
- Treatment	25	0.8616	3.3521	19.4800	48.6800	24.0000	24.7200	0.5200
- Control	33	0.8612	3.3943	19.8485	52.0000	24.2424	24.7576	0.4848
Macroeconomics								
- Treatment	30	0.8756	3.0464	20.2000	38.2000	23.1333	23.4000	0.7000
- Control	40	0.8256	3.1955	19.9750	33.9750	23.9500	24.0250	0.5500
Microeconomics								
- Treatment	26	0.6192	2.6554	20.3243	43.4615	21.9615	22.6539	0.6538
- Control	37	0.6830	2.9737	20.0385	36.2973	24.3243	24.1351	0.6216
Business Law								
- Treatment	13	0.8105	2.8552	21.3077	76.0000	23.3077	22.5385	0.6154
- Control	13	0.7878	3.0977	20.6923	79.4615	24.6923	23.5385	0.8462

Mean values of consenting students with ACT scores. "Final % Score" is the final percentage earned in the course. "GPA" is the student's overall college GPA prior to entering the course, and "Earned Credits" is the number of college credits earned prior to entering the course. "ACT Math" and "ACT Comp" represent the ACT math and composite score for the student. Overall, we have 131 males and 86 females in this dataset. Race information was not collected as the university is overwhelmingly white and a race variable could uniquely identify a student.

The average difference in final score percentage between the treatment and control groups is approximately 0.001; when limiting observations to just those with ACT scores, this difference is about 0.002.

of the semester. Finally, we provide the percent of the class that is male to show the sex breakdown by class. All observations were reported to us as either male or female from the administrative data. While the column was named "gender" in the administrative data, we refer to this column as "sex" as the data was generated from the student's university application form where they were only offered 'male' and 'female' as options.

Table 3 shows mean values of all students with ACT scores in all courses (217 observations; 94 treatment, 123 control). In addition to the columns reported in Table 2, we report the average ACT Math and Composite score for each section. We do not report "Prior Sequence Grade," because Microeconomics and Business Law do not have a prior sequence course. While these mean values varied substantively from course to course, they varied far less between treatment and control of any given course. Students often take these courses in a prescribed order; this results in less variation in age, earned credits, and GPA (as the GPA contains many of the same classes).

Materials and Procedure

For each course included in the study, two sections of the same course were taught in the same semester. For both sections, the course grade was administered using total points rather than percentage grades. Each assignment was worth a fixed number of points, and the point value indicated the value of the assignment relative to the final grade. In the control sections (i.e. counting up), students began the term with 0 points and were awarded points after completing assignments or exams. Their cumulative grade was the accumulation of these points at the end of the semester, which was then compared to the grade scale to determine the course grade.

In the treatment sections (i.e. counting down), students were endowed with the full number of possible points on the first day of the semester. When students completed each assignment or test, they had points taken away when their work merited less than full credit. After all graded events, the student's remaining points were compared to the same grade scale as the control section to determine the final grade. In both treatment and control sections, homework and exams were worth the same amount of points, and the same homework assignments and exams are administered independent of section.

All assignments and tests were identically timed across sections for each subject. The grading scale was identical across section pairs, with the exception of one section administering points in the more traditional pattern, and the other section administering points to exploit students' potential loss aversion. There was no indication of the section's grading policy or that it was part of a quasi-experiment when the students registered for classes. The treatment section of each class-pair was assigned at random. No student switched from a control to treatment section (or vice versa) after the grading policy was revealed.

The only observable difference between the control and treatment sections was the point allocation method, where treatment sections frame full points in the course as the baseline reference point, and control sections frame zero points as the baseline reference point. In order to minimize the variation between each pair of sections, each course was taught by the same instructor in an identical classroom; each course was offered as a two-day-per-week, 75-minute class starting between 10:30 AM and 1:30 PM.

In order to obtain consent in each of the eight sections, a table was set up outside one of the classroom doors where the students took the final exam. Compliance with IRB protocol required a specific table setup. First, in all sections, the table was set up such that a student could walk out of one of the two doors without walking past the table. Second, all consent forms were collected by the principal investigator (PI) on the IRB protocol; the PI was not the instructor of record in any of the courses involved in the study. Third, students were informed that their consent would not be known to their instructor until after grades were submitted and thus their consent could not influence their final grade. Finally, student IDs were collected without collecting names as a final protection of our students' identities. We have been in contact with our IRB administrator throughout this quasi-experiment and analysis, and they approve of all procedures employed in this study.

Data Analysis

Due to length, our data analysis methods have been dramatically shortened from the working paper version of this study. Needless to say, our statistical approach has been verified with selection-correction models (Heckman, 1976), alternative estimators and estimates on sub-samples of the data. For this article, however, we will focus exclusively on our primary regression results. For a more complete description of all methods employed in this study, please see the working paper posted on SSRN:
`http://dx.doi.org/10.2139/ssrn.3048028`.

Primary specification. We estimated the following empirical model using Ordinary Least Squares for our primary specification:

$$\text{Final Class Score}_{ij} = \hat{\beta}_0 + \hat{\beta}_1 \text{Counting Down}_{ij} + \text{FE}_j + \hat{\beta}_2 \mathbf{X}_i + \varepsilon_{ij} \tag{1}$$

where "Final Class Score_{ij}" is the final score of student i in course j ranging from 0-1, "Counting Down_{ij}" is binary, taking a value of one if student i's section of course j implemented the counting down grading procedure, and "FE_j" is the course fixed effect to control for course and instructor effects (omitted for Accounting II as it is the baseline course). $\mathbf{X}_i$ is a set of controls that acts as proxy for student i's ability in field j.

As it is possible that a student's selection of class time of a given course is non-random, we used the following variables to control for ability in the subject matter of the course: GPA, prior sequence grade, ACT scores, earned credits, age, and sex. GPA is the weighted average college GPA of each student at the time they entered the course. While GPA is highly correlated to performance in any course, it is a measure of overall ability, not ability in any particular field. Moreover, the student's GPA is sensitive to the difficulty of their course of study (which may vary across students). Prior sequence grade is the most direct measure of the student's ability in the subject of the course. Unlike GPA, which is a noisy signal of a student's overall ability, prior sequence grade is generated by students' performance in the same course for all students in the subject area of the current course. ACT scores were included as an instrument common to all students in the second and third specifications. In the specifications without a prior sequence grade variable, ACT scores partially controlled for differences in the data generating process of the GPA. Finally, earned credits, age, and sex are maturity proxies that captured differences across sections that might not be otherwise absorbed by prior control variables. We included these maturity proxy variables only for completeness.

The differential impact of the treatment by sex specification. The work of Apostolova-Mihaylova et al. (2015) indicated that female students might perform worse under the inverted grading method, as opposed to the male students. Using our dataset, we were interested in replicating this result. In order to estimate the Apostolova-Mihaylova et al. model, we modified our reduced form from equation 1 as follows:

$$\begin{aligned}\text{Final Class Score}_{ij} = {} & \hat{\beta}_0 \\ & + \hat{\beta}_1 \text{Counting Down}_{ij} + \hat{\beta}_2 \text{Female}_{ij} \times \text{Counting Down}_{ij} \\ & + \text{FE}_j + \hat{\beta}_3 \mathbf{X}_i + \varepsilon_{ij}\end{aligned} \tag{2}$$

where "$\text{Female}_{ij} \times \text{Counting Down}_{ij}$" takes a value of one when a student is female and in a treated section. We used this reduced form equation to determine if we find the same differential treatment effect as Apostolova-Mihaylova et al. (2015); β_2 was the coefficient of interest for this specification.

Results

Primary Specifications

In Table 4 we present three regression results using equation 1. Moving from the left to the right of the table, the first specification used only the observations from the two sequenced classes. To use all observations, we could not include the ACT controls as they were absent for 20 students. The second (middle) specification used all observations in the sequenced classes where the students had a recorded ACT score. Finally, the specification on the far right included all students with ACT scores from the four classes. Because all four classes were used, we cannot utilize the "Prior Sequence Grade" control.

The results in Table 4 indicate that counting down is correlated with a 2.6 to 4.2 percentage point increase in final score. These results are significant at the 1% level using the specifications with the prior sequence class grade control and at the 5% level with all class observations.

The Differential Impact of the Treatment by Sex

We present our estimates from equation 2 in Table 5. The results indicate that for male students exposed to the counting down treatment, there was a 4 to 4.7 percentage

Table 4
Regression Results

	Sequence[a]		Sequence w/ ACTs[a]		All w/ ACTs[b]	
	Coef.	B. P-Value	Coef.	B. P-Value	Coef.	B. P-Value
Counting Down ($\hat{\beta}_1$)	0.0354	0.0020	0.0422	0.0008	0.0257	0.0190
Prior Sequence Grade	0.0671	0.0000	0.0581	0.0000		
GPA	0.0511	0.0000	0.0474	0.0016	0.0772	0.0000
Earned Credits	0.0005	0.1580	0.0003	0.4860	0.0000	0.9086
Male	0.0071	0.5158	-0.0055	0.7076	0.0068	0.5764
Age	-0.0001	0.9344	0.0053	0.4694	0.0126	0.0012
Macroeconomics Fixed Effect	0.0069	0.6026	-0.0005	0.9898	0.0067	0.6284
ACT Math			0.0075	0.0104	0.0071	0.0018
ACT Composite			-0.0028	0.2994	0.0035	0.1058
Microeconomics Fixed Effect					-0.1603	0.0000
Business Law Fixed Effect					-0.0451	0.0040
Observations		148		128		217
Adjusted R^2		0.4395		0.4648		0.6325[c]

For all three regressions, the dependent variable is the final class score. "Coef." is the coefficient estimate, and "B. P-Value" is the bootstrapped p-value. In all models presented, a Box-Cox transformation was estimated and in all versions a linear model was the closest common functional form. Pairwise bootstrap results are presented in the table due to normality issues. However, all conclusions hold using corrected standard errors. The baseline student is a female in the control group of the Accounting II course. An intercept (or constant) term was used in all regressions, but the results are omitted for brevity. [a] The two sequence courses are Macroeconomics and Accounting II where there is a prior sequence course (respectively, Microeconomics and Accounting I) to control for ability in the specific subject area. We present two versions of this model, one without ACT controls with all observations and one with ACT controls but with 20 fewer observations due to unavailable ACT scores.
[b] The third specification contains all four courses but must drop "prior sequence grade" as a control variable as Microeconomics and Business Law have no prior sequence course. We control for ability with the addition of ACT controls. The p-values in this specification are the result of a block bootstrap procedure (Künsch, 1989). Some students were enrolled in both Microeconomics and either Accounting II or Macroeconomics. Therefore, we have 206 unique students. To maintain the bootstrap's independence assumption, we drew students (with replacement) and extract the corresponding row(s). This procedure produces results nearly identical to the standard pairwise bootstrap. We present these results instead of the pairwise bootstrap as they are technically more correct. [c] This adjusted R^2 value includes variation introduced by the large difference in difficulty of the two non-sequence courses. This variation was then explained by the course fixed effect resulting in an increased adjusted R^2 value.

point positive impact ($\hat{\beta}_1$) on the final score; using the sequenced course specifications where we controlled for domain knowledge, we saw a positive 3 to 3.6 percentage point effect ($\hat{\beta}_1 + \hat{\beta}_2$) for female students in response to the counting down treatment (an effect that is jointly statistically significant based on testing the significance of $\hat{\beta}_1 + \hat{\beta}_2$, see table note *c* of Table 5). In the final specification, males were impacted by the treatment by about 4 percentage points and female students were not impacted at all. However, in all three specifications the female/treatment interaction term is statistically insignificant at the 5% level (and insignificant at any conventional level in two specifications).

Discussion

Our results suggest that inverting the gradebook and placing the students into a loss frame results in an increase in student performance by about a half-letter grade. While our

Table 5

Regression Results with a Sex Interaction Term

	Sequence[a]		Sequence w/ ACTs[a]		All w/ ACTs[b]	
	Coef.	B. P-Value	Coef.	B. P-Value	Coef.	B. P-Value
Counting Down[c] ($\hat{\beta}_1$)	0.0395	0.0126	0.0468	0.0108	0.0422	0.0066
Female×Counting Down[c] ($\hat{\beta}_2$)	-0.0098	0.6958	-0.0109	0.6984	-0.0424	0.0580
Prior Sequence Grade	0.0669	0.0000	0.0578	0.0000		
GPA	0.0512	0.0002	0.0474	0.0018	0.0771	0.0000
Earned Credits	0.0005	0.1548	0.0003	0.4746	0.0001	0.8556
Male	0.0028	0.8400	-0.0100	0.6110	-0.0113	0.4668
Age	-0.0002	0.9108	0.0053	0.4604	0.0126	0.0024
Macroeconomics Fixed Effect	0.0067	0.6184	-0.0006	0.9836	0.0062	0.7024
ACT Math			0.0074	0.0104	0.0069	0.0024
ACT Composite			-0.0026	0.3188	0.0040	0.0650
Microeconomics Fixed Effect					-0.1596	0.0000
Business Law Fixed Effect					-0.0425	0.0056
Observations		148		128		217
Adjusted R^2		0.4362		0.4610		0.6367[d]

For all three regressions, the dependent variable is the final class score. "Coef." is the coefficient estimate, and "B. P-Value" is the bootstrapped p-value. In all models presented, a Box-Cox transformation was estimated and in all versions a linear model was the closest common functional form. Pairwise bootstrap results are presented in the table due to normality issues. However, all conclusions hold using corrected standard errors. The baseline student is a female in the control group of the Accounting II course. An intercept (or constant) term was used in all regressions, but the results are omitted for brevity. [a] The two sequence courses are Macroeconomics and Accounting II where there is a prior sequence course (respectively, Microeconomics and Accounting I) to control for ability in the specific subject area. We present two versions of this model, one without ACT controls with all observations and one with ACT controls but with 20 fewer observations due to unavailable ACT scores.
[b] The third specification contains all four courses but must drop "prior sequence grade" as a control variable as Microeconomics and Business Law has no prior sequence course. We control for ability with the addition of ACT controls. The p-values in this specification are the result of a block bootstrap procedure (Künsch, 1989). Some students were enrolled in both Microeconomics and either Accounting II or Macroeconomics. Therefore, we have 206 unique students. To maintain the bootstrap's independence assumption, we drew students (with replacement) and extract the corresponding row(s). This procedure produces results nearly identical to the standard pairwise bootstrap. We present these results instead of the pairwise bootstrap as they are technically more correct. [c] In the two sequence specifications, we tested if $\hat{\beta}_1 + \hat{\beta}_2$ was jointly statistically different from zero. In the sequenced courses without ACT controls, the joint test produced a p-value of 0.0598. When including the ACT controls the resulting p-value was 0.0432. A significance test for the right-most specification isn't necessary as the joint effect is $\hat{\beta}_1 + \hat{\beta}_2 = -0.0002$ and clearly not statistically significant. [d] This adjusted R^2 value includes variation introduced by the large difference in difficulty of the two non-sequence courses. This variation was then explained by the course fixed effect resulting in an increased adjusted R^2 value.

results have withstood numerous robustness checks, the limitations of this study must be acknowledged. First, while we have attempted to control for all possible differences in the classes, it is always possible that an unobserved variable was responsible for the treatment effect. Second, the treatment effect could be the result of novelty, particularly if a student would typically use heuristics to decide how much effort to expend on a given graded event.

As described in Kahneman (2003), system one thinking is based on heuristics. It is highly likely that the quasi-experimental course is the first class the treated student has taken with an inverted grading scheme. Because this is a new grading scheme for the student, their established heuristics would not apply, forcing the student to switch to a more systematic approach. Abd-El-Fattah (2011) and Hacker, Bol, and Bahbahani (2008) suggested that this study's results might be from increases in the ability of students to calibrate accurately. It may be that loss aversion requires more attention on the part of the students, causing them to be more thoughtful as they determine their effort on any given assignment or exam. If counting down were widely adopted, we might see the positive effect diminish with repeated treatment. Determining if this alternative explanation has merit would be difficult. However, it could be achieved with a long term study where the students are consistently graded using the counting down method throughout their course of study.

In contrast with Apostolova-Mihaylova et al. (2015), our results suggest limited evidence of a differential impact of the treatment on the two sexes. Some statistical considerations must be taken into account when examining these two studies. Like most Business and Economics courses, both our sample and the sample studied in Apostolova-Mihaylova et al. (2015) are dominated by male students. In our regression results using sequenced courses, 57% of the observations come from male students (56% when using ACT scores); in our specification using all courses, 60% of the observations come from male students. The sample in Apostolova-Mihaylova et al. (2015) was 69% male observations.

Additionally, a substantive difference in how men and women react to the treatment in our model occurs only where a control for domain knowledge (prior sequence grade) is omitted; even then, the sex difference term is insignificant at the 5% level. A power analysis is uninteresting in our primary specification, as we reject the null hypothesis of no effect (thus type II error cannot have occurred), but it may be helpful for this model, since we have failed to find a statistically significant female/treatment interaction effect. We performed a simple power analysis assuming the point estimate in Apostolova-Mihaylova et al. (2015) is truth ($\beta_2 = -0.0678$, where the lack of a hat indicates an assumed true underlying value) and using our heteroscedasticity-corrected standard errors, as they are marginally larger than our bootstrapped standard errors. Using results from the two specifications that control for domain knowledge, we find a 20-25% chance that we falsely accept the null of no effect (75-80% power). More to the point, looking at our observed point estimates for $\hat{\beta}_2$, there is only a 1% (specification one: left columns of Table 5), and 2% chance (specification two: middle columns of Table 5) that we would observe a point estimate this high or higher if we assume that the point estimate in Apostolova-Mihaylova et al. (2015) is truth ($P(\hat{\beta}_2 \geq -0.0098|\beta_2 = -0.0678) = 0.0091$, and $P(\hat{\beta}_2 \geq -0.0109|\beta_2 = -0.0678) = 0.0163$). The authors believe the difference in outcomes for the two studies results from the current study's improved control for student ability. This study controls for students' subject-specific domain knowledge/ability through the prior sequence grade. Moreover, in our weakest specification (where prior sequence grade is unavailable) we used ACT scores to control for differences in the GPA data generating process (e.g. math ability). Neither prior sequence grade nor ACT scores were used in the prior study's regression models.

On balance, counting down is positively correlated with the final score in the class. There might be a difference in how men and women react to the inverted grading scheme, but it is unlikely that there exists any overall negative effect for women. Additionally, the literature on loss aversion provides little indication that women would be less loss averse than men: some studies find no heterogeneous sex effect (Abdellaoui, Bleichrodt, & L'Haridon, 2008; Gächter, Johnson, & Herrmann, 2010), while others find that women are more loss adverse (Booij & Van de Kuilen, 2009; Brooks & Zank, 2005; Rau, 2014; Schmidt & Traub, 2002).

Pedagogical Implications

The treatment demonstrated in this study requires no class time and is not tied to any course content. Moreover, the primary result was found using the dataset from four different courses. Some of these courses are highly focused on memorization while other courses are focused on learning about a specific scientific approach. Additionally, loss aversion should impact student motivation and thus should be immune to differences in class content. Therefore, we believe our result is generalizable, acknowledging the methodological caveats discussed at the beginning of this section.

Instructors of all disciplines can apply the approach outlined in this study to motivate their students. However, we suspect the instructor need not be ‘all-in’ to enjoy at least some of the benefits. Instead of offering donuts if the class exceeds a particular exam threshold, one could include the donuts in the syllabus and take them away if the class scores below a given threshold. Similarly, even in a traditionally graded class, the instructor could re-frame all bonus points as a loss; the points could be entered at the beginning of the semester where certain actions (or lack of actions) would result in their loss on particular dates.

Conclusion

In this quasi-experiment, we find that an inverted grading scheme where students are endowed with all potential points at the start of the semester might result in increased student performance. With strong ability controls in our sequenced classes, we find little evidence of a substantial or statistically significant sex-based difference. The evidence from our quasi-experiment suggests that both sexes benefit from counting down, but that male students might benefit more.

While this study provides evidence of increased student performance using inverted grading, it provides no information about student perceptions. Perceptions are an important consideration to the adopting instructor, particularly because of the frequent use of student evaluations in measuring job performance. Even when this is not the case, positive perceptions of introductory courses are likely associated with an increased ability to attract students to a particular discipline. While outside the scope of this project, the authors believe a study focusing on how students perceive inverted grading would be helpful to both instructors adopting the grading scheme as well as departments.

References

Abd-El-Fattah, S. M. (2011). The effect of test expectations on study strategies and test performance: A metacognitive perspective. *Educational Psychology*, *31*(4), 497-511.

Abdellaoui, M., Bleichrodt, H., & L'Haridon, O. (2008). A tractable method to measure utility and loss aversion under prospect theory. *Journal of Risk and Uncertainty*, *36*(3), 245-266.

Apostolova-Mihaylova, M., Cooper, W., Hoyt, G., & Marshall, E. C. (2015). Heterogeneous gender effects under loss aversion in the economics classroom: A field experiment. *Southern Economic Journal*, *81*(4), 980-994.

Bean, J. P., & Bradley, R. K. (1986). Untangling the satisfaction-performance relationship for college students. *The Journal of Higher Education*, *57*(4), 393-412.

Bies-Hernandez, N. (2012). The effects of framing grades on student learning and preferences. *Teaching of Psychology*, *39*(3), 176-180.

Booij, A. S., & Van de Kuilen, G. (2009). A parameter-free analysis of the utility of money for the general population under prospect theory. *Journal of Economic Psychology*, *30*(4), 651-666.

Brooks, P., & Zank, H. (2005). Loss averse behavior. *Journal of Risk and Uncertainty*, *31*(3), 301-325.

Cathey, C. L., Sly, J. S., & Barry, H. L. (2018). To randomize or not: Examining effects of exam item order. *Scholarship of Teaching and Learning in Psychology*, *4*(2), 115-119.

Cerbin, W. (2018). Improving student learning from lectures. *Scholarship of Teaching and Learning in Psychology*, *4*(3), 151-163.

Clayson, D. E., Frost, T. F., & Sheffet, M. J. (2006). Grades and the student evaluation of instruction: A test of the reciprocity effect. *Academy of Management Learning & Education*, *5*(1), 52-65.

Gächter, S., Johnson, E. J., & Herrmann, A. (2010). *Individual-level loss aversion in riskless and risky choices* (Tech. Rep.). CeDEx discussion paper series.

Grimes, P. W. (2002). The overconfident principles of economics student: An examination of a metacognitive skill. *The Journal of Economic Education*, *33*(1), 15-30.

Hacker, D. J., Bol, L., & Bahbahani, K. (2008). Explaining calibration accuracy in classroom contexts: The effects of incentives, reflection, and explanatory style. *Metacognition and Learning*, *3*(2), 101-121.

Heckman, J. J. (1976). The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. In *Annals of economic and social measurement* (Vol. 5, p. 475-492).

Hoyert, M. S., O'dell, C. D., & Hendrickson, K. A. (2012). Using goal orientation to enhance college retention and graduation rates. *Psychology Learning & Teaching*, *11*(2), 171-179.

Kahneman, D. (2003). Maps of bounded rationality: Psychology for behavioral economics. *The American Economic Review*, *93*(5), 1449-1475.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica: Journal of the Econometric Society*, *47*, 263-291.

Künsch, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *The annals of Statistics*, 1217-1241.

Lo, C. C. (2010). How student satisfaction factors affect perceived learning. *Journal of the Scholarship of Teaching and Learning*, *10*(1), 47-54.

Quinn, P. D., & Duckworth, A. L. (2007). Happiness and academic achievement: Evidence for reciprocal causality. In *The annual meeting of the american psychological society* (Vol. 24).

Rau, H. A. (2014). The disposition effect and loss aversion: Do gender differences matter? *Economics Letters*, *123*(1), 33-36.

Robinson, C., Lee, M. G. L., Dearing, E., & Rogers, T. (2017). *Reducing student absenteeism in the early grades by targeting parental beliefs* (Tech. Rep.). HKS Faculty Research Working paper.

https://research.hks.harvard.edu/publications/workingpapers/Index.aspx.

Schmidt, U., & Traub, S. (2002). An experimental test of loss aversion. *Journal of Risk and Uncertainty*, *25*(3), 233-249.

Smith, B., White, D., Kuzyk, P., & Tierney, J. (2018). Improved grade outcomes with an e-mailed "grade nudge". *The Journal of Economic Education*, *49*(1), 1-7.

Tversky, A., & Kahneman, D. (1991). Loss aversion in riskless choice: A reference-dependent model. *The Quarterly Journal of Economics*, *106*(4), 1039-1061.

Footnotes

[1]See `https://www.bi.team/our-work/publications/` for a list of publications based on the work of the Behavioural Insights Team