I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 12983

Ask and You Shall Receive? Gender Differences in Regrades in College

Cher Hsuehhsiang Li
Basit Zafar

FEBRUARY 2020

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 12983

Ask and You Shall Receive? Gender Differences in Regrades in College

Cher Hsuehhsiang Li
Colorado State University and IZA

Basit Zafar
Arizona State University and IZA

FEBRUARY 2020

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Ask and You Shall Receive? Gender Differences in Regrades in College*

Using administrative data from a large 4-year public university, we show that male students are 18.6 percent more likely than female students to receive favorable grade changes. These gender differences cannot be explained by observable characteristics of the students, instructors, and the classes. Surveys of students and instructors reveal that regrade requests are prevalent, and that male students are more likely than female students to ask for regrades on the intensive margin. We corroborate the gender differences in regrade requests in an incentivized controlled experiment where participants receive noisy grade signals, and where they can ask for regrades: we find that males have a higher willingness to pay (WTP) for regrades. Almost half of the gender difference in the WTP is due to gender differences in confidence, uncertainty in beliefs, and the Big Five personality traits.

JEL Classification: C9, I2, J7

Keywords: education, gender, inequality, negotiation, experiment

Corresponding author:
Cher H. Li
Colorado State University
1771 Campus Delivery
Fort Collins, CO 80523
USA

E-mail: cher.hh.li@gmail.com

* We are grateful to Colorado State University for providing access to the administrative data. We thank Christine Exley, participants in the Allied Social Science Association Annual Meeting, Asian & Australasian Society of Labour Economics Conference, NBER Education Meeting, Southern Economic Association Annual Meeting, UCLA Anderson, University of Wisconsin-Madison Alumni Conference, Colorado Summer Workshops, and seminar participants at San Francisco State University and University of West Georgia for valuable comments. All errors are our own.

1 Introduction

Women and men differ across a variety of behaviors, including their tendency to negotiate (Bowles et al., 2007; Leibbrandt and List, 2014; Small et al., 2007). The difference in propensity to negotiate has been argued to contribute to the gender gap in labor market outcomes (Babcock and Laschever, 2003; Artz et al., 2018). This project examines whether such gender differences have consequences when individuals are still in school, prior to entry in the labor force. Specifically, we examine whether male and female students experience different rates of successful grade changes in college. If men are more aggressive than women in bargaining for better grades (either on the extensive or intensive margin), they may be more likely to convince their instructors to alter their grades which serve as productivity signals to potential employers. Gender differences in willingness to ask and to negotiate may then put equally capable female students at a relative disadvantage in the job market.

The starting point of our analysis is a unique administrative dataset that contains not only the final grade records but also any grade changes related to the records from a large 4-year public university. The dataset includes different reasons of grade changes that allow us to distinguish changes that are due to student actions, university rules, or instructor initiations. Assuming that the distribution of grading errors is the same for both male and female students, we would expect to observe a similar grade correction pattern initiated by instructors for both male and female students. The overwhelming majority of grade changes (95 percent) led to an improvement in the grade. Our analysis based on the administrative records reveals that although women made up 53.4 percent of the grade records, they represented only 49.1 percent of the favorable grade changes initiated by instructors. The gender differences in students' grade changes persist across colleges and departments, and are robust to inclusion of class and student characteristics.

Grade changes in the administrative records are, unsurprisingly, rare events: less than 0.5 percent of grades are changed. The administrative data is not very informative of the prevalence of regrading requests. By construction, it contains no information on any requests that students made but resulted in no grade change. In addition, students might have asked for regrades all along the semester—analysis of the administrative data by itself is likely to underestimate the extent of regrading requests and gender differences in them. Therefore, the small fraction (less than 0.5%) of grade changes and its consequential effects on GPA (an increase of 0.0035 points for females and 0.0041 points for males) observed in the administrative records should be considered a lower bound for regrade requests and changes in GPA. Administrative data is also limited in what it can tell us about the origins of the gender differences in regrades. For example, they may be a result of at least three distinct

scenarios: 1) male students are more likely than female students to ask instructors for grade changes although instructors treat all requests equally (i.e., a gender difference in willingness to ask); 2) the propensity to ask is the same for both male and female students, but the outcomes are more favorable for males than for females when they ask (i.e., discrimination); and 3) female students make regrade requests *during* the semester which in turn lowers their demand for regrade requests at the *end* of the semester. Administrative records do not contain any information that allows us to understand which scenario accounts for these unequal outcomes. Knowing the origin of these differences is relevant for policy. For example, evidence consistent with the second scenario would be indicative of some kind of preferential treatment towards males. Evidence consistent with the first scenario would be suggestive of a preference-based explanation.

To understand these patterns, we next administered surveys to instructors and students. These surveys elicited respondents' recollections on regrade requests in the past semesters. Forty percent of students reported that they approached instructors for grade changes *at some point* during their college life. We treat this high propensity of asking (i.e. 40%) as the upper bound of regrade requests. Conditional on asking, over 60 percent of the requests lead to an improvement in some grade component (e.g. exams, quizzes, and assignments), and over 30 percent of these requests result in better final grade outcomes for students. Therefore, these regrade requests may have a profound impact on grades for students. Indeed, using the statistics reported by students, our back-of-the-envelope calculation shows that the upper bound of the increase in GPA conditional on asking could be as high as 0.43 points for females and 0.47 points for males. Importantly, the high prevalence of regrade requests in college suggests that restricting the analysis to the administrative data would have shown a very incomplete picture. The surveys reveal that there is little gender difference in the propensity of males and females to ever ask (or consider asking) for regrades. However, male students ask (or consider asking) for regrades in a larger number of classes. That is, there is a gender difference on the intensive margin. We also find that these request patterns persist throughout the semester. Hence, even if instructors change the grades for male and female students at the same rate, the outcome may still favor male students simply because they ask more frequently. We also find no gender difference in the expected change in grades, conditional on asking.

To investigate why men and women differ in their propensity to ask, we next rely on an incentivized controlled experiment where payoff depends on one's revealed performance in a quiz. Students are, however, initially given noisy signals of their quiz performance. They are then given the option to request a regrade at varying levels of costs. Specifically, we used a multiple-price-list variation of the Becker-DeGroot-Marschak method (Andersen et al., 2006)

to elicit the individual's willingness to pay (WTP) to ask for a regrade: individuals are asked to decide whether they want to make a regrade request at 10 varying cost scenarios, with the costs varying from the respondent paying \$3.50 to −\$1.00 (that is, getting paid \$1.00) in increments of \$0.50. If a participant decides to pay a given cost to ask for a regrade, the true grade is revealed. Regrades potentially affect not only the final shown grade but also the student's monetary payoff. The laboratory setting is advantageous over observational field evidence because it allows us to quantify gender differences in the WTP to ask, and to investigate its correlates.

Our results are striking. There are no gender differences in the willingness to ask when the cost of doing so is zero or negative (that is, when the respondent is being paid to ask). However, a large gender difference emerges once the cost of asking is positive: nearly half of males are willing to pay a positive cost for regrades, versus only a third of females. As a result, the average WTP differs by gender: it is \$0.25 for males, compared with \$0.05 for females (the difference is significant at the 5-percent level). We also validate our experimental measure of WTP by showing that it positively correlates with the survey data on regrade considerations in actual courses, that is, students who exhibit a higher WTP to ask in the experiment are also more likely to have considered asking their instructors for a regrade. The higher willingness to ask on the part of men, however, does not necessarily translate into increased monetary returns for them once the regrading costs are taken into account. For low-cost values, men's higher propensity to ask for regrades does make them better off. However, at high-cost values, men's higher propensity to ask can actually make them worse off. As a result, males are both more likely than females to see a payoff increase when the cost is low and a payoff decrease when the cost is high. This is consistent with the survey evidence from instructors that male students are much more aggressive in their requests, and with the survey evidence from the students themselves where male students are more likely to report regrets for asking too aggressively for regrades.

So what explains the gender difference in the WTP to ask? We see that females, relative to males, have lower average beliefs about their score on the test. However, the gender gap in beliefs mimics the gender gap in actual performance that we observe on the quiz. It is worth noting that both genders overestimate their performance on the quiz.[1] We find that females are, on average, more under-confident and more uncertain about their performance than males. This is consistent with prior evidence that men exhibit a greater degree of overconfidence in their relative ability (Barber and Odean, 2001; Niederle and Vesterlund, 2007) or that females tend to be less confident (Sarsons and Xu, 2015). Likewise, consistent

[1]It is indeed the case that the WTP to ask is positively related to the gap between the guessed score and the shown score; the relationship does not differ by gender.

with prior literature (Croson and Gneezy, 2009; Eckel and Grossman, 2008), we find that females are more risk averse. Our survey data reveal systematic gender differences in the Big Five personality traits, with females being more agreeable and neurotic; this is consistent with prior literature (Costa Jr et al., 2001; Weisberg et al., 2011). We find that gender differences in confidence, uncertainty in beliefs and personality traits can explain 43 percent of the gender difference in the propensity to ask; we find a small role of risk preferences. Nonetheless, the majority of the gender gap remains unexplained, suggestive of preference-based explanations.

Our survey provides further evidence on why females are less likely to ask. They are significantly more likely to report higher psychic costs—specifically, fear of rejection and stress—when having to ask an instructor for a regrade. Thus, policies that make regrading policies explicit and lower the (psychic) costs of asking could reduce gender differences in asking. In addition, reducing under-confidence and uncertainty in beliefs (through policies that provide more accurate signals to individuals) is likely to reduce gender differences in the WTP to ask.

It is rather remarkable that across three very different types of data (administrative, survey, and experimental), we get consistent results. We view that as a contribution in itself. Our paper is related to the literature on gender differences in asking. Empirical findings often show that women are much less likely than men to initiate negotiations and attain less favorable outcomes when they do negotiate. For instance, Small et al. (2007) find that women were more likely than men to accept a low reward for their participation in a study without bargaining. Similarly, Babcock and Laschever (2003) find that among graduates from a prestigious MBA program, only 7 percent of the women negotiated their wage offers, while 57 percent of the men did so, contributing to the gender wage gap in the starting salary of the MBA graduates. Leibbrandt and List (2014) conducted a field experiment by posting job advertisements and observed the negotiation behaviors of real job applicants. They find that when salaries were not explicitly made negotiable in the advertisement, men were significantly more likely to negotiate for a higher wage. However, the gender differences in wage negotiation disappeared when salaries were posted as negotiable. Exley et al. (2019) find that women attain worse returns from negotiation when it is mandatory than when it is optional, suggesting that women know when to ask. This has some parallels with our finding that women are less likely to see a payoff decrease than men (for positive costs of asking). However, we also find that females are less likely to experience payoff gains when the costs of asking are low.

We complement this literature by showing that females are simply less likely to ask, especially when there is a positive cost for doing so. Our rich data also allow us to shed light

on the underlying reasons for the gender gap. Prior studies primarily focus on the impact of gender differences in negotiations on outcomes *after* individuals enter the labor force. There is little work on how gender differences in negotiations and willingness-to-ask *prior* to labor market entry may impact outcomes. Our proposed project fills this gap by investigating whether the gender differences in propensity to bargain have implications for outcomes in college. These pre-market gender differences are important because they could have direct implications in potential labor market outcomes for college graduates. Employers frequently require candidates who apply for entry-level positions to provide their transcripts, and many competitive positions require a minimum GPA. The difference in regrades by gender may keep qualifying females below potential employers' GPA threshold. In fact, Kessler et al. (2019) show that college GPAs play a critical role in employers' recruitment assessment—a 1 point increase in GPA results in an increase in employers' hiring interest by 0.8 standard deviations. In an audit study, Quadlin (2018) also finds that the callback rate for job applicants increases from 9.6 percent to 16.3 percent as their GPA increases from 2.50 to 3.59. Hence, gender gaps in regrades could place women in a relative disadvantaged position.[2]

The remainder of the paper is structured as follows. The next section presents the analysis of the administrative data. Section 3 analyzes the survey data that we collected from instructors and students, while Section 4 presents the analysis of the laboratory experiment. Finally, Section 5 concludes with a discussion of the results and their implications.

2 Analysis of Administrative Records

2.1 Data

We analyze a unique administrative dataset from Colorado State University (CSU), a large 4-year public university. CSU was ranked 129^{th} among all public and private universities nationwide in 2016 (US News and World Report, 2016). Fall 2016 enrollment at CSU consisted of 23,768 on-campus undergraduate students. On average, the freshmen admitted in Fall 2016 had a 3.6 high-school GPA, a 25.2 ACT composite score (compared to the national average of 20.8), a 566.5 SAT critical reading score, and a 575.5 SAT math score (compared to national averages of 494 and 508, respectively). Among these freshmen, approximately 56 percent were female and 25 percent were minorities (CSU, The Fact Book 2016–17). Although the gender distribution of CSU freshmen was comparable to the national average (56

[2]We need not take a stand on what exactly grades measure. Whether grades serve as a proxy of unobservable cognitive skills or non-cognitive traits (including negotiation skill) is an interesting question but one that is beyond the scope of this paper. What matters for the relevance of our findings is that grades have consequences for outcomes, including labor market outcomes.

percent), CSU was less ethnically diverse than the average university (43 percent).[3]

Key to our purposes, this administrative dataset records not only the final grades but also any grade changes. In the case of a grade change, a reason code is also provided. We focus primarily on the change reasons including grade entry errors, instructor corrections, and re-calculations—the only three options available to instructors when they submit grade changes in the computing system. The distribution of instructors' grade change reasons is: 26 percent entry errors, 59 percent instructor corrections, and 15 percent recalculations. There are no clear instructions on the choices of the grade change codes and no verification mechanism in place to distinguish the assignment of regrade reasons. Therefore, instructors have the flexibility in assigning the grade change reasons among the three options. For this reason, we treat grade changes based on any of the three reasons as regrades by instructors. The results are robust to exclusion of individual reasons among the three change codes one-at-a-time.[4] If the grade changes are based on students' own actions (e.g., taking a repeat-and-delete option by re-taking the same course to override the original grade with a new grade or receiving a final letter grade for initially incomplete credits) or university rules (e.g., grades were automatically changed to F when students fail to meet the higher requirements for college writing and mathematics courses), they are not considered regrades made by instructors.

2.2 Descriptive Statistics

The administrative dataset contains 1,341,552 credited student-class records with letter grades of 64,857 students taught by 3,726 instructors during the years between 2010 and 2016.[5] Excluding grade updates for the incomplete credits after students completed their work, there were 6,225 grade changes (0.46 percent) made by instructors during this time. Among the grade changes initiated by instructors, 94.6 percent (5,886 records) of the grades were corrected upward (i.e., the initial grade was changed to a better grade). The overwhelmingly large fraction of upward corrections among grade changes indicates that the

[3]The national average test scores were published on the ACT and SAT websites (The ACT Profile Report—National Graduating Class 2016 and 2016 College-Bound Seniors SAT Total Group Profile Report). The source for the national enrollment rate by gender and race is the National Center for Education Statistics, Table 306.10. Total fall enrollment in degree-granting postsecondary institutions, by the level of enrollment, sex, attendance status, and race/ethnicity of student: Selected years, 1976 through 2016.

[4]We also analyze gender differences in regrades by individual reasons (e.g., including only grade changes due to "entry errors" (or "instructor corrections", or "recalculations") and grade records with no change), and find that the gender gap in regrades persists for both "entry errors" and "instructor corrections" but not for "recalculations".

[5]44,481 withdrawal observations are excluded from the sample. The withdrawal rate is higher among males (2.7%) than among females (2.3%). If students withdrew from classes to avoid a low grade which may be subsequently more likely to lead to a regrade request, the withdrawal selection would bias our estimate of the gender gap towards 0.

risk of receiving a downward grade change was relatively small when students made regrade requests. Although women made up 53.4 percent of the grade records, they represented only 49.2 percent of the upward grade changes initiated by instructors. Table 1 shows the summary statistics by students' gender. Conditional on students' gender, the rate of upward grade changes initiated by instructors was 0.479 percent (2,991 records) for male students and 0.404 percent (2,895 records) for female students. Although grade changes were rare events, among these upward changes, the 0.075 percentage points difference represented that men were 18.6 percent more likely than women to receive an upward grade change by instructors. On the other hand, in the extremely sparse events (339 observations) of downward grade corrections (i.e., the initial grade was changed to a lower grade), male students were 50-percent more likely than their female counterparts to receive such an adverse outcome (0.03 percent versus 0.02 percent).

Students showed additional differences in their course records by gender as presented in Table 1. For instance, female students took half of their classes from female instructors, while male students took classes from female instructors only 38.9 percent of the time – potentially due to gender sorting into different college majors where the gender distribution of instructors is also uneven. For an average academic term, female students attained a higher average GPA by 0.2 points when compared with their male counterparts. The gender distribution of students also varies substantially across classes offered by different colleges. Women were more likely than men to take classes offered by the Colleges of Agriculture, Human Sciences, Liberal Arts, Veterinary Sciences, and Intra-University. Nevertheless, the gender gap in grade changes was pervasive across colleges and departments.[6]

Students may have different propensities to request regrades based on the grade they originally received. Figure 1 shows that indeed students who received A+ or A ("A students") as their initial grade are very unlikely to experience a positive grade change: this is mechanical since their grade has already hit the upper bound. Students who received an initial grade of D have the largest probability (1.49 percent) of getting a boost in their final grade, followed by those with an initial grade of C+ (0.92 percent), B− (0.9 percent), A− (0.79 percent) and F (0.72 percent). The relatively high frequencies observed for those receiving an initial grade of D or F are not surprising. Some departments (e.g., Economics) require students to complete their courses at a grade of C and above in order to fulfill the major and/or minor requirements. Furthermore, if a student's average GPA falls below 2.0 (equivalent to a C), the student will be on academic probation and be dismissed if the probation has persisted for two semesters. However, the high relative frequencies of grade changes for those receiving A−, B−, and C+ also indicate that the upward grade changes are not

[6]The gender differences in grade changes by college are detailed in Appendix Tables B1.

limited to failing students. Frequently, students are motivated to argue for a better grade when the grade they received is below their expectations. Indeed, the likelihood of favorable regrades increases when a student's grade is below their average GPA.[7] It is worth noting that the propensities of grade changes are similar in both large classes (i.e., defined as classes with over 90 students) and small classes.[8]

2.3 Empirical Analysis

To examine whether gender differences are present in grade changes among college students, we analyze our data with the following specification:

$$Y_{ijt} = \alpha_0 + \alpha_1 Male_i + \alpha_2 W_{ijt} + \alpha_3 X_{it} + \alpha_4 Z_{jt} + \alpha_5 Grade_{ijt} \times GPA_{it} \times Dept_{jt} + \epsilon_{ijt}, \quad (1)$$

In this equation, Y_{ijt} is a binary variable and takes the value of one if student i received a grade change (e.g., a upward grade change or a downward grade change) in class j during semester t and zero otherwise; W_{ijt} includes time and class varying variables such as the initial grade student i received in class j during semester t; X_{it} are characteristics of student i in semester t, such as the student's class standing (i.e., freshman, sophomore, junior, and senior) and cumulative GPA;[9] Z_{jt} captures class-specific information, such as the instructor's gender and rank (i.e., tenure-track assistant professor, tenure-track associate professor, tenure-track full professor, non-tenure-track instructor, graduate teaching assistant), department, and college; $Grade_{ijt} \times GPA_{it} \times Dept_{jt}$ includes the full interactions of indicators of students' initial grades, discretized cumulative GPAs, and the indicator of departments. We saturate the model with these interactions because the incentive to ask for a regrade may vary across departments (since many majors have certain GPA and course grade requirements); and ϵ_{ijt} is an error term. $Male_i$ is an indicator for male students and α_1 captures any gender differences in the regrades. This is the parameter of interest.

We first present the analysis of equation 1 for upward grade changes in Panel A of Table 2 (i.e., the dependent variable is an indicator that equals 1 if the individual experiences a favorable grade change). Column 1 shows that the raw gender difference (that is, the estimate

[7] We find that the upward grade changes were near zero when the student's performance in a given class was better than her average performance in other classes. However, when the student's grade in a class is below her average GPA, the propensity to receive an upward grade correction increases in the difference between the GPA and the class grade.

[8] Approximately 28-percent of administrative records were from large classes (i.e., classes of more than 90 students) where grade changes occurred 0.45 percent of the time, compared with 0.47 percent for the remaining classes.

[9] In our primary analysis, cumulative GPAs are discretized into 8 intervals corresponding to letter grades: $[0, 1), [1, 2), [2, 2.334), [2.334, 2.667), [2.667, 3), [3, 3.334), [3.334, 3.667)$, and $[3.667, 4]$. Results are robust to alternative continuous values.

of α_1 without any controls) in upward grade change is 0.0748 percentage points. This raw gender difference implies that the male-to-female odds ratio of a favorable regrade outcome is 1.186 (reported in square brackets in the table). In other words, for a male, the odds of receiving an upward grade change are 1.186 times as large as the odds for a female receiving such an outcome.[10] Controlling for students' characteristics (i.e., initial grade, GPA, and class standing) that are unequal across gender (e.g., females tend to have higher grades and GPA and shorter time-to-graduation than males), Column 2 shows that the gender gap reduces to 0.0665 percentage points. In Column 3, controlling for class characteristics (i.e., colleges and departments in which the courses were offered, and the instructors' gender and rank) reduces the gender gap to 0.0685 percentage points which is mainly attributable to the differential rates of grade changes by departments and students' gender-sorting into different majors. In Column 4, controlling for both student and class characteristics fails to explain the gender gap in regrade results. The specification in Column 4 implicitly assumes that the grade change rates are similar for students with the same grade/GPA regardless of which departments offered the courses. This assumption may not hold when the grading standards and major requirements vary across departments.[11] In Column 5, we fully interact discretized GPA, initial grade, and the department in which the class is offered—the triple interactions create 4,500 mutually exclusive cells—to control for varying regrade consideration across the interaction dimensions. The gender gap estimate is reduced slightly to 0.0698 percentage points. Thus, the finding that males are more likely to have a favorable grade change is remarkably robust.

We analyze the same equation with downward regrades as the outcome in Panel B of Table 2. As stated earlier, male students were also more likely (by 0.0102 percentage points as reported in Column 1 of Panel B) than female students to receive a downward regrade. Although the result is insensitive to the inclusion of the student's class standing, GPA, and initial grades, the relative adverse outcome for males disappears when we control for the class characteristics.

Including all the aforementioned control variables (i.e., colleges, departments, instructors' gender and rank, students' class standing, grade, and GPA, and the interactions of students' GPA, initial grades, and departments in which the classes are offered), Column 1 of Panel A in Table 3 reports the baseline estimate of a 0.0698 percentage points male advantage in upward regrades that cannot be explained by class and student observable characteristics. To investigate the sensitivity of the analysis, we estimate the model using a logit regression

[10]The odds ratio is calculated as $\frac{Pr_{Male}(Y=1)/Pr_{Male}(Y=0)}{Pr_{Female}(Y=1)/Pr_{Female}(Y=0)}$, holding all else at the mean values.

[11]For example, students majoring in economics have to pass core economics courses with a minimum grade of C to receive the bachelor's degree, while students majoring in agricultural economics only need to pass these courses to meet their degree requirements.

in Column 2. Because a large number of observations in the triple interaction cells do not have any grade change records, we are unable to reliably estimate the gender differences in regrades because these observations will be dropped from the logit regression. Alternatively, we use the continuous values of GPA and initial grade for the logit regression analysis. The gender gap remains significant at the value of 0.0829 percentage points from the logit model in Column 2. We then remove students who received multiple upward regrades from the analysis. We find that the gender gap reduces but remains substantial at 0.0583 percentage points in Column 3 of Panel A. This result suggests that approximately 18 percent of the male advantage was due to the fact that male students were more likely (8.03 percent) to receive multiple upward grade changes than were their female counterparts (7.29 percent) during their time in college.

If male class enrollment is correlated with instructor leniency in regrades (e.g., through knowledge of instructor reputation), a concern may be that gender differences in regrades are driven by a handful of instructors who make frequent grade changes. Column 4 of Panel A in Table 3 examines the sensitivity of the result by removing observations if the instructor made more than 10 percent changes in grades; we see the exclusion explains little of the gender difference (i.e., the gender difference reduces slightly from 0.0698 to 0.0565 percentage points and remains significant).

Next, we consider that students whose grade has hit the upper bound (A+ and A) may not need to request grade changes. We remove them from the analysis in Column 5 of Panel A in Table 3. The gender gap actually increases to 0.105 percentage points because female students are over-represented in this high-end grade group. On the other end of the grade spectrum, students who hit the lower bound (F) may have greater incentives to request regrades because they face no downside risks. By removing students who received a failing grade, the gender gap in Column 6 becomes 0.0715 percentage points—similar to the magnitude in the baseline model. In Column 7, we restrict the sample to large classes (i.e., classes with more than 90 students), and the magnitude of gender gap is 0.0724 percentage points—statistically indistinguishable from the gender gap in smaller classes.

We conduct a similar sensitivity analysis for downward grade changes in Panel B of Table 3. When we control for all these aforementioned independent variables, the gender difference in downward regrades becomes trivial (-0.0008 percentage points) and insignificant as presented in the baseline model in Column 1 of Panel B in Table 3. The absence of gender differences in downward regrades is persistent across the sensitivity analyses (Columns 2 through 7).

In summary, we find robust gender differences in regrades, with males significantly more likely to experience upward grade changes. The underlying mechanisms are not clear. For

example, the gender differences in regrades may be a result of at least three distinct scenarios: 1) male students are more likely than female students to ask instructors for grade changes although instructors treat all requests equally (i.e., a gender difference in willingness to ask); 2) the propensity to ask is the same for both male and female students, but the outcomes are more favorable for males than for females when they ask (i.e., discrimination); and 3) female students make regrade requests *during* the semester which in turn lowers their demand for regrade requests at the *end* of the semester. Understanding the mechanisms is relevant for appropriate policy recommendations. For example, if these patterns were driven by the third explanation, the observed patterns should not be reason for concern. Administrative records do not contain any information that allows us to understand which scenario accounts for these unequal outcomes. The administrative data also are limited in informing us about the prevalence of regrade requests in college—students might have asked all along during the semester. In addition, the data contain no information on any requests that were made but resulted in no grade change.

To investigate the prevalence of regrade requests more generally and understand the gendered patterns, we next report results from surveys that we administered to instructors and students. These surveys elicited respondents' recollections on regrade requests in the past semesters. Then, in the subsequent section, we report results from an experimental study that further allows us to quantify the underlying mechanisms.

3 Surveys of Instructors and Students

3.1 Samples

For the instructor survey (Appendix A1), invitational emails were disseminated to all faculty members and graduate students across all colleges at CSU. For eligibility, instructors must have taught at least one undergraduate course and experienced regrade requests at the end of a semester. The email contained an online survey link. 154 instructors (including graduate teaching assistants) completed the instructor survey between October and December, 2018. The average (median) time taken to complete the survey was 11.94 (8.71) minutes. The respondents were over-represented by female instructors (58 percent female in the survey compared to 48 percent in the administrative records) and under-represented by non-tenure tracked instructors (26 percent non-tenure tracked instructors in the survey as opposed to 49 percent in the administrative records). Weighted by class sizes, male students represented 52.5 percent of the classes in the sample which was slightly higher than the 46.6 percent in the administrative records. Instructors reported that 5.94 percent of their students requested

to change their final grade at the end of the semester, and 11.2 percent of students requested regrades during the semester. Weighted by class sizes, instructors who had experienced regrade requests reported that 0.727 percent of the grades were corrected to a better grade at the end of the semester. The reported upward corrections were more frequent than the corresponding proportion of 0.439 percent in administrative records because the instructor survey only elicited participation of instructors who had experienced regrade requests.[12]

For the student survey (Appendix A2), recruiting emails were sent to all undergraduate students on campus. The email mentioned that the survey was about gauging students' experiences with courses, and that the student must have completed at least three courses with letter grades at the time of survey. To prevent selection on our outcome of interest, the invitation did not mention the focus on regrade requests. The email contained the online survey link. As an incentive, respondents were offered a chance to win one out of the ten prizes of $100.[13] The average (median) time taken to complete the survey was 9.44 (7.99) minutes.[14]

1,295 students completed the online student survey during April–May, 2019 (459 respondents started the survey but did not complete it, and hence are not included in the analysis). In the student survey, we asked students about their past behaviors, subjective beliefs, and outcomes regarding grade change requests. The student survey, similar to the administrative records, show a monotonic increase in the share of students by class standing.[15] Although the student survey was over-represented by female students (67 percent), the course-taking patterns appear to be comparable to the underlying population. Similar to the administrative records, among the reported courses, 50 percent of female students were taught by female instructors, while only 35 percent of male students were taught by female instructors. For students who indicated consideration of regrade requests during past semesters, we further elicited the frequencies of such a consideration and collected information about the instructor and grade outcomes for up to three classes. Importantly, we collected data on students' perception regarding the outcomes of a regrade request: specifically, we elicited

[12] In the sub-sample collected between November 30 and December 9, 2018, we allowed all instructors to participate. The reported upward regrades comprises 0.354 percent of the students in the sub-sample—close to the statistic from the administrative records. This exercise confirms that the difference between the instructor survey and the administrative records is due to the exclusion of instructors who had never experienced regrade requests in the survey. To keep the sample consistent, we report all the results conditional on instructors who had experienced some regrade requests at the end of the semester.

[13] The payments are made in in Ram Cash or Amazon electronic gift cards, depending on students' choices. Ram Cash is a near-cash payment credited to the CSU student accounts that participants can use for consumption on campus.

[14] Time spent on a page is truncated at 10 minutes. The truncation affects fewer than 1% of the pages.

[15] The distribution of participants in the student survey (administrative records) is as follows: 16.5% (9.86%) freshmen, 24.4% (20%) sophomores, 28.8% (25%) juniors, and 30.3% (45.1%) seniors. Any student who has not graduated beyond the fourth year in college is counted as a senior.

students' subjective beliefs about the regrade requests that result in favorable outcomes (a better grade) and unfavorable outcomes. This allows us to understand if male and female students have substantially different beliefs—that is, whether students assess the upside and downside risks differently by gender, and whether the fear of backlash (i.e., students receiving a lower grade after requesting a grade change) is a factor driving gender differences in making grade change requests.

Students reported high frequencies of ever requested regrades—16.8 percent of students reported that they requested regrades at the end of a semester and 29.9 percent during the semester at some point. These statistics are not comparable to those from the instructor survey—the statistics in the student survey measure students' extensive margin of ever making regrade requests in any past semesters, while the statistics in the instructor survey measure the average regrade-request rate in a class.[16]

3.2 Analysis

3.2.1 Are men more likely than women to ask?

We first examine the instructor survey by comparing the gender distribution among regrade requests against the gender distribution in class. Panel A of Table 4 shows that male students made up 57.1 percent of regrade requests at the end of the semester—a larger proportion than the male class representation in the instructor survey (52.5 percent).

Results from the student survey offer further insights into gender differences in asking. First of all, asking for grade changes is fairly common among college students. Panel A1 of Table 5 shows that a sizable proportion of students have asked an instructor for a regrade at some point. Among the 1,295 student participants, 39 percent of females and 41 percent of males *ever* asked for regrades during or at the end of a semester (the difference is statistically insignificant). Turning the focus to the number of classes students considered (asked) for regrades, a gender difference emerges: male students considered asking (actually asked) for regrades in 1.30 classes (0.65 classes) versus 1.04 classes (0.56 classes) for females; the gender difference is significant at the 1-percent (10-percent) level. These differences are driven by the gender gap on the intensive margin (Panel A2 of Table 5). Conditional on considering (asking) for regrades, females considered (asked) regrades requests in 1.78 (1.43)

[16]To make the student survey's statistics comparable to those in the instructor survey, we multiply the fraction of students who reported ever asking for regrades at the end of/during the semester (0.168/0.299) by the fraction of the courses (up to three courses considered for regrades) in which they made such requests (0.209/0.392) to get the average student requests at the end of/during the semester as 3.5% ($0.168 \times 0.209 = 0.035$)/11.7% ($0.299 \times 0.392 = 0.117$). The regrade request rates from the student survey using this back-of-the-envelop calculation are close to the statistics (5.94 percent at the end of the semester and 11.2 percent during the semester) reported by instructors.

classes which is lower than the 2.14 (1.59) classes considered (asked) by male students. Figure 2 shows the gender-specific distributions of regrade consideration; the distributions are statistically different by gender. Note that the number of classes students could have reported for regrade consideration in the survey was capped at "5 or more classes", and we only collected information on whether students actually asked for regrades in up to 3 classes. Hence, the gender difference in the reported regrade requests is likely under-stated. To mitigate this issue, we imputed the number of classes where students asked for regrades if students reported that they considered asking for regrades in four or more classes.[17] The unconditional (conditional) imputed number of classes in which students actually asked for regrades is 0.97 (1.65) for females and 1.12 (1.84) for males, and the difference is significant at the 5-percent level. This result indicates that the gender difference in propensity to ask is primarily driven by the difference on the intensive margin.

The survey data reveal that (1) regrade requests are very prevalent, and (2) while there is little gender difference in asking on the extensive margin, males are significantly more likely to ask on the intensive margin; using the three different measures (the number of classes considered for regrades, the number of classes reported for regrade requests, and the imputed number of classes of regrade requests), we see that males request (or consider requesting) about 16 to 25 percent more than females do. Interestingly, the magnitude is similar to the evidence from the administrative records where male students are 18.6 percent more likely than their female counterparts to receive favorable regrade outcomes.

3.2.2 Do male students have more favorable outcomes when they ask?

To answer this question, we turn to the instructor survey. The survey asked the following questions: "*Now consider all regrade requests that are made by [**female/male**] students [**DURING/at the END of**] the semester. On average, what percent of them result in an improvement, no change, and a deterioration of the grade? Please note that your answers need to sum to 100.*" With regards to the outcomes of regrade requests, Figure 3 shows that the change patterns conditional on student requests were indistinguishable for male and female students at the *end* of the semester in the instructor survey. Although the regrade outcome distributions are different for male and female students *during* the semester (Panels C and D), the outcomes are actually more favorable for female than for male students.

[17] Because we only have information for up to three classes, we focus on the subsample where the number of classes students considered for regrades is between 1 and 3 to calculate the request ratio ($\frac{\text{number of classes students actually asked for regrades}}{\text{number of classes students considered for regrades}}$) by students' gender. We then impute the requests by multiplying the gender-specific ratio by the number of classes considered for regrades that is greater than three (i.e., 4 or 5).

The student survey presents a consistent pattern in regrade outcomes that are observed in the instructor survey. Summarized in Panel B of Table 5, conditional on asking, students experienced a higher success rate in a grade increase during the semester than at the end of the semester regardless of students' gender. During the semester, over 60 percent of regrade requests (65 percent of male requests and 60 percent female requests) resulted in an increase in the grade component (e.g., quiz, midterm, homework assignments) and less than 2 percent of requests led to an adverse outcome (0 percent of male requests and 1.44 percent for female requests). At the end of the semester, male (female) students reported that about 32 percent of the cases resulted in final grade increases and 6 percent (3 percent) of the cases led to a lower grade. The gender differences in regrade outcomes are not statistically significant for grade improvements. Although women reported a higher rate of receiving a decrease in their grade component, men experienced a higher rate of grade decreases in the final grade. Note that the fractions of grade changes reported in the student survey are higher than those reported in the instructor survey because we only collected information for up to 3 courses. If the remaining (imputed) classes in which the students asked for regrades result in no grade changes, the statistics across the instructor and student surveys would be very similar.[18]

Furthermore, Panel A3 of Table 5 shows that the subjective beliefs about potential outcomes of regrade requests were also very similar across gender. On average, male (female) students reported a 38.3 percent (37.0 percent) probability of receiving a grade increase, a 55.9 percent (57.2 percent) probability of the grade remaining unchanged, and a 5.83 percent (5.74 percent) probability of receiving a grade decrease.[19] The subjective probability distributions in regrade outcomes are statistically indistinguishable between male and female students. It is, for example, not the case that female students perceive a higher downside risk of asking.

As far as perceptions are concerned, it does not seem that the two genders differ in terms of how they believe their requests would be treated. However, female students report a significantly higher stress level than males do for having to ask the instructor for a grade change: an average response of 5.01 (on a 1–7 scale) versus 3.88 for males.[20] This suggests

[18] If the remaining (imputed) classes in which the students asked for regrades result in no grade change, students would experience grade increases in 18% of the time at the end of the semester and 37% of the time during the semester conditional on asking—very close to the change rates (17% for males and 17.9% for females at the end of the semester, and 32.8% for males and and 37.2% for females during the semester) reported by instructors.

[19] Each student was asked about the likelihood of potential regrade outcomes conditional on making a regrade request, regardless of whether the student actually ever made or considered making a regrade request. Hence, these data do not suffer from the selection in who asks for a regrade. We take each student's average subjective probabilities for the three potential regrade outcomes (i.e., increase, no change, or decrease in grade) across all scenarios.

[20] Students were either asked "*On a scale from 1 (not stressed at all) to 7 (extremely stressed), how stressed*

that females have higher psychic costs of asking. We discuss the implications of gender differences in stress and personality traits on regrades in Sections 4.4 and 5.

3.2.3 Does the timing of regrade requests differ by gender?

The gender differential outcomes at the end of the semester may be a result of differences in the timing of regrade requests by students' gender. For example, it could be that males are more likely than females to procrastinate, and put off such requests until the end of the semester; the literature on gender differences in procrastination has found mixed results (Balkis and Duru, 2009; Haycock et al., 1998; Milgram et al., 1995; Onwuegbuzie, 2004; Senecal et al., 1995; Solomon and Rothblum, 1984).

The surveys do not show any evidence of this. In the instructor survey (Panel A of Table 4), male students were still over-represented in the regrade requests during the semester (58 percent of the requests versus 52.5 percent of their representation in class). In the student survey (Panel A of Table 5), 29 percent of females and 32 percent of males reported ever requesting regrades during the semester.

The instructor and student surveys indicate that the most likely explanation of the gender gap in regrades is that male students are more likely than female students to ask for regrades on the intensive margin, and that these request patterns persist throughout the semester. Hence, even if instructors change the grades for male and female students at the same rate, the outcome may still favor male students simply because they ask more frequently. The survey data also show that regrade requests are quite prevalent and may have serious implications for final grades. These data in fact can give us an upper bound of GPA changes as a result of regrade requests—assuming that 1) the ratio of classes reported for regrade requests to classes considered for such requests (up to 3 courses) represents students' average requesting intensity, and 2) the magnitudes of grade changes students reported in their final grade also apply to the (unobserved) changes made *during* the semester. The average effects of regrade requests on GPA are 0.11 points for females and 0.13 points for males. Conditional on asking, the upper bounds of the GPA increases are 0.43 points for females and 0.47 points for males.[21]

would you have felt if you had asked the instructor for a regrade?" or "*On a scale from 1 (not stressed at all) to 7 (extremely stressed), how stressed did you feel when you asked the instructor for a regrade?*" The former was asked to students who did not ask (or never considered asking) an instructor for a regrade.

[21] We calculate the upper bound of the unconditional effect of regrade requests on GPA by gender as the likelihood of the student ever having asked for a regrade × fraction of courses asked for regrades × grade change: $\sum_t Pr(EverAsked_{gt}=1) \times \frac{ClassesAsked_{gt}}{ClassesConsidered_{gt}}|_{EverAsked_{gt}=1} \times [\sum_s Pr(Outcome=s|_{Asked_{gt}=1}) \times \Delta Grade_{gts}]$, where $Pr(EverAsked_{gt})$ is the fraction of students in the gender group $g \in \{male, female\}$ ever asking for regrades in time $t \in \{mt, fn\}$ either *during* the semester (mt) or at the *end* of semester (fn), $\frac{ClassesAsked_{gt}}{ClasseesConsidered_{gt}}$ is the ratio of the number of classes students reported that they actually asked for regrades

Using the student survey, Appendix Table B2 offers some insights into the reasons for why students asked or did not ask for regrades. The reasons for regrade requests are mostly similar across gender, except that females are more likely than males to ask when they received a grade relatively lower than their performance in other classes and when the low grade is likely to delay their graduation. The primary reasons for the gender difference in not asking for regrades include females more likely to report that: (1) instructors would never change the grade, (2) a higher stress level resulting from asking, and (3) a stronger fear for rejection ("*I would have been embarrassed if the instructor had rejected my request.*") or punishment ("*I was worried the instructor might lower the grade further, or punish me in some other way.*"). To investigate why men and women differ in their propensity to ask, we next rely on an incentivized controlled experiment to help us disentangle how a variety of factors, such as gender differences in risk attitudes, confidence, optimism, perceived uncertainty, and personality traits affect behavior.

4 Laboratory Experiment

To understand what gives rise to gender differences in regrade requests, we designed a laboratory experiment where grades serve as imperfect and noisy signals of individual performance in a task, and allow participants to request a regrade at varying levels of costs. The shown grade matters for the individual's monetary payoff. If a participant decides to pay a given cost to ask for a regrade, the true grade is revealed and the payoff is adjusted accordingly. The laboratory setting is advantageous over observational field evidence because we are able to control and eliminate factors, such as fear of backlash, dynamics of gender interactions, or gender differences in negotiation skills, that may contribute to the gender differences in regrade requests. In addition, behavior in this setting is incentivized (since a regrade request may impact not only the revealed final score but also the payoff), so if we find results consistent with our survey evidence, it would give us more confidence in the conclusions.

We first examine whether male and female participants differ in asking for regrades. Because participants' risk preferences and subjective beliefs about potential outcomes may dictate their regrade decisions, we elicit the measures via incentivized question designs. We further elicit students' beliefs about their performance in the experiment to construct

to those they considered for regrade requests—up to the 3 courses reported, and $Pr(Outcome = s|_{Asked_{gt}=1})$ is the probability the outcome for gender g conditional on asking at time t resulting in either an increase or decrease (s) in the grade, and $\Delta Grade_{gts}$ is the average reported final grade changes given the outcome scenario s at time t for group g—assuming the same magnitude of the final grade change also applies to the grade changes *during* (mt) the semester. We calculate the regrade effects on GPA conditional on asking as: $\sum_t \frac{ClassesAsked_{gt}}{ClassesConsidered_{gt}}|_{EverAsked_{gt}=1} \times [\sum_s Pr(Outcome = s|_{Asked_{gt}=1}) \times \Delta Grade_{gts}]$ following the same definitions of the terms described above.

measures of over-confidence, under-confidence, over-optimism, uncertainty, and downside risk in beliefs. These measures allow us to quantify the effects of the various factors on the gender differences in asking. At the end of the experiment, participants also completed the student survey which collected their regrade experience in class and psychological characteristics such as the Big Five personality traits. As a result, we are able to investigate how participants' choices in the laboratory setting relate to their actual regrade experience in class and whether additional psychological factors help explain the gender differences.

4.1 Laboratory Experiment Design

The experiment consists of 5 tasks:[22]

Task 1 (risk preferences): Following Eckel and Grossman (2008), we ask participants to choose their most favorable option from the following five lottery choices:
1) 50% chance of receiving \$2 (Event A) and 50% chance of receiving \$2 (Event B),
2) 50% chance of receiving \$3 (Event A) and 50% chance of receiving \$1.50 (Event B),
3) 50% chance of receiving \$4 (Event A) and 50% chance of receiving \$1 (Event B),
4) 50% chance of receiving \$5 (Event A) and 50% chance of receiving \$0.5 (Event B),
5) 50% chance of receiving \$6 (Event A) and 50% chance of receiving \$0 (Event B).
A random draw decides whether the event A or B occurs, and participants are paid according to their lottery choice.

Task 2 (performance): Students complete a quiz by answering 20 multiple-choice IQ questions derived primarily from Russell and Carter (2006).[23] Questions appear one at a time, and students have 45 seconds to answer each question. Students are informed about the payoff scheme before taking the quiz: a piece rate of \$0.50 for each answer graded as correct, and an additional \$2.00 bonus when their grade exceeds a threshold (i.e., a bonus payment of \$2 for getting 6–10 correct answers, \$4 for getting 11–15 correct answers, and \$6 for getting 16–20 correct answers) that resembles the step function of letter grades. As a result, the total payoff for the quiz could range from \$0 to \$16.

When taking the quiz, after each question, students are also asked to provide their best guess (on a 0–100 scale) of the chance they think the answer is correct. One of the 20 guesses is randomly chosen to reward the accuracy of the guess for a bonus payment up to \$1, following the formula in Witkowski et al. (2018).

[22] In the experiment, we break each of Tasks 3 and 5 into two decisions, and hence participants were asked to complete 7 tasks. Task 3 here corresponds to Tasks 3 and 4, Task 4 corresponds to Task 5, and Task 5 corresponds to Tasks 6 and 7 in the experiment instrument.

[23] We sample few questions online and shorten some questions in Russell and Carter (2006) to allow participants to answer them within the time limit for each question.

Task 3 (prior beliefs): After completing the quiz, students are asked for 1) their "prior" belief about the number of questions (out of 20) they thought they answered correctly; and 2) the uncertainty around the prior belief—students are asked for the likelihood that the actual grade is in five intervals centered around their guess (i.e., the actual grade is 5 or more questions below their guess, 2–4 questions below their guess, between −1 to +1 question of their guess, 2–4 questions above their guess, and 5 or more questions above their guess). The probabilities assigned to the five bins have to sum to 100. These guesses are also incentivized for up to a $5 payment depending on the accuracy of the guesses following the formula in Witkowski et al. (2018).

Task 4 (willingness to ask for regrades): Before the score is revealed to the student, she is informed that 3 out of the 20 questions would be drawn and graded randomly, and therefore the score she would be shown may be higher, lower, or the same as her true performance. To imitate the asymmetric upward and downward regrade patterns we observe in the administrative records, we assigned the noise in the score as follows: for each of these three randomly drawn questions, if a student's answer is actually correct, there is a 1/3 chance that it would be graded as correct and 2/3 chance that it would be graded as incorrect. In contrast, if the student's answer is incorrect, there would be a 2/3 chance that it would be graded as incorrect and 1/3 chance that it would be graded as correct. In other words, for these three questions, a correct answer has twice as much chance of being graded as incorrect (false outcome) than as correct (true outcome), and an incorrect answer has twice as much chance of being graded as incorrect (true outcome) than as correct (false outcome). Therefore, when a student requests a regrade, in expectation, the chance of receiving a higher grade is higher than the chance of receiving a lower grade.

Once the initial score is revealed, the detailed grading (i.e., graded as correct or incorrect) and the subjective probability that the student assigns to each answer being correct is shown on the screen for each of the 20 questions. The student is then informed that she could make a regrade request. Specifically, we use a multiple-price-list variation of the Becker-DeGroot-Marschak method to elicit the individual's willingness to ask for a regrade: the participant is asked to decide whether she wants to make a regrade request at 10 varying cost scenarios. The costs vary from paying $3.50 to getting paid $1.00.[24] One of the cost scenarios is randomly drawn for implementation. If the student chooses to request a regrade, she pays

[24]The 10 regrade cost scenarios are having to pay $3.50, $3.00, $2.50, $2.00, $1.50, $1.00, $0.50, $0.00, −$0.50 (that is, getting paid $0.50), and −$1.00 (getting paid $1.00). This set-up allows us to estimate the respondent's willingness to pay (WTP) to ask for a regrade. For example, if the individual chose to make a regrade request for costs of up to $2.00, then her WTP must be at least $2.00 but less than $2.50. In this case, we assign the lowest possible value $2 as her WTP. Note that the highest possible gain from asking for a regrade is $3.5 (which happens if all three correct answers were graded as incorrect, and the score change is around a jump in the step function).

the associated regrade cost (deducted from the show-up fee) and the true grade is revealed as her final grade. If the student chooses not to request a regrade, she pays zero cost and the initial grade becomes the final grade.

Task 5 (posterior beliefs): Using the same approach as for Task 3, we re-elicit participants' "posterior" beliefs about the number of accurate answers and the uncertainty around the posterior belief. By construction, the posterior uncertainty is within 3 questions above and below the initially revealed score. Therefore, students are asked for the likelihood that the actual grade is in seven bins centered around the initial score (i.e., the actual grade is 3 questions below the initial score, 2 questions below, 1 question below, same as the initial score, 1 question above the initial score, 2 questions above, and 3 questions above). The incentives formula is the same as for the prior beliefs mentioned earlier. One of the four subjective beliefs (the two beliefs in each of Tasks 3 and 5) is randomly drawn to determine a bonus payment of up to $5.00.

The experiment also had built-in checks and test questions to make sure that participants understood the instructions. Appendix A3 shows the instrument of the experiment.

4.2 Sample and Administration

The experiment was created using oTree (Chen et al., 2016). We recruited participants by sending emails to all undergraduate students on campus to invite them to take part in an incentivized experiment of economic decision-making and a survey related to college course experiences. 563 CSU undergraduate students participated in the online experiment during April–May 2019. After the conclusion of the experiment, respondents were automatically forwarded to the survey. Thus, for these respondents, we have both experimental and survey data.

The experiment (excluding the survey) took 24.5 minutes, on average, to complete (see footnote 14 for details). Combining the payoff for the quiz, bonuses for guesses, and a $5 guaranteed payment for completing the experiment, participants could earn between $5 and $34. The average compensation was $20.14, with the 10^{th} percentile being $14.50 and the 90^{th} percentile being $26.

We remove 47 observations (8.35 percent) where students made inconsistent regrade requests (for example, asking for a regrade at $3 but then choosing not to ask for a regrade at a cost of $2.5).[25] The results of our analyses are robust to inclusion of these students (see

[25]This inconsistency is within the range of other studies that use multiple price list elicitation methods. For instance, Holt and Laury (2002) find that up to 10% of participants switched their lottery choices. The share of inconsistent respondents was about 2% in Allcott and Kessler (2015) and 15% in Cullen and Perez-Truglia (2018).

Appendix Tables B3 and B4). The final sample consists of 516 participants, with 283 female students (55 percent) and 233 male students (45 percent). The gender distribution closely mirrors the gender distribution of the CSU undergraduate population.

4.3 Key Measures

4.3.1 Risk aversion

We assume that participants' utility function is constant relative risk averse (CRRA): $U = \frac{C^{1-\gamma}}{1-\gamma}$, where γ measures the CRRA coefficient and C is the payment the participant receives. The expected utility from the chosen option is $E(U) = \left(0.5 \times \frac{C_A^{1-\gamma}}{1-\gamma}\right) + \left(0.5 \times \frac{C_B^{1-\gamma}}{1-\gamma}\right)$, where C_A and C_B are the payments from the two events A and B, respectively. Participants' lottery choices reveal the range of risk aversion coefficients that are consistent with the utility maximization assumption. For simplicity, we take the midpoint of each interval as the risk aversion coefficient for each corresponding lottery choice.[26] Consistent with findings in prior literature, Panel A of Table 6 shows that female participants are more risk averse than males (an average risk aversion coefficient of 1.12 versus 0.86 for males), with the gender difference significant at the 1-percent level.

4.3.2 Over-confidence and under-confidence

After answering each question, respondents were asked to assign a probability that their answer is correct. Over-confidence is constructed using the questions the respondent answered incorrectly. It is defined as the average belief assigned to these questions of the answers being correct, and takes a value between 0 and 1 (beliefs are elicited on a 0–100 scale, but we scale them down here). This measure will be 0 (1) for an individual who guesses that there is a 0 (100) percent chance of her being correct for each question that she ends up actually answering incorrectly. Under-confidence is similarly constructed by using only correct answers for the measure and is the average absolute gap between the reported belief and 1 (that is, the answer was correct). Thus, if the participant guesses that there is a 100 percent chance that she answered all these questions correctly, the value of her under-confidence measure would be 0. In contrast, if the participant has no confidence in her correct answers and assigns the probability of being correct as 0, the under-confidence value would be 1. These variables are conditional on the accuracy of the question. In the analysis, as we discuss below, we also control for the beliefs regarding the number of questions answered correctly.

[26] See Table 1 of Eckel and Grossman (2008) for the CRRA ranges. The mid-point values of risk aversion coefficient are 2, 1.335, 0.525, 0.29, and 0.2 for the lottery choices 1 through 5 in the ascending order.

Panel A of Table 6 shows that male and female participants appear to be similarly overconfident when their answers were wrong, while females are more likely than males to be under-confident when their answers are actually correct. This is consistent with the finding by Sarsons and Xu (2015) that tenured female economists tend to be less confident in the accuracy of their answers to survey questions about the economy, relative to their male counterparts. This also seems to be in line with Exley and Kessler (2019), who find that females rate their performance less favorably than equally performing men.

4.3.3 Over-optimism, uncertainty, and downside risk in beliefs

We elicited participants' prior belief about their performance: Panel A of Table 6 shows that the average prior belief is higher for males (14.35 versus 13.21 for females, with the difference statistically significant at the 1-percent level). Panel C shows that the true score was 12.98 for males and 11.57 for females. Thus, the average belief is higher than the actual score for both genders by about 1.5 points. Note that the objective performance gap between males and females in the quiz plays an immaterial role in the regrade decisions. As long as male and female participants are similarly optimistic (i.e., their expected score is higher than their true score), they face similar incentives for regrade requests. Appendix Figure C1 shows the relationship between the prior guess and actual performance. Beliefs are indeed positively related with actual performance for both genders. Each additional point of guessed score is associated with 0.47 (0.55) points in actual score for female (male) students, and the gender difference is statistically insignificant.

We use participants' prior belief of the number of questions they thought they answered correctly to construct our measure of over-optimism. Specifically, it is the gap between the prior guessed score and the actual score. Most participants are over-optimistic about their performance in the quiz. Table 6 shows that both genders are, on average, equally over-optimistic: female (male) students believed they answered 1.64 (1.37) more questions correctly than their true score.

In addition, the experiment elicited the probability that the participant believes the actual score was within one point of the guessed outcome. If the participant is absolutely certain about their guess, they should assign a probability close to 100%.[27] We capture the uncertainty in beliefs as the gap between 1 and the assigned probability (scaled down to 0-1) to the predicted outcome. Hence, the measure takes values between 0 and 1, with a value closer to 1 indicative of the participant being very uncertain about her prediction. Table 6 shows that, on average, female participants are more uncertain (0.51) than their

[27] We also consider an alternative uncertainty measure using the variance of their expectations based on the subjective probability distributions. Results are insensitive to the choice of uncertainty measures.

male counterparts (0.47), and the gender difference is significant at the 10-percent level.

We also consider that participants may respond differently to upside risk (probability of favorable outcomes) versus downside risk (probability of unfavorable outcomes) as suggested by Comeig et al. (2018). Note that the probability of the guessed outcome, favorable outcome (i.e., the actual score is more than one point higher than the guessed score), and unfavorable outcome (i.e., the actual score is more than one point lower than the guessed score) sum to 1 (100%). As a result, we include only the uncertainty measure for the guessed outcome and the downside risk measure to avoid perfect multi-collinearity. We capture the prior downside risk as the summation of probabilities (between 0 and 1) participants assigned to the scenarios where the actual score is: 5 questions below the guessed score, or 2-4 questions below the guessed score before the revelation of the initial score. Table 6 shows that female students perceive a higher downside risk in their prior beliefs than male students do (0.33 versus 0.29, and the difference is significant at the 5-percent level).

Posterior beliefs and uncertainty were elicited in a similar way. Panel A of Table 6 shows that the posterior belief is lower than the prior for both males and females. Both genders revised their guessed score downward by a similar magnitude after receiving the initial grade signals (1.07 for males and 1.11 for females, and the difference is statistically insignificant). Posterior beliefs, on average, remain higher than the true score (12.10 for females compared to their true score of 11.57, and 13.28 for males compared to their true score of 12.98). We see that the posterior downside risk (that is, the probability assigned to the actual score being 1-3 questions below the noisy shown score) is significantly higher for females than for males (an average probability of 0.28 versus 0.22 for males). This could be consistent with gender differences in response to noisy feedback (see, for example, Shastry et al. (2019)).

4.3.4 Personality traits

We also collected data on the Big Five personality traits—extroversion, agreeableness, conscientiousness, neuroticism, and openness. These traits were collected using standard survey methods (Rammstedt and John, 2007). Panel B of Table 6 shows the average values of these traits by gender. Consistent with prior literature findings (Costa et al., 2001; Weisberg et al., 2011), we find that females are significantly more agreeable, neurotic, and open. For example, the gender difference in neuroticism is 0.63, almost 61% of the standard deviation (1.04) in this measure. These personality traits could have direct connections to individuals' negotiation and conflict management styles. For instance, individuals with high degrees of agreeableness and neuroticism tend to avoid confrontations (Antonioni, 1998), and the high level of agreeableness is particularly detrimental in bargaining outcomes in zero-sum negotiations (Barry and Friedman, 1998). Based on these findings, there is good reason to believe

that students with higher levels of agreeableness and neuroticism may be less likely to ask.

4.4 Analysis

Figure 4 shows interesting gender differences in the willingness to ask for regrades. When the cost is zero or negative (i.e., participants are getting paid for requesting regrades), there is no discernible gender difference in the willingness-to-pay (WTP) for regrades. However, females are much more reluctant than males to request regrades when the cost is positive. For instance, when the regrade cost is \$0.50, 47 percent of male students asked for regrades, while only 36 percent of female students asked for regrades. This 10-percentage-point difference (significant at the 5-percent level) shows that males are 28.5 percent more likely than females to request regrades. The fraction of students asking for regrades declines monotonically as the cost increases. We take the highest value of the cost a participant is willing to pay for regrades as the measure of her WTP.[28] Table 6 shows that the average WTP is higher for males (\$0.25) than for females (\$0.05) and the difference is significant at the 5-percent level. In addition, while nearly half (47%) of males are willing to pay a positive cost for regrades, only a third of females are willing to do so. Participants' WTP is positively correlated with their perceived performance relative to the initial scores they receive. Figure 5 shows that the WTP increases as the gap between the guessed score (prior belief) and the initial score widens. The slope coefficients are similar across gender.

We next investigate the drivers of gender differences in the WTP. Column 1 of Table 7 regresses the indicator of willing to pay a positive cost for regrades on respondents' gender. Men are 10.4 percentage points more likely than women to pay a positive cost for regrades. While we find that risk preferences differ by gender, it explains little of the gender difference in willingness to pay for regrades (Column 2 of Table 7). Columns 3 and 4 show that over-confidence and under-confidence explain about 7.7 percent and 19.2 percent of the gender gap, respectively. Because women are less confident about their answers when they are actually correct, under-confidence can explain a non-trivial fraction of the gender difference in the willingness to pay.

Column 5 of Table 7 shows that over-optimism, uncertainty in prior beliefs of quiz performance, prior guessed score and prior downside risk combined also explain about 30 percent of the gender gap. Column 6 controls for the posterior guessed score and the posterior downside risk measure; the gender coefficient remains similar to the coefficient in Column 5 where we only include the measures based on prior beliefs. We also find that the effect

[28] If a person always chooses not to request regrades, we know her WTP should be lower than −\$1.00. In this case, we assign the WTP to be −\$1.50. If a participant always chooses to ask, her WTP must be \$3.50 or higher. We assign her WTP to be \$3.50. Results are robust to other values.

of downside risk on WTP is similar for both male and female students (results available from the authors). Column 7 shows that all the aforementioned covariates combined explain about 35 percent of the gender gap in willingness to pay for regrades, and the gender gap becomes statistically insignificant.

Column 8 examines whether the Big Five personality traits help explain the gender differences in regrade requests. Controlling for the Big Five personality traits, the gender difference in asking for regrade at a positive cost reduces to 8.3 percentage points (although the coefficients on the personality traits are imprecise and insignificant). Finally, the last column shows that including all the covariates and the Big Five personality traits explains 45.2 percent of the gender gap. Thus, the majority of the gender gap in the willingness to pay persists even after controlling for a rich and extensive set of controls. Appendix Table B5 shows similar patterns when we use the WTP as the dependent variable instead.

The high willingness to ask on the part of men, however, does not necessarily translate into increased monetary returns for men, once the regrading costs are taken into account. Panel A of Figure 6 shows that indeed male students become better off financially when the regrading cost is positive but low (i.e., \$0.50). However, panel B shows that the fraction of male students who become worse off monetarily exceeds the fraction of female students facing such an adverse outcome when the costs are greater than \$1.00. Averaging over the 10 cost scenarios, Panel C of Table 6 shows that the fraction of students becoming better (worse) off is higher for males, though the gender differences are not statistically significant. In summary, the laboratory analysis corroborates the evidence from the instructor and student surveys that males have a higher propensity than females to ask for regrades. In particular, women are less likely to ask for regrades when they face positive costs. For low cost values, this hurts women relative to men. However, at high cost values, the higher propensity of men to ask can actually make them worse off.

Finally, a natural question to ask is whether our experimental measure of WTP to ask correlates with actual regrade requests by students. Figure 7 shows that it is indeed the case: the experiment-based WTP is positively correlated with the consideration for regrades in classes, on both the extensive and intensive margins.

5 Conclusion

Women's lower tendency to ask (for promotions, salary increases, etc.) has been considered as one explanation for gender differences in labor market outcomes (Babcock and Laschever, 2003; Rigdon, 2012; Card et al., 2016). Here, our focus is on whether this tendency could lead to differential outcomes *prior* to individuals' entry in the labor market.

We start out by documenting the empirical fact that males are significantly more likely to have grade changes on their transcripts. The overwhelming majority of these grade changes are grade improvements. Class and student characteristics hardly explain the favorable regrade outcomes that male students enjoy in the administrative records. This empirical fact by itself does not necessarily mean that men are more likely to ask, nor does it tell us about the prevalence of asking for regrades in college. To shed further light on this, we conduct surveys of students and instructors. The surveys show that regrade/grade change requests are prevalent in college with 40 percent of students reporting that they approached instructors for grade changes which subsequently result in a better grade 30 percent of the time conditional on asking. The surveys further show that there is no gender difference in asking on the extensive margin. However, male students are disproportionately more likely than female students to make regrade requests in more classes (the intensive margin).

To shed light on the potential mechanisms, we rely on an incentivized laboratory experiment where initial performance is graded with error. The student's monetary payoff depends on the grade in the experiment. Students are then given the choice to request a regrade at different costs. A regrade request may subsequently affect the final grade and payoff. Our experiment reveals that males have a higher propensity to ask, both in cases where it makes economic sense to do so and where it does not. This is consistent with the findings from the surveys: instructors report that male students are more aggressive than female students in asking for regrades (Panel B of Table 4), and male students also indicate stronger regrets for asking too aggressively for regrades (Panel A3 of Table 5). Thus, a coherent picture emerges: men are simply more willing to ask. When the regrade cost is low, male students benefit more. However, such male advantage diminishes quickly as the cost of regrades increases.

Why is it that females are less willing to ask? Although fear of backlash, the dynamics of gender interactions, and gender differences in negotiation skills may play some roles in the gender differences in regrades in real life, we purposely eliminate these factors in the laboratory setting. Our experimental evidence suggests that gender differences in risk preferences and over-confidence explain little of the gender gap in asking. In contrast, we find that under-confidence and uncertainty, guessed score, and perceived downside risk in (prior and posterior) subjective beliefs explain up to 30 percent of the gender differences in asking. Interestingly, we find that personality traits also explain a large fraction (20 percent) of the gender differences in asking. All of the above measures combined account for over 45 percent of the gender differences in asking.

Thus, the majority of the gap remains unexplained even after controlling for a rich set of variables. Our survey data shed further light on the unexplained part of the gap. The survey asked students how stressed they felt on a 1–7 scale (1: not stressed at all; 7:

extremely stressed) when/if they asked instructors to reconsider their grade. On average, female students indicate a higher stress level than males do (5.01 versus 3.88, as shown in Table 5). Stress explains nearly half of the gender difference in the number of classes students considered for regrades. We find that one's stress level is negatively correlated to the control one perceives they have over the direction of their life and positively correlated to the measure of neuroticism (one of the Big Five personality traits). If women have internalized gender norms that "nice girls don't ask"(Babcock et al., 2003), their deviation from such norms could lead to a high stress level (Reidy et al., 2018). Future work that sheds light on why females find the process of asking stressful would be valuable.

While we intentionally shut down these channels in the experiment, the survey data also reveal that, as reasons for not asking for a regrade, females are more likely to report that instructors would never change the grade, or that instructors may lower the grade further or punish them in some other ways. Females were also more likely to cite embarrassment if the instructor had rejected their request as a reason for not asking. At the same time, however, we find no gender difference in students' subjective beliefs regarding potential outcomes if they were to ask for a regrade. These all suggest higher psychic costs on part of females for asking. One policy implication of our results is clear: given the aversion of females to ask when costs of doing so are positive, the perceived cost of asking needs to be lowered. For example, instructors should make regarding policies explicit.[29] It should, however, be pointed out that in our experiment, the gains from asking are positive in expectation–that need not be the case in all environments. We certainly do not want to push individuals into over-asking (Exley et al., 2019). Even in our setting, we find that the greater willingness of males to ask works both against and in favor of them, depending on the associated financial costs. In addition, we find that higher levels of under-confidence and uncertainty in beliefs of females can explain 20 to 30 percent of the gender gap in asking. Policies that provide stronger signals to individuals about their ability could also allow them to make informed decisions about when to ask.

[29]Consistent with this idea, Leibbrandt and List (2014) find that men and women are similarly likely to negotiate when salaries are explicitly made negotiable in the job postings, but gender differences in asking emerge when the negotiability is not explicit.

References

Allcott, H. and J. Kessler (2015). The welfare effects of nudges: A case study of energy use social comparisions. no. w21671.

Andersen, S., G. W. Harrison, M. I. Lau, and E. E. Rutström (2006). Elicitation using multiple price list formats. *Experimental Economics 9*(4), 383–405.

Antonioni, D. (1998). Relationship between the big five personality factors and conflict management styles. *International journal of conflict management 9*(4), 336–355.

Artz, B., A. H. Goodall, and A. J. Oswald (2018). Do women ask? *Industrial Relations: A Journal of Economy and Society 57*(4), 611–636.

Babcock, L. and S. Laschever (2003). *Women don't ask: Negotiation and the gender divide*. Princeton University Press.

Babcock, L., S. Laschever, M. Gelfand, and D. Small (2003). Nice girls don't ask. *Harvard Business Review 81*(10), 14–16.

Balkis, M. and E. Duru (2009). Prevalence of academic procrastination behavior among pre-service teachers, and its relationship with demographics and individual preferences. *Journal of Theory & Practice in Education (JTPE) 5*(1).

Barber, B. M. and T. Odean (2001). Boys will be boys: Gender, overconfidence, and common stock investment. *The quarterly journal of economics 116*(1), 261–292.

Barry, B. and R. A. Friedman (1998). Bargainer characteristics in distributive and integrative negotiation. *Journal of personality and social psychology 74*(2), 345.

Bowles, H. R., L. Babcock, and L. Lai (2007). Social incentives for gender differences in the propensity to initiate negotiations: Sometimes it does hurt to ask. *Organizational Behavior and human decision Processes 103*(1), 84–103.

Card, D., A. R. Cardoso, and P. Kline (2016). Bargaining, sorting, and the gender wage gap: Quantifying the impact of firms on the relative pay of women. *The Quarterly Journal of Economics 131*(2), 633–686.

Chen, D. L., M. Schonger, and C. Wickens (2016). otree—an open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance 9*, 88–97.

Comeig, I., C. Holt, and A. Jaramillo-Gutiérrez (2018). Upside versus downside risk: Gender, stakes, and skewness.

Costa Jr, P. T., A. Terracciano, and R. R. McCrae (2001). Gender differences in personality traits across cultures: robust and surprising findings. *Journal of personality and social psychology 81*(2), 322.

Cullen, Z. and R. Perez-Truglia (2018). How much does your boss make? the effects of salary comparisons. Technical report, National Bureau of Economic Research.

Eckel, C. C. and P. J. Grossman (2008). Forecasting risk attitudes: An experimental study using actual and forecast gamble choices. *Journal of Economic Behavior & Organization 68*(1), 1–17.

Exley, C. L. and J. B. Kessler (2019). The gender gap in self-promotion. NBER Working Paper No. 26345.

Exley, C. L., M. Niederle, and L. Vesterlund (2019). Knowing when to ask: The cost of leaning in. *Journal of Political Economy Forthcoming*.

Haycock, L. A., P. McCarthy, and C. L. Skay (1998). Procrastination in college students: The role of self-efficacy and anxiety. *Journal of counseling & development 76*(3), 317–324.

Holt, C. A. and S. K. Laury (2002). Risk aversion and incentive effects. *American economic review 92*(5), 1644–1655.

Kessler, J. B., C. Low, and C. D. Sullivan (2019). Incentivized resume rating: Eliciting employer preferences without deception. *American Economic Review 109*(11), 3713–3744.

Leibbrandt, A. and J. A. List (2014). Do women avoid salary negotiations? evidence from a large-scale natural field experiment. *Management Science 61*(9), 2016–2024.

Milgram, N., S. Marshevsky, and C. Sadeh (1995). Correlates of academic procrastination: Discomfort, task aversiveness, and task capability. *The Journal of psychology 129*(2), 145–155.

Niederle, M. and L. Vesterlund (2007). Do women shy away from competition? do men compete too much? *The quarterly journal of economics 122*(3), 1067–1101.

Onwuegbuzie, A. J. (2004). Academic procrastination and statistics anxiety. *Assessment & Evaluation in Higher Education 29*(1), 3–19.

Quadlin, N. (2018). The mark of a woman's record: Gender and academic performance in hiring. *American Sociological Review 83*(2), 331–360.

Rammstedt, B. and O. P. John (2007). Measuring personality in one minute or less: A 10-item short version of the big five inventory in english and german. *Journal of research in Personality 41*(1), 203–212.

Reidy, D. E., P. D. Kernsmith, C. A. Malone, A. M. Vivolo-Kantor, and J. P. Smith-Darden (2018). Feminine discrepancy stress and psychosocial maladjustment among adolescent girls. *Child Psychiatry & Human Development 49*(2), 176–186.

Rigdon, M. L. (2012). An experimental investigation of gender differences in wage negotiations. *Available at SSRN 2165253*.

Russell, K. and P. Carter (2006). *Discover Your IQ Potential: Over 500 Tests of Your Mental Agility*. Arcturus Publishing.

Sarsons, H. and G. Xu (2015). Confidence men? gender and confidence: Evidence among top economists. *Harvard University, Department of Economics, Littauer Center*, 1–26.

Senecal, C., R. Koestner, and R. J. Vallerand (1995). Self-regulation and academic procrastination. *The journal of social psychology 135*(5), 607–619.

Shastry, G. K., O. Shurchkov, and L. Xia (2019). Luck or skill: How women and men react to noisy feedback.

Small, D. A., M. Gelfand, L. Babcock, and H. Gettman (2007). Who goes to the bargaining table? the influence of gender and framing on the initiation of negotiation. *Journal of personality and social psychology 93*(4), 600.

Solomon, L. J. and E. D. Rothblum (1984). Academic procrastination: Frequency and cognitive-behavioral correlates. *Journal of counseling psychology 31*(4), 503.

Weisberg, Y. J., C. G. DeYoung, and J. B. Hirsh (2011). Gender differences in personality across the ten aspects of the big five. *Frontiers in psychology 2*, 178.

Witkowski, J., R. Freeman, J. W. Vaughan, D. M. Pennock, and A. Krause (2018). Incentive-compatible forecasting competitions. In *Thirty-Second AAAI Conference on Artificial Intelligence*.

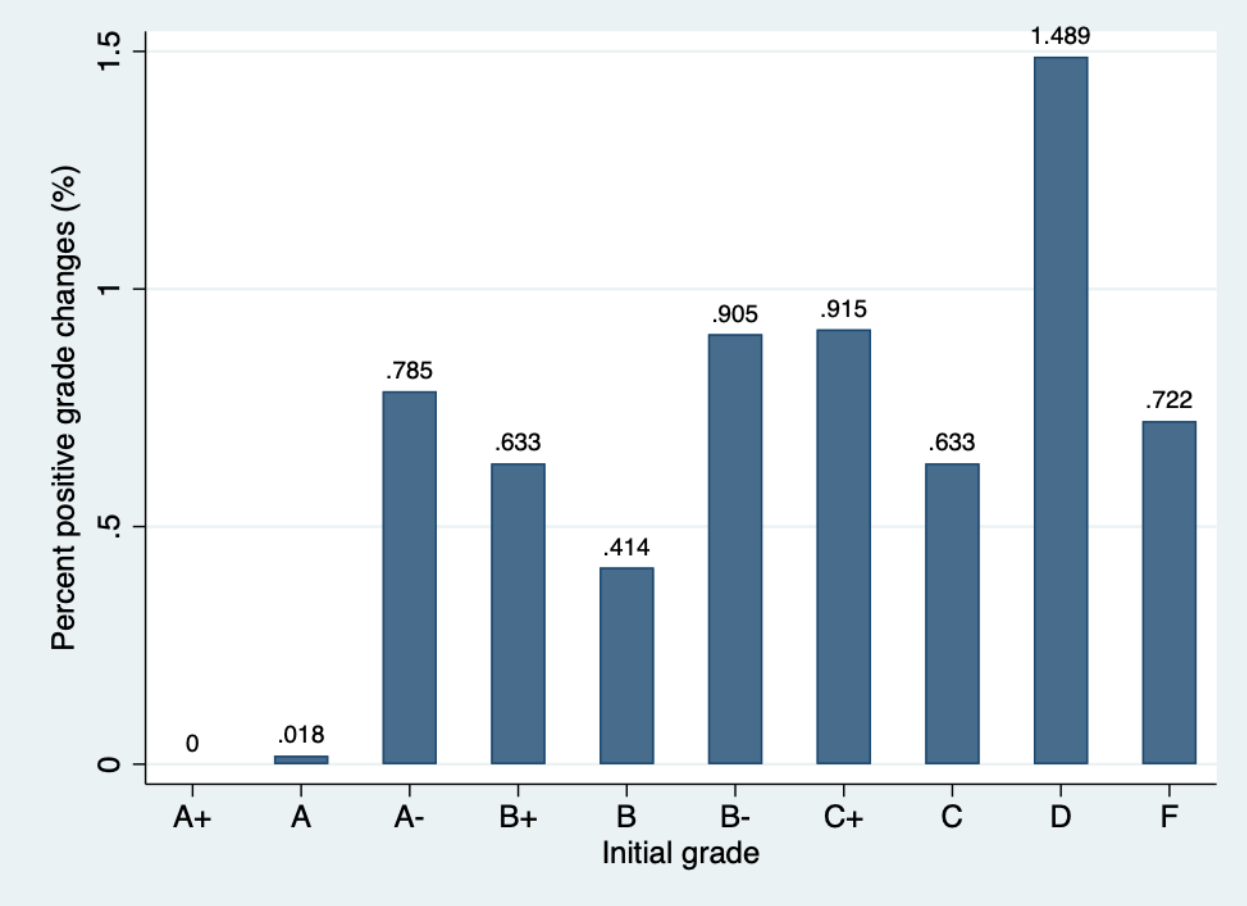

Figure 1: Percent upward grade changes conditional on initial grades

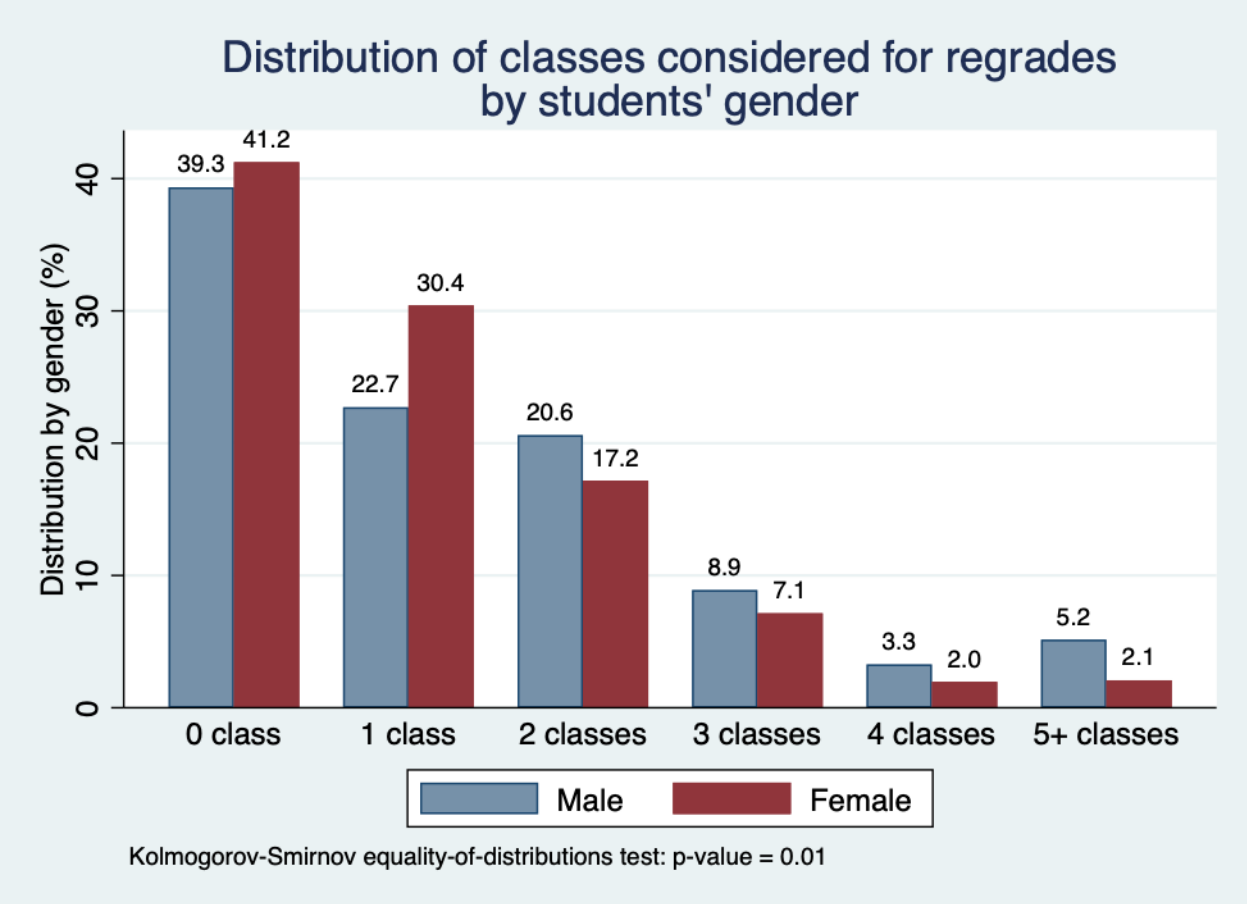

Figure 2: Student survey: number of classes students considered for regrade requests by students' gender

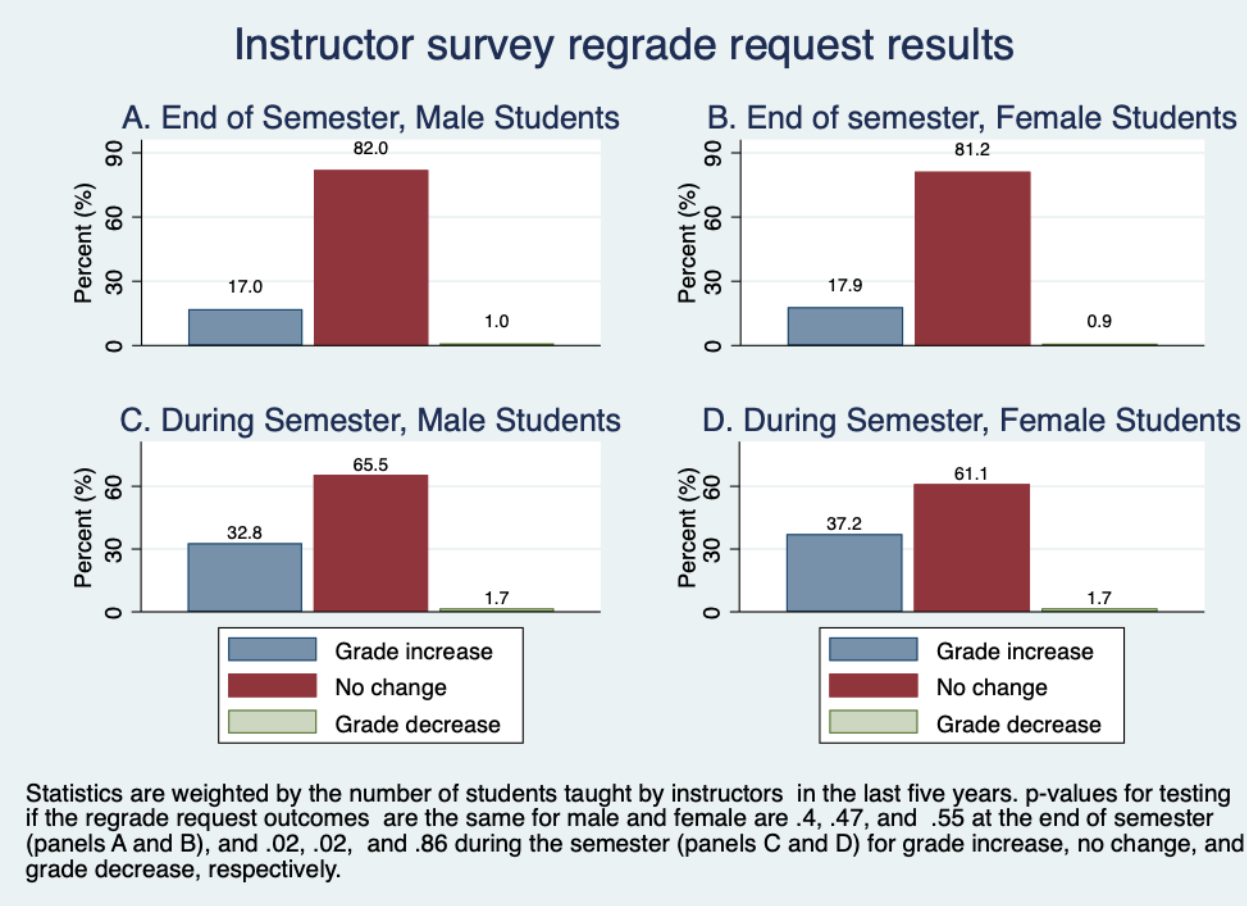

Figure 3: Instructor survey: regrade results by timing and students' gender

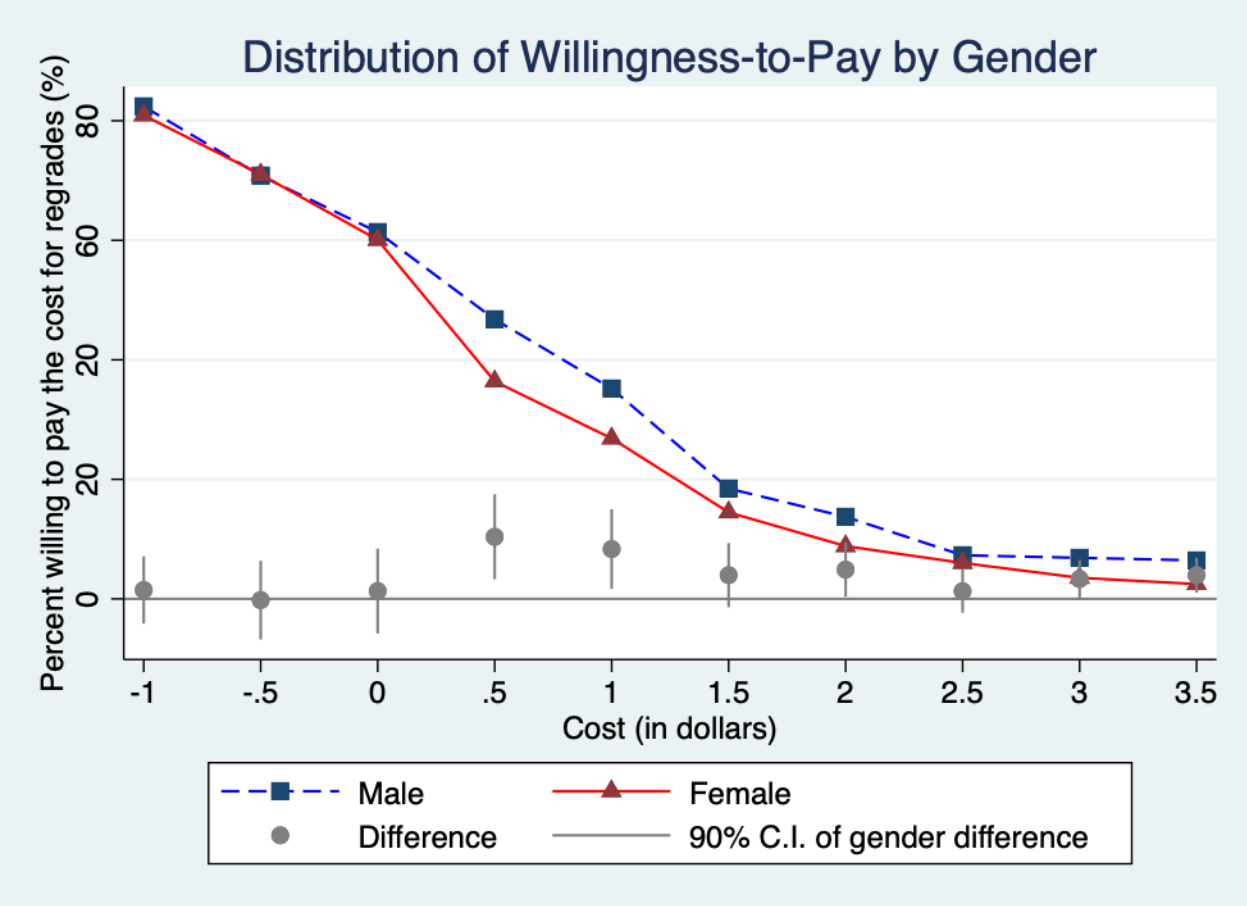

Figure 4: Regrade requests at varying costs by students' gender

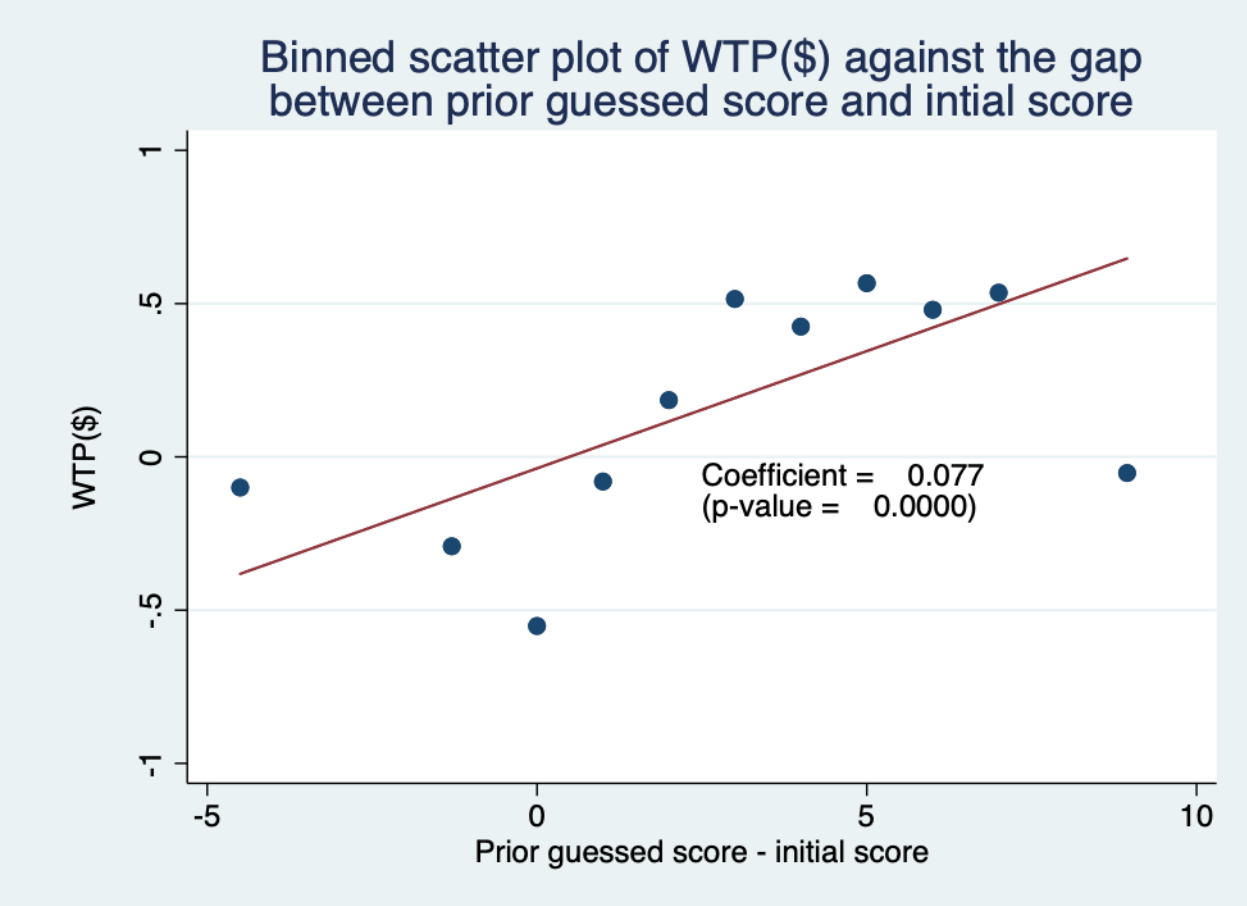

Figure 5: Binscatter plot of WTP($) against the gap between prior guessed score and intial score

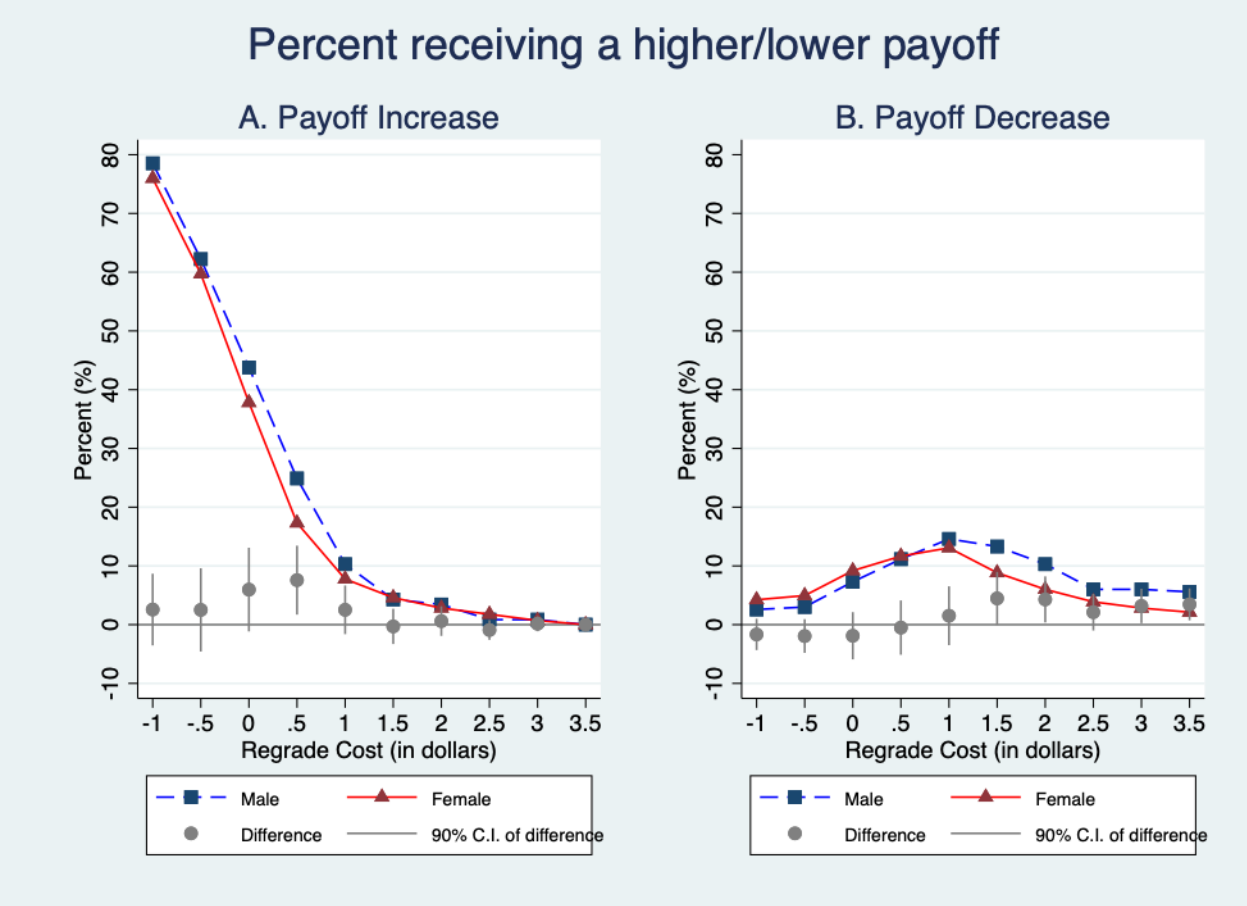

Figure 6: Percent grade/payoff increase/decrease at varying regrade costs by students' gender

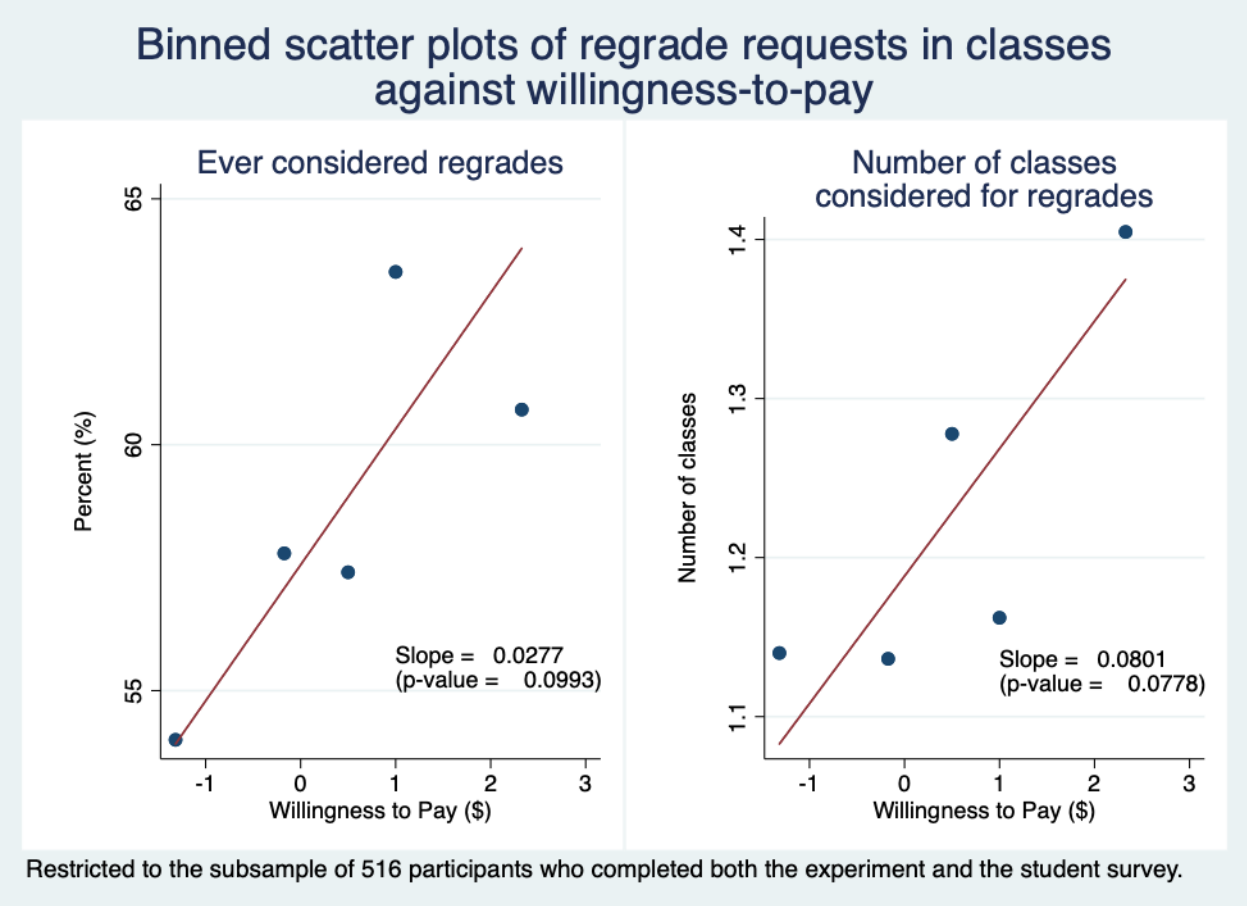

Figure 7: Correlation between the WTP and regrade experience in class

Table 1: Summary statistics of administrative data

	Students			
	All	Female	Male	Gender Difference
Percent (%) of records	100	53.4	46.6	
Percent (%) grades with a...				
Grade change	0.464	0.424	0.509	-0.085***
	(6.80)	(6.50)	(7.12)	[0.012]
Positive grade change	0.439	0.404	0.479	-0.075***
	(6.61)	(6.34)	(6.90)	[0.011]
Negative grade change	0.025	0.021	0.031	-0.010***
	(1.59)	(1.43)	(1.75)	[0.003]
Female instructor[a]	44.9	50.0	38.9	11.1***
	(49.7)	(50.0)	(48.8)	[0.088]
Days between grade changes	25.8	24.9	26.7	-1.80
	(56.6)	(55.0)	(58.0)	[1.43]
GPA	3.04	3.13	2.94	0.18***
	(0.59)	(0.57)	(0.60)	[0.01]
Percent (%) in...				
College of Agriculture	4.5	4.9	4.0	0.87***
College of Business	10.8	8.6	13.2	-4.58***
College of Engineering	4.8	1.9	8.2	-6.35***
College of Human Sciences	12.6	15.5	9.1	6.40***
College of Liberal Arts	32.2	33.6	30.6	3.07***
College of Natural Resources	4.6	3.5	5.9	-2.32***
College of Natural Sciences	23.3	23.2	23.4	-0.25***
College of Veterinary Sciences	3.4	4.4	2.3	2.07***
Intra-University	3.9	4.4	3.3	1.07***
N	1,341,552	716,772	624,780	

*$p < 0.10$, **$p < 0.05$, ***$p < 0.01$. Standard deviations are in parentheses. Standard errors are in the square brackets. a. 85,250 administrative records have missing values for instructors' gender. Excluding these data points, the numbers of observations are 671,276 and 585,026 for female and male student records, respectively.

Table 2: Linear probability regression analysis of administrative data

	[1]	[2]	[3]	[4]	[5]
A. Dependent variable: Upward grade change $Y \in \{0, 1\}$					
Male student	0.000748***	0.000665***	0.000685***	0.000763***	0.000698***
	(0.000118)	(0.000120)	(0.000127)	(0.000129)	(0.000128)
	[1.186]	[1.164]	[1.167]	[1.188]	[1.171]
Observations	1,341,552	1,341,552	1,256,302	1,256,302	1,256,302
B. Dependent variable: Downward grade change $Y \in \{0, 1\}$					
Male student	0.000102***	0.000085***	0.000023	0.000001	-0.000008
	(0.000029)	(0.000028)	(0.000028)	(0.000028)	(0.000028)
	[1.499]	[1.399]	[1.091]	[1.005]	[0.969]
Observations	1,341,552	1,341,552	1,256,302	1,256,302	1,256,302
Student controls	No	Yes	No	Yes	Yes
Class controls	No	No	Yes	Yes	Yes
GPA × Grade × Dept	No	No	No	No	Yes

*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$. Standard errors are clustered at the student level and reported in parentheses.
Male-to-female odds ratios (calculated as $\frac{Pr_{Male}(Y=1)/Pr_{Male}(Y=0)}{Pr_{Female}(Y=1)/Pr_{Female}(Y=0)}$ assuming the mean values for all variables other than *Male* student are reported in brackets.
a. Student controls include: students' class standing (i.e., freshman, sophomore, junior, senior), discretized GPA (corresponding to letter grades), and initial grade.
b. Class controls include: college, department, and instructors' gender and rank.
c. GPA × Grade × Dept includes 4,500 indicators for the full interactions of discretized GPAs (corresponding to letter grades), initial grades, and departments.

Table 3: Sensitivity analysis of linear probability model using administrative data

	[1] Baseline	[2] Logit[a]	[3] Excl. students w/2+ changes	[4] Excl. instructors w/10%+ changes	[5] Excl. A+ and A	[6] Excl. F	[7] Large class[b]
A. Dependent variable: Upward grade changes $Y \in \{0,1\}$							
Male student	0.000698***	0.000829***	0.000583***	0.000565***	0.001045***	0.000715***	0.000724***
	(0.000128)	(0.000132)	(0.000112)	(0.000122)	(0.000191)	(0.000128)	(0.000219)
	[1.171]	[1.208]	[1.165]	[1.152]	[1.170]	[1.179]	[1.188]
Observations	1,256,302	1,255,910	1,243,781	1,252,854	819,624	1,212,024	360,618
B. Dependent variable: Downward grade changes $Y \in \{0,1\}$							
Male student	-0.000008	0.000008	-0.000016	-0.000019	0.000016	-0.000008	0.000024
	(0.000028)	(0.000038)	(0.000028)	(0.000023)	(0.000033)	(0.000029)	(0.000063)
	[0.969]	[1.025]	[0.938]	[0.899]	[1.064]	[0.969]	[1.080]
Observations	1,256,302	1,028,958	1,255,987	1,252,854	819,624	1,212,024	360,618

$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$. Standard errors are clustered at the student level and reported in parentheses.
Male-to-female odds ratios (calculated as $\frac{Pr_{Male}(Y=1)/Pr_{Male}(Y=0)}{Pr_{Female}(Y=1)/Pr_{Female}(Y=0)}$ assuming the mean values for all variables other than *Male* student are reported in brackets.
All models control for students' class standing (i.e., freshman, sophomore, junior, senior), discretized GPA (corresponding to letter grades), and initial grade, college, department, and instructors' gender and rank, and 4,500 fully interacted indicators of GPA × Grade × Department.
a. Logit regression uses continuous values of GPA and Grade for convergence consideration.
b. Classes with more than 90 students.

Table 4: Summary statistics of the instructor survey

	End of semester	During semester	Difference
A. Regrade requests			
Percent requested regrades in a class (%)	5.94	11.17	5.23***
	(7.18)	(11.32)	[1.01]
Male students among requests (%)	57.08+	58.01++	0.93
	(24.69)	(21.44)	[2.09]
B. Instructors' report on students aggressiveness in regrade requests			
Males more aggressive (%)	51.18	37.70	-13.48**
	(50.19)	(48.66)	[5.68]
Similarly aggressive (%)	37.77	48.40	10.63*
	(48.68)	(50.18)	[5.69]
Females more aggressive (%)	11.05	13.90	2.85
	(31.48)	(34.74)	[2.95]

$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}\ p < 0.01$ indicate the significance level of the differences between end of the semester and during semester.
$^{+}p < 0.10$, $^{++}p < 0.05$, $^{+++}p < 0.01$ indicate the significance level of differences between regrade requests at the end of/during the semester against the male representation (52.5%) in the class. Standard deviations are in parentheses. Standard errors are in square brackets. All statistics are weighted by class size.

Table 5: Summary statistics of the student survey

A. All students	Female	Male	Difference
A1. Regrade requests			
Percent ever considered regrades (%)	58.76	60.66	-1.90
	(49.26)	(48.91)	[2.90]
Percent ever asked for regrades (%)	38.94	40.98	-2.04
	(48.79)	(49.24)	[2.89]
Percent ever asked for regrades, end of semester (%)	16.47	17.33	-0.86
	(37.12)	(37.90)	[2.21]
Percent ever asked for regrades, during semester (%)	29.03	31.62	-2.58
	(45.42)	(46.55)	[2.71]
Number of classes considered regrades	1.04	1.30	-0.25***
	(1.17)	(1.41)	[0.07]
Number of classes asked for regrades	0.56	0.65	-0.09*
	(0.81)	(0.91)	[0.05]
Imputed number of classes asked for regrades	0.97	1.12	-0.15**
	(0.99)	(1.07)	[0.06]
A2. Conditional on considered/asked for regrades			
Number of classes considered regrades	1.78	2.14	-0.25***
	(1.02)	(1.22)	[0.07]
Number of classes asked for regrades	1.43	1.59	-0.09*
	(0.66)	(0.73)	[0.05]
Imputed number of classes asked for regrades	1.65	1.84	-0.15**
	(0.74)	(0.74)	[0.06]
A3. Student Perceptions			
Probability of grade increased if asked (%)	37.04	38.25	-1.21
	(28.51)	(27.98)	[1.68]
Probability of grade unchanged if asked (%)	57.22	55.91	1.31
	(28.32)	(28.50)	[1.68]
Probability of grade decreased if asked (%)	5.74	5.83	-0.10
	(11.72)	(11.45)	[0.69]
Regret asking aggressively, during semester[a]	1.49	1.63	-0.14**
	(1.09)	(1.22)	[0.07]
Regret asking aggressively, end of semester[a]	1.44	1.60	-0.15**
	(1.07)	(1.22)	[0.07]
Control over direction of life[b]	5.14	5.28	-0.14*
	(1.23)	(1.33)	[0.07]
Stress from asking for regrades[c]	5.01	3.88	1.12***
	(1.63)	(1.89)	[0.10]
Number of participants	868	427	
B. Sub-sample of regrade requests	Female	Male	Difference
Percent grade component increased (%)	59.59	65.23	-5.65
	(49.12)	(47.71)	[3.65]
Percent grade component decreased (%)	1.44	0.00	1.44**
	(11.94)	(0.00)	[0.71]
Percent final grade increased (%)	31.34	31.54	-0.20
	(46.44)	(46.55)	[3.49]
Percent final grade decreased (%)	2.68	6.09	-3.41**
	(16.17)	(23.96)	[1.46]
Number of participants	338	175	
Number of observations	485	279	

*$p < 0.10$, **$p < 0.05$, ***$p < 0.01$ indicate the significance level of gender differences. Standard deviations are in parentheses. Standard errors are in square brackets. a. Scale ranges between 1 (never true to me) and 7 (always true to me). b. Scale ranges between 1 (I don't have any control) and 7 (I have full control). c. Scale ranges between 1 (not stressed at all) and 7 (extremely stressed).

Table 6: Summary statistics of the experiment

		Female	Male	Difference
A.	Risk aversion coefficient	1.12	0.86	0.25***
		(0.75)	(0.73)	[0.07]
	Over-confidence[a]	0.53	0.54	-0.02
		(0.19)	(0.18)	[0.02]
	Under-confidence[b]	0.21	0.17	0.04***
		(0.16)	(0.14)	[0.01]
	Over-optimism[c]	1.64	1.37	0.27
		(3.28)	(2.73)	[0.27]
	Uncertainty[d]	0.51	0.47	0.04*
		(0.23)	(0.21)	[0.02]
	Prior guessed score	13.21	14.35	-1.14***
		(3.30)	(2.69)	[0.27]
	Prior downside risk[e]	0.33	0.29	0.04**
		(0.23)	(0.20)	[0.02]
	Posterior guessed score	12.10	13.28	-1.19***
		(3.11)	(2.90)	[0.27]
	Posterior downside risk[e]	0.28	0.22	0.05**
		(0.26)	(0.23)	[0.02]
	Belief revision[f]	-1.11	-1.07	-0.04
		(2.73)	(2.33)	[0.23]
B.	Big 5 trait: Extroversion[g]	3.03	2.96	0.07
		(0.97)	(0.94)	[0.08]
	Big 5 trait: Agreeableness[g]	3.62	3.48	0.13*
		(0.85)	(0.82)	[0.07]
	Big 5 trait: Conscientiousness[g]	3.74	3.64	0.10
		(0.77)	(0.80)	[0.07]
	Big 5 trait: Neuroticism[g]	3.52	2.89	0.63***
		(1.03)	(0.94)	[0.09]
	Big 5 trait: Openness[g]	3.71	3.47	0.24***
		(0.90)	(0.83)	[0.08]
C.	True score of the quiz	11.57	12.98	-1.41***
		(3.19)	(2.87)	[0.27]
	Original score of the quiz	10.84	12.08	-1.24***
		(2.87)	(2.69)	[0.25]
	Final score of the quiz	11.13	12.45	-1.32***
		(3.00)	(2.91)	[0.26]
	WTP ($)	0.05	0.25	-0.19*
		(1.22)	(1.38)	[0.11]
	% Willing to Pay Positive Cost	36.40	46.78	-10.39**
	Percent payoff increased (%)[h]	20.85	22.92	-2.07
	Percent payoff decreased (%)[h]	6.68	7.98	-1.30
	Observations	283	233	

* $p < 0.10$; ** $p < 0.05$, *** $p < 0.01$. Standard deviations are in parentheses. Standard errors are in square brackets. a. *Over-confidence* measures participants' probability assignments to the *wrong* answers in the quiz (0 = zero over-confidence to 1 = completely over-confident). b. *Under-confidence* measures the gap (between 0 and 1) between participants' probability assignments and 1 (100%) of the *correct* answers in the quiz (0 = zero under-confidence to 1 = completely under-confident). c. *Over-optimism* measures the gap between participants' guessed score (prior belief) and true score of the quiz. d. *Uncertainty* measures how certain (in probability) they are about their guessed outcome (0 = completely certain to 1 = completely uncertain). e. *Prior* (*posterior*) *downside risk* measures the prior (posterior) probability (between 0 and 1) participants assigned to adverse outcomes (i.e., true score is below the guessed score). f. *Belief updates* measures the updates in the guessed score (posterior guessed score - prior guessed score). g. *Extroversion*, *Agreeableness*, *Conscientiousness*, *Neuroticism*, *Openness* are the Big Five personal traits with the value ranging between 1 and 5. h. Mean value over the 10 cost scenarios.

Table 7: Logit regression of willingness to pay a positive cost for regrades

	Dependent Variable: Participants who are willing to pay a positive cost for regrades $\in \{0, 1\}$								
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]
Male student	0.104**	0.099**	0.096**	0.084*	0.073*	0.080*	0.068	0.083*	0.057
	(0.043)	(0.044)	(0.043)	(0.044)	(0.044)	(0.044)	(0.044)	(0.047)	(0.047)
Risk aversion[a]		-0.018					-0.013		-0.014
		(0.029)					(0.029)		(0.029)
Over-confidence[b]			0.345***				0.219		0.235*
			(0.113)				(0.133)		(0.134)
Under-confidence[c]				-0.450***			-0.035		-0.036
				(0.152)			(0.188)		(0.189)
Over-optimism[d]					0.001		-0.003		-0.003
					(0.008)		(0.012)		(0.012)
Uncertainty[e]					-0.289*		-0.276		-0.271
					(0.167)		(0.170)		(0.170)
Prior guessed score					0.017*		0.017		0.016
					(0.009)		(0.016)		(0.016)
Prior downside risk[f]					-0.038		-0.014		-0.021
					(0.172)		(0.174)		(0.175)
Posterior guessed score						0.013*	-0.004		-0.005
						(0.008)	(0.013)		(0.013)
Posterior downside risk[f]						-0.166*	-0.011		-0.003
						(0.094)	(0.103)		(0.104)
Extroversion[g]								-0.008	-0.000
								(0.024)	(0.024)
Agreeableness[g]								-0.025	-0.028
								(0.026)	(0.026)
Conscientiousness[g]								-0.007	-0.010
								(0.028)	(0.028)
Neuroticism[g]								-0.020	-0.007
								(0.023)	(0.023)
Openness[g]								-0.014	-0.015
								(0.025)	(0.024)
Observations	516	516	515	516	516	516	515	516	515
F-test (p-value): all coefficients (other than Male student and Constant) = 0							.0065	.842	.041

* $p < 0.1$, ** $p < 0.05$, and *** $p < 0.01$ indicate the significance level of gender differences. Marginal effects are reported in the table. Standard errors are reported in parentheses. a. *Risk aversion* measures the risk aversion coefficient. b. *Over-confidence* measures participants' probability assignments to the *wrong* answers in the quiz (0 = zero over-confidence to 1 = completely over-confident). c. *Under-confidence* measures the gap (between 0 and 1) between participants' probability assignments and 1 (100%) of the *correct* answers in the quiz (0 = zero under-confidence to 1 = completely under-confident). d. *Over-optimism* measures the gap between participants' guessed score (prior belief) and true score of the quiz. e. *Uncertainty* measures how certain (in probability) they are about their guessed outcome (0 = completely certain to 1 = completely uncertain). f. *Prior* (*posterior*) *downside risk* measures the prior (posterior) probability (between 0 and 1) participants assigned to adverse outcomes (i.e., true score is below the guessed score). g. *Extroversion*, *Agreeableness*, *Conscientiousness*, *Neuroticism*, *Openness* are the Big Five personal traits with the value ranging between 1 and 5.