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

Cher Hsuehhsiang Li*
Colorado State University

Basit Zafar†
Arizona State University

Preliminary (Do Not Cite or Circulate)

August 23, 2019

Abstract

Using transcript 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 initiated by instructors. These gender differences cannot be explained by observable characteristics of the students, instructors, and the classes. To understand the mechanisms underlying these gendered outcomes, we conduct surveys of students and teachers, which reveal that not only are regrade requests prevalent but that male students are also more likely than female students to ask for regrades on the intensive margin. Finally, we corroborate the gender differences in regrade requests in an incentivized controlled experiment where participants receive noisy signals of their performance, and where they can ask for regrades: we find that males have a higher willingness to pay (WTP) to ask, both in cases where it makes economic sense to do so and where it does not. About a third of the gender difference in the WTP is due to gender differences in under-confidence, uncertainty in prior beliefs about ability, and the Big Five personality traits.

*E-mail:.hsueh-hsiang.li@colostate.edu.

†E-mail: basitak@gmail.com.

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 contributes to a sizable portion of the gender gap in salaries and career advancement (Babcock and Laschever, 2009). 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.

To investigate this, we analyze a unique transcript 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. This transcript dataset includes different reasons of grade changes that allow us to distinguish changes that are resulted by 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. 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.

These 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; 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 transcript 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. Another limitation of the transcript data is that it, by construction, it contains no information on any requests that were made but resulted in no grade change.

To understand the origins of these patterns, we next administered surveys to instructors and students. These surveys elicited respondents' recollections on regrade requests in the past semesters. The surveys reveal 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 (primarily 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.

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 performance on 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 (BDM) method 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. 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.

We find no gender difference 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. The higher 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. 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. 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 that male students are more likely to report regret for asking for regrades too aggressively.

So what explains the gender differences in the WTP to ask? 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; Sarsons

and Xu, 2015). Likewise, consistent 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 beliefs and personality traits can explain about a third of the gender difference in the propensity to ask; we find no role of risk preferences. Nonetheless, the remaining two-thirds of the gender gap remains unexplained. Our survey evidence shows that females report higher psychic costs (specifically, stress) when having to ask an instructor for a regrade. Thus, policies that lower costs of asking are likely to reduce gender differences in asking. In addition, reducing under-confidence and uncertainty in beliefs (through policies that provide more accurate signals to individuals) are likely to reduce gender differences in the WTP to ask.

Our paper is related to the literature on gender differences in asking. We know 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 the study without bargaining. Similarly, Babcock and Laschever (2009) 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; this gender difference in engaging in wage negotiations contributes 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 observing the negotiation behaviors by real job applicants. They find that when salaries were not explicitly made negotiable in the advertisement, men still negotiated for a higher wage, but women inclined to signal their willingness to accept a lower wage offer. However, the gender differences in wage negotiation disappears when salaries were posted as negotiable. Exley et al. (2016) find that women attain worse returns from negotiation when it is mandatory than when it is optional. We complement this literature by showing that females are simply less likely to ask, especially when there is a positive cost for doing so. Prior studies primarily focus on the impact of gender differences in negotiations on the 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 both before and after individuals enter the labor market. 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 the 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.

The next section presents the analysis of the administrative transcript 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 transcript 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 thinking score, and a 575.5 SAT math score (compared to national averages of 494 and 508, respectively). Among these freshmen, approximate 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 percent), CSU was less ethnically diverse than the average U.S. university (43 percent).[1]

Key for our purposes, this administrative dataset recorded 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 submitted grade changes in the computing system. There were 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 had 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. If the grade changes were based on students' own actions (e.g., taking a repeat-and-delete option by re-taking the same course to override

[1]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 races 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.

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. 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., when an initial grade was changed to a better grade). The overwhelmingly large fraction of upward corrections among grade changes indicates that the 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 (0.03 percent) were 50-percent more likely than their female counterparts (0.02 percent) to receive such an adverse outcome.

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, Health and Human Sciences, Liberal Arts, Veterinary Medicine and Biomedical Sciences, and Intra-University. Nevertheless, the gender gap in grade changes was pervasive across

colleges and departments.[2]

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. Other than these "A students", the positive grade changes do not demonstrate particular patterns conditional on the initial grades. 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 from 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 frequencies of grade changes for those receiving A-, B-, and C+ also indicate that the upward grade changes are not limited to the 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.[3] Among all grade changes, over 55 percent of the grade changes are moved up by one letter grade. The intensity of the grade changes on the margin implies that instructors may be sympathetic towards students when their grade was close to the margin between two letter grades. It is also worth noting that fewer grade changes are observed in large classes (i.e., defined as classes with over 180 students) where computerized grading system is frequently used and little instructor latitude is exercised.[4]

[2]The gender differences in grade changes by college and department are detailed in the Appendix.

[3]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 during the same semester. However, when the student's grade in a class is below her average GPA in the same semester, the propensity to receive an upward grade correction increases in the difference between the GPA and the class grade. This pattern is robust to excluding individual class grades from the GPA calculations in the same semester.

[4]Approximately 9-percent of transcript records were from large classes (i.e., classes of more than 180 students) where grade changes occurred only 0.22 percent of the time, compared with 0.49 percent for the remaining classes.

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:

$$Y_{ijt} = \Phi(\alpha_0 + \alpha_1 Male_i + \alpha_2 W_{ijt} + \alpha_3 X_{it} + \alpha_4 Z_{jt} + \epsilon_{ijt}), \tag{1}$$

In this equation, Y_{ijt} is a binary variable and assumes the value of one when the 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 for those with no such changes; 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 that in semester t, such as the student's class standing (i.e., freshman, sophomore, junior, and senior) and term GPA; Z_{jt} captures class-specific information, such as the instructor's gender and rank (i.e., tenured tracked assistant professor, tenured tracked associate professor, tenure-tracked full professor, non-tenure tracked instructor, graduate teaching assistant), department, and college, 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.

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 in upward grade change is 0.0748 percentage points. Adding controls for the colleges and departments in which the courses were offered reduces the gender gap slightly to 0.0705 percentage points (Column 2). In Column 3, controlling for the instructor's gender and rank, the estimate is reduced slightly to 0.0675 percentage points. This reduction in gender gap suggests some gender sorting among students into classes taught by different types of instructors who may have varying degrees of leniency in regrades. Column 4 controls for the students' class standing (freshman, sophomore, junior, and senior) because students may learn and improve their odds of successfully convincing instructors to change their grades over the years in college. Since the average time-to-graduation is longer for males then for females, the increasing presence of male students in upper-class may overstate the gender bias. We find that the gender gap in upward grade corrections is hardly changed when we control for students' class standing. Finally, adding additional controls for students' initial grade and GPA in Column 5 of Panel A results in the widening of the gender gap because of differential regrade rates by gender within the initial grade. Although both the GPA and the initial grades significantly influence the likelihood of upward grade changes for students, they fail to explain the gender gap in upward regrades.

We analyze the same equation with the downward regrade outcomes 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 instructor's gender and rank and students' class standing, GPA, and initial grades, the relative adverse outcome for males disappears when we control for the college and department fixed effects.

Including all the aforementioned control variables (i.e., colleges, departments, instructor's gender and rank, students' class standing, GPA, and initial grade), Column 1 of Panel C in Table 2 reports the baseline estimate of a 0.0780 percentage points male advantage in upward regrades that cannot be explained by class and student observable characteristics. To investigate the sensitivity of the estimate, we first remove students who received multiple upward regrades from the analysis. We find that the gender gap reduces but remains substantial at 0.0642 percentage points in Column 2 of Panel C. 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 3 of Panel C in Table 2 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 0.078 to 0.0651 percentage points and remains significant).

Next, we consider that students whose grade hit the upper bound (A+ and A) may not need to request grade changes. We remove them from the analysis in Column 4 of Panel C in Table 2. The gender gap actually increases to 0.118 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 becomes 0.083 percentage points — similar to the magnitude in the baseline model. Because a large number of instructors and classes do not have any grade change records, we are unable to reliably estimate the gender differences in regrades by controlling for instructor or class fixed effects because these observations will be dropped from the probit or logit regression.[5] To test how sensitive the results are to the idiosyncratic

[5]The transcript records show that 28.3 percent of the records would be dropped because they were taught by instructors who never changed grades, and 55.6 percent of the records would be dropped because there was no grade correction made in these classes.

variation across instructors and classes, we estimate a linear probability model instead. We find that instructor and class fixed effects reduce the gender gap in regrades but it remains substantial and significant.[6]

We conduct a similar sensitivity analysis for downward grade changes in Panel D of Table 2. When we control for all these aforementioned independent variables, the gender difference in downward regrades becomes trivial (0.0006 percentage points) and insignificant as presented in the baseline model in Column 1 of Panel D in Table 2. The absence of gender differences in downward regrades is persistent across the sensitivity analyses (Columns 2 through 5).

In summary, we find robust gender differences in regrades, with males significantly more likely to experience any grade changes, in particular, upward grade changes. 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; 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 transcript records do not contain any information that allows us to understand which scenario accounts for these unequal outcomes. Another limitation of the administrative data is that it, by construction, contains no information on any requests that were made but resulted in no grade change.

To understand these 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

This section reports results from the instructor and student surveys, that enable us to better understand the patterns documented in the administrative data.

[6]Instructor fixed effects reduce the gender gap to 0.059 percentage points and the class fixed effects reduces this gap to 0.046 percentage points, but the gender gap in upward grade changes remain statistically significant at the 1-percent level.

3.1 Samples

For the instructor survey, 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 slightly 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.[7]

For the student survey, 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.[8] The average (median) time taken to complete the survey was 9.44 (7.99) minutes.[9]

[7] 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.

[8] 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.

[9] We measure the time participants (both participants who completed the survey only and participants who completed both the survey and experiment) spent on the survey pages. 86 out of 16,042 pages (0.53%) show a duration lasting 10 minutes or more. In this case, participants were likely taking a break on pages where we did not impose a time limit. We capped the time spent on a page at 10 minutes in this case.

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).[10] 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.[11] Although the student survey was slightly 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' perceptions regarding the outcomes of a regrade request. Specifically, we elicited 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. Compared with the instructor survey, students reported higher frequencies of regrade requests – 16.8 percent of students reported that they requested regrades at the end of a semester and 29.9 percent during the semester. This difference may be a result of the gap in respondents' understanding of the definition of regrade requests. For instance, while many students reported regrade requests due to grading errors, instructors may dismiss simple error corrects as regrades.

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. Table 3 shows that male students made up

[10] 1,304 participants completed the survey (741 completed the survey only, and 563 completed both the survey and the experiment). Among these entries, 10 participants submitted their responses through both links (one for the survey only, and the other for the combination of the survey and the experiment). We removed the duplicate entries from the 10 participants by keeping only their entries in the combination of the survey and experiment.

[11] The distribution of participants in the student survey (transcript 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.

57.1 percent of regrade requests at the end of the semester — a larger proportion than the male representation in class (52.5 percent) in the instructor survey.

Results from the student survey offer further insights into gender differences in asking. First of all, asking for grade changes are fairly common among college students. Table 4 shows that 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). However, a gender difference emerges on the intensive margin: 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. Figure 3 shows the distributions of regrade consideration and the distributions are statistically different by gender. Note that the number of classes students may report 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.[12] The imputed number of classes in which students actually asked for regrades is 0.59 for females and 0.73 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, and not on the extensive 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) of the intensive margin of regrade requests, males request (or consider requesting) about 16 to 24 percent more than females do. Interestingly, the magnitude is similar to the transcript evidence where male students are 18.5 percent more likely than their female counterparts to receive favorable regrade outcomes.

3.2.2 Are male students treated more favorably when they ask?

With regards to the outcomes of regrade requests, Panel C of Table 3 and 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, the outcomes are actually more favorable for female than for male students. Panel D of Table 3

[12]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 0 and 3 to calculate the request ratio (number of classes students actually asked for regrades / number of classes students considered for regrades) by students' gender. We then impute the requests by multiplying the gender-specific ratio (depending on students' gender) by the number of classes considered for regrades that is greater than three (i.e., 4 or 5).

shows instructors' overall impressions about the success of regrade requests. The majority (75.9 percent) of the instructors (weighted by class sizes) further reported an impression that male and female students were equally successful in regrade requests. 21.8 percent of instructors (weighted by class sizes) reported that female students were more successful than male students in regrades, while only 2.27 percent of instructors (weighted by class sizes) reported that male students were more successful than female students in regrade outcomes. These results from the instructor survey suggest that most instructors treated regrade requests by male and female students equally. If there is any difference, it seems to favor female students.

The student survey presents a consistent pattern in regrade outcomes that are observed in the instructor survey. Summarized in Panel B of Table 4, conditional on asking, students experienced a higher success rate in a grade increase during the semester than the final grade increase at the end of the semester regardless of students' gender. During the semester, over 60 percent of regrade requests (67 percent of male requests and 61 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 (2 percent of male requests and 0 percent for female requests). At the end of the semester, male (female) students reported that 34 percent (31 percent) of the cases resulted in final grade increases, 60 percent (66 percent) of the cases did not observe any grade changes, and 6 percent (3 percent) of the cases led to a lower grade. Note that the gender differences in regrade outcomes are not statistically significant. Furthermore, Panel A of Table 4 shows that the subjective beliefs in potential outcomes of regrade requests were also very similar across the two genders. 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.[13] 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. However, female students do report a significantly higher stress level than males 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. This suggests that females have higher psychic costs of asking. We discuss the implications of gender differences in stress and personality traits on regrades in Section 5.

[13]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. 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.

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 procrastinate more than females, and put off such requests till the end of the semester; the literature on gender differences in procrastination has found mixed result (Balkis and Duru, 2009; Haycock et al., 1998; Milgram et al., 1995; Onwuegbuzie, 2004; Senecal et al., 1995; Solomon and Rothblum, 1984).

The survey does not show any evidence of this. In the instructor survey (Panel A of Table 3), 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, 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 (primarily 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. 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, and uncertainty perceptions, 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 and allow participants to request a regrade at varying levels of costs. If a participant decides to pay a given cost to ask for a regrade, the true grade is revealed. The laboratory setting is advantageous over observational field evidence because we are able to control and eliminate factors, such as fear of backlash or gender interactions, that may contribute to the gender differences in regrades. 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 measured students' risk preferences through their incentivized choices of different lottery offers following Eckel and Grossman (2008) to test how much of the gender

differences in grade change requests are attributable to gender differences in risk aversion. We further elicit students' beliefs in their performance in the experiment to construct measures of over-confidence, under-confidence, over-optimism, and uncertainty in beliefs. These measures allow us to decompose the effects of the varying 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 settings 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 7 tasks:

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 \$1 (Event B), and 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 IQ multiple-choice questions derived primarily from Russell and Carter (2006).[14] 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).

Tasks 3 and 4 (prior beliefs): After completing the quiz, students are asked for 1) their "prior" belief about the number of questions that thought they answered correctly; and

[14]We sample few questions online and shorten some questions in Russell and Carter (2006) to allow participants to answer them under the time limit for each question (45 seconds).

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 5 (willingness to ask for regrade): 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 transcript records, we assigned the uncertainty/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 to receive a higher grade would be higher than receiving a lower grade.

Once the initial score is revealed, the detailed grading (i.e., correct or incorrect) and the subjective probability that the student assigns to each answer being correct is shown on the screen for each question. The student is then informed that she could make a grade request. Specifically, we use a multiple-price-list variation of the Becker-DeGroot-Marschak (BDM) 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.[15] One of the cost scenarios is randomly drawn for implementation. If the student chooses to request a regrade, she pays the associated regrade cost 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.

[15]The 10 regrade cost scenarios were 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 gain possible 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).

Tasks 6 and 7 (posterior beliefs): Using the same approach as for Task 2, we re-elicit participants' "posterior" beliefs about the number of accurate answers and the uncertainty around the posterior belief. The posterior uncertainty is within 3 questions above and below the initially revealed score by construction. Therefore, students are asked for the likelihood that the actual grade is in seven cases centered around the initial score (i.e., the actual grade is 3 questions below the initial score, 2 questions below the initial score, 1 question below the initial score, same as the initial score, 1 question above the initial score, 2 questions above the initial score, and 3 questions above their guess). The incentives formula is the same for the posterior beliefs as for the prior beliefs mentioned earlier. One of the four subjective beliefs (Tasks 3, 4, 6, and 7) is randomly drawn to determine a bonus payment up to $5.00 for participants.

The experiment also has built-in checks and test questions to make sure that participants understand the instructions. The Appendix shows the screenshots 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 taking 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 a subset of the respondents, we have both experimental and survey data. The experiment (excluding the survey) took 24.5 minutes, on average, to complete.[16] 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).[17] The results of our analyses are robust to inclusion of these students (see the Appendix C2 and C3). The final sample consists of 516 participants with 283 female

[16] We measure the time participants spent on the experiment pages. 68 out of 27,587 pages (0.25%) show a duration lasting 10 minutes or more. In this case, participants were likely taking a break on pages where we did not impose a time limit. We capped the time spent on a page at 10 minutes in this case.

[17] 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).

students (55 percent) and 233 male students (45 percent), and 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.[18] For simplicity, we take the midpoint of each interval as the risk aversion coefficient for each corresponding lottery choice.[19] Consistent with findings in prior literature, Panel A of Table 5 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 (though beliefs are elicited on a 0-100 scale, 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 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 she answered all these questions correctly, the value of her under-confidence measure would be close to 0. In contrast, if the participant has no confidence in her correct answers and assigns the probability of being correct close to 0, the under-confidence value would be close to 1. Table 5 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.

[18] See Table 1 of Eckel and Grossman (2008) for the CRRA ranges.

[19] 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.

4.3.3 Over-optimism and uncertainty in beliefs

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 5 shows that both genders are, on average, equally over-optimistic: female (male) students believed they answered 2.37 (2.27) more questions correctly than their true score.

In addition, the experiment elicited the probability that the participant believes the guessed outcome would occur (prior belief). If the participant is fairly certain about their guess, they would assign a probability close to 1 (that is, 100%).[20] We capture the uncertainty in beliefs as the gap between 1 and the assigned probability to the predicted outcome. Hence, the measure takes values between 0 and 1, with a value closer to 1 indicative of participants being very uncertain about her prediction. Table 5 shows that, on average, female participants are more uncertain (0.51) than their male counterparts (0.47), and the gender difference is significant at the 10-percent level (Table 5).

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 5 shows the average values of these traits by gender. We see that females are significantly more agreeable, neurotic, and open. For example, the gender difference in neuroticism is 0.58, almost 56% of the standard deviation (1.04) in this measure. These personality traits 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).

4.4 Analysis

Panel C of Table 5 shows that male students performed better than female students on the quiz (an average of 12.98 correct answers for males versus 11.57 for females). Turning to the willingness to ask for regrades, Figure 4 shows interesting gender differences. When the

[20] We also consider an alternative uncertainty measure using the variance of their expectations based on the subjective probability distributions reported by the participants, and the results are insensitive to the choice of uncertainty measures.

cost is zero or negative (i.e., participants are getting paid for requesting regrades), there are no discernible gender differences in their 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 female 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 willingness-to-pay (WTP). Table 5 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.

We next investigate the drivers of gender differences in the WTP. Column 1 of Table 6 regresses the indicator of willing to pay a positive cost for regrades respondent gender. Men are 10.3 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 6). Columns 3 and 4 show that over-confidence and under-confidence explain about 7.8 percent and 18.4 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 6 shows that over-optimism is not a determinant and cannot explain any of the gender difference in the willingness to pay a positive cost for regrades. Column 6 shows that the more uncertain the participant is about the predicted outcome, the less likely it is that she requests a regrade. Uncertainty alone explains 13.6 percent of the gender gap in willingness to pay for regrades. Column 7 shows that all the aforementioned covariates combined explain about 20 percent of the gender gap in willingness to pay for regrades.

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 percent although the coefficients on the personality traits are imprecise and insignificant. Column 9 shows that under-confidence, uncertainty, and the Big Five personality traits combined explain 33 percent of the gender gap and reduce the gender gap to 6.7 percent (statistically insignificant). Finally, the last column shows that the including all other variables in addition to under-confidence, uncertainty, and the Big Five personality traits does not further reduce the unexplained gender gap. Appendix Table C1 shows similar patterns when we use the WTP as the dependent

variable instead.

Although men are more likely than women to ask for regrades, the regrades do not necessarily make men better off. By the asymmetric experiment design, participants are more likely to see their grade improved than deteriorated after requesting a regrade. Because male students are more likely than female students to request regrades at a given cost, Panel A of Figure 5 shows that the regrade outcomes are favorable to men especially when the cost is positive.[21]

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 C of Figure 5 shows that indeed male students become better off financially when the regrading cost is positive but low (i.e., \$0.50). However, panel D 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. In summary, 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.

A natural question to ask is whether our experimental measure of WTP to ask correlates with actual regrade requests by students. Figure 6 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. For example, amongst those with a non-positive WTP, 56% of students reported ever considering making a regrade request, versus 61% of students with a positive WTP.

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.

5 Discussions and Conclusion

We find persistent gender differences in regrades in all three different sources of data. The magnitudes of gender differences (18.5 percent in transcript data, 16-24 percent in the student survey, and 28.5 percent in the laboratory experiment) are also comparable across these

[21] When the cost is positive (i.e., the participant has to pay a cost to request a regrade), except for the cost at \$1.50 and \$2.50, males are statistically significantly more likely than females to receive an improved grade as a result of regrade requests.

data sources. We find that class and student characteristics hardly explain the favorable regrade outcomes male students enjoyed in the transcript records. Instructor and student surveys show that regrade/grade change requests are prevalent in college. Forty percent of students reported that they approached instructors for grade changes at some point during their college life. These regrade experiences include grade corrections during the semester and changes to the final grade that may occur before the instructors officially submitted the final grades. Conditional on asking, over 30 percent of these requests result in better grade outcomes for students. Therefore, these regrade requests may have profound impact on grades for students. The survey results from instructors and students provide evidence that male students are disproportionally more likely than female students to make regrade requests on the intensive margin which result in favorable outcomes for men.

To shed light on the potential mechanisms, we rely on surveys and a laboratory experiment. The evidence points to the gender differences in regrades being primarily driven by gender differences in the propensity to ask. We find that gender differences in risk preferences do not explain the gender differences in asking. We also find that male and female participants are similarly overconfident and over-optimistic and hence these two factors only explain a small fraction (below 4 percent) of the gender gap in asking. In contrast, we find that under-confidence and uncertainty in subjective beliefs explain 13.6 – 18.4 percent of the gender differences in asking. Interestingly, we find that personality traits also explain a large fraction (19.4 percent) of the gender differences in asking. All of the above measures combined account for about a third of the gender differences in asking, with two-thirds of the gap unexplained.

Conversely, we find that subjective probability beliefs in possible outcomes and risk aversion explain little of the gender differences in regrade consideration using the student survey data. In the student survey, we asked students how how stressed they felt when/if they asked instructors to reconsider their grade on a 1-7 scale (1: not stressed at all; 7: extremely stressed). Table 4 shows that, on average, female students indicate a higher stress level than do males (5.01 versus 3.88). We find that stress explains nearly half of the gender difference in the number of classes students consider for regrades. One's stress level is negatively correlated to the control one perceives she has over the direction of her life and positively correlated to the measure of neurotism (one of the Big Five personality traits). These gender differences in personality traits and psychic costs seem to be a factor in the gender differences in WTP to ask for regrades.

Although fear of backlash and the dynamics of gender interactions may play some roles in the gender differences in regrades in real life, we purposely eliminate these factors in the laboratory setting. Without relying on these two explanations, our experimental evidence

suggests that substantial gender differences in willingness to pay for regrades persist. Although, in the transcript data, we observe that the male advantage in regrade outcomes is most salient when male students are in male instructors' classes (compared with female students in female instructors' classes, female instructors are 0.044 percentage points more likely to change male students' grade upward and male instructors are 0.075 percentage points more likely to do so), our experiment design shuts down this channel. Future work that sheds light on this would be valuable.

Another potential explanation for gender differences in asking is the ambiguity in appropriateness of asking. Recall that 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. In our experiment, subjects are explicitly asked to make the regrade decisions at each cost level. Similar to the finding in Leibbrandt and List (2014), we find no gender differences in regrade requests when participants bear zero cost of asking. However, the gender differences start to surface when we introduce a small cost of asking. Female reluctance in paying a cost appears to be the primary driving force of the gender differences in asking.

Our experiment shows that males have a higher propensity to ask, both in cases where it makes economic sense to do so and where it does not. Compared to the data in the instructor and student surveys, instructors reported that male students are more aggressive than female students to ask for regrades (see Table 3), and male students also indicate stronger regrets than female students for asking for regrades too aggressively (Table 4). Thus, a coherent picture emerges: men are simply more willing to ask, regardless of whether it makes sense to do so. When the regrade cost is small, it creates a substantial advantage for male students because they are more likely to ask for regrades. However, such male advantage diminishes quickly as the cost of regrades increases. In real life, because the regrade costs are very small for students, it creates an advantage for male students. If employers rely on college transcripts as a screening mechanism, this gender gap may contribute to the unfavorable treatment to women in the labor market.

One policy implication of our results is clear: given the aversion of females to ask when costs of doing so are positive, the cost of asking needs to be lowered. Policy aiming at reducing under-confidence and uncertainty in beliefs for females would also help mitigate the adverse outcomes for females resulted by the gender differences in asking. For instance, in the educational context, when a regrade is granted to a student, the instructor may inform the class about the grade correction that other students may also be eligible for. In the labor market context, when a promotion opportunity is available, the employers may actively reach out to eligible female workers to encourage their applications.

References

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

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

Babcock, L. and S. Laschever (2009). *Women don't ask: Negotiation and the gender divide*. Princeton University Press.

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.

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.

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.

Croson, R. and U. Gneezy (2009). Gender differences in preferences. *Journal of Economic literature 47*(2), 448–74.

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., M. Niederle, and L. Vesterlund (2016). Knowing when to ask: The cost of leaning in. Technical report, National Bureau of Economic Research.

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.

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.

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.

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.

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*.

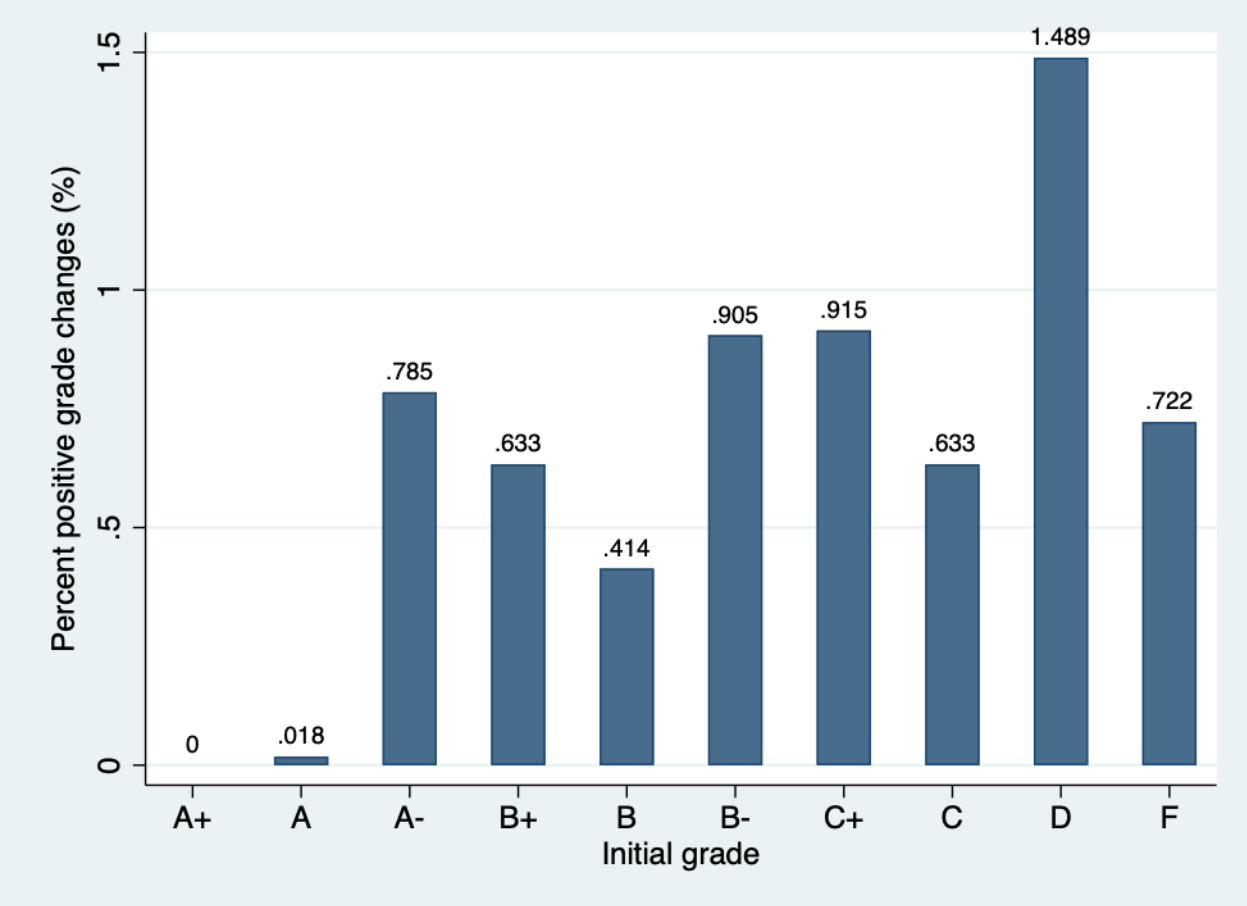

Figure 1: Percent upward grade changes conditional on initial grades

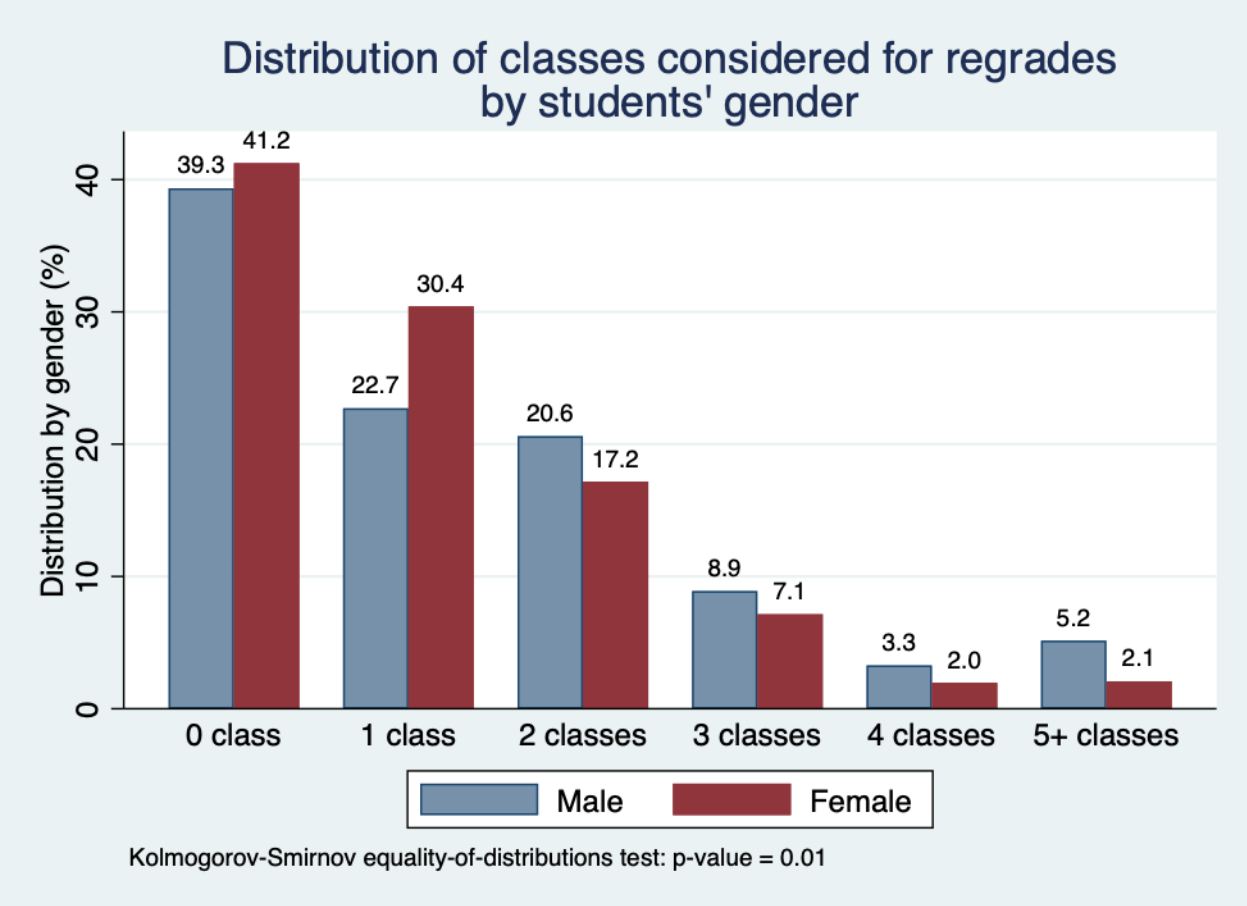

Figure 2: Student survey: number of classes students considered for regrade requests by students' gender

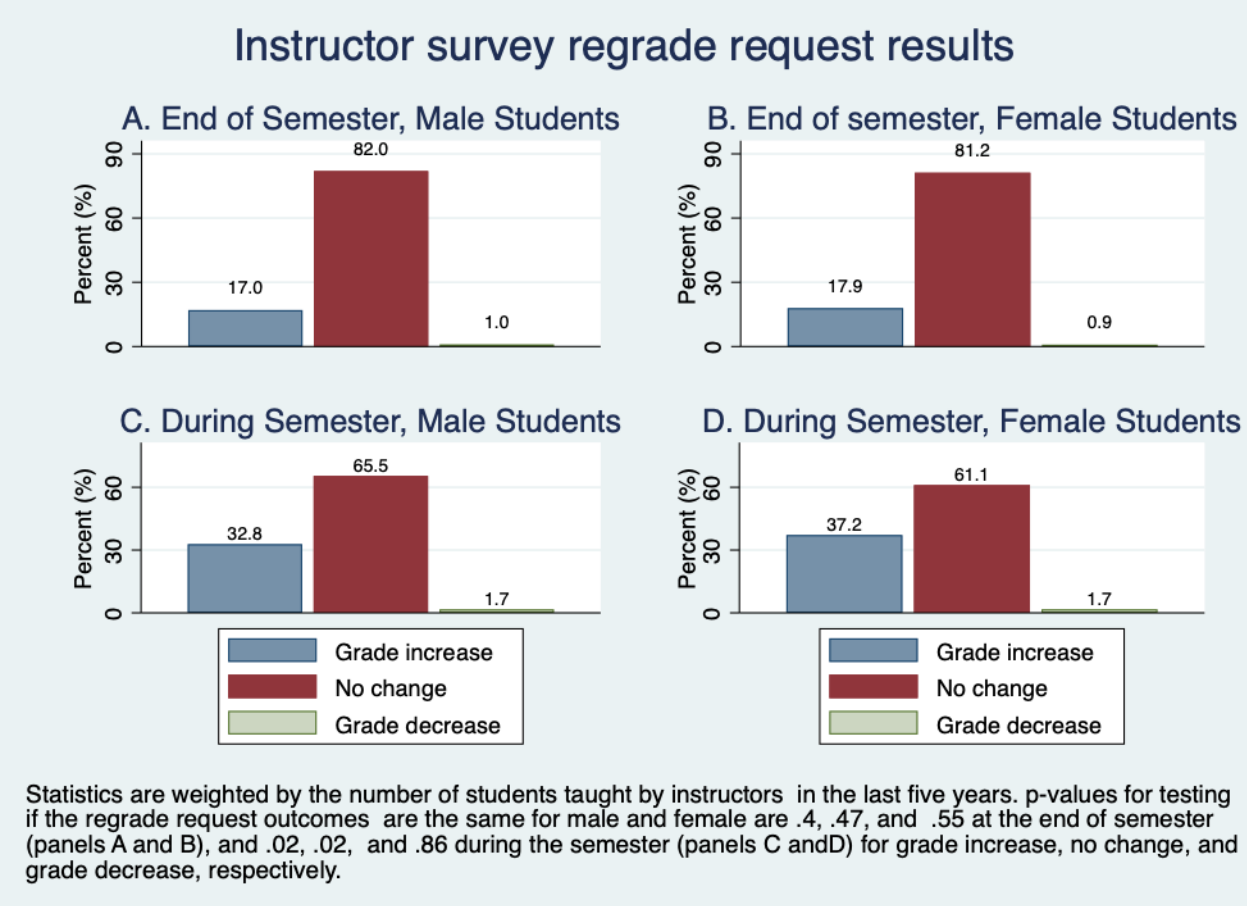

Figure 3: Instructor survey: regrade results by timing and students' gender

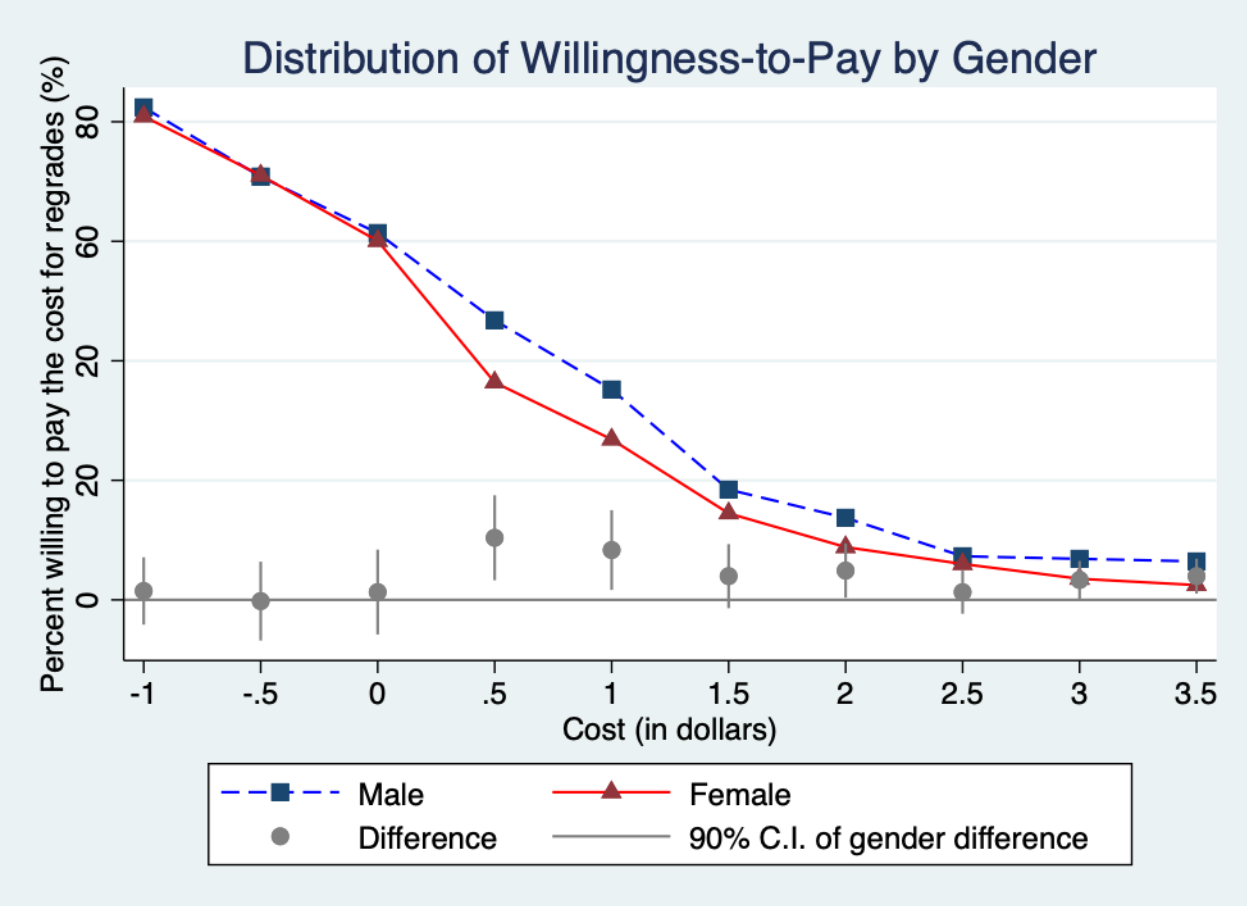

Figure 4: Fraction requesting regrades at varying costs by sstudents' gender

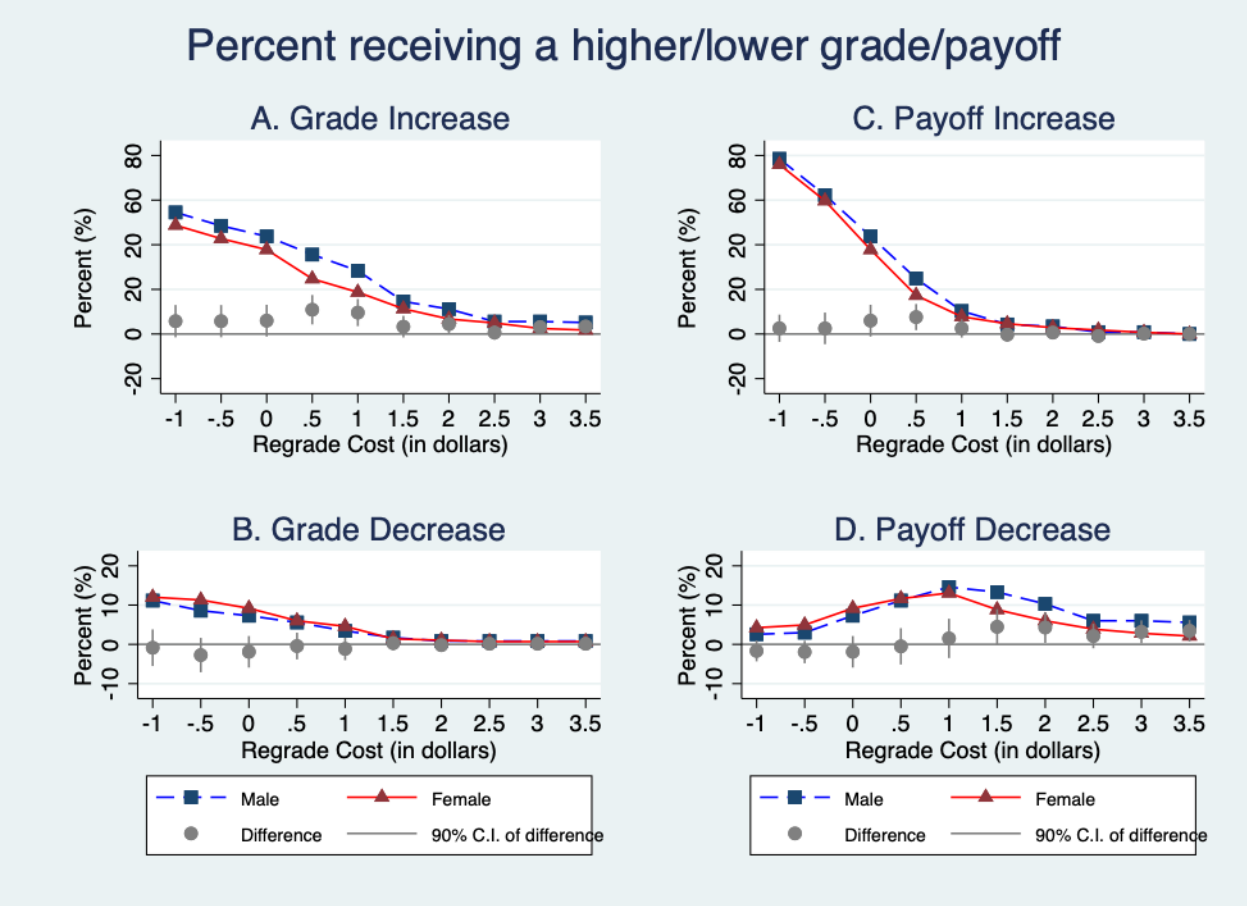

Figure 5: Fraction of grade/payoff increase/decrease at varying regrade costs by sstudents' gender

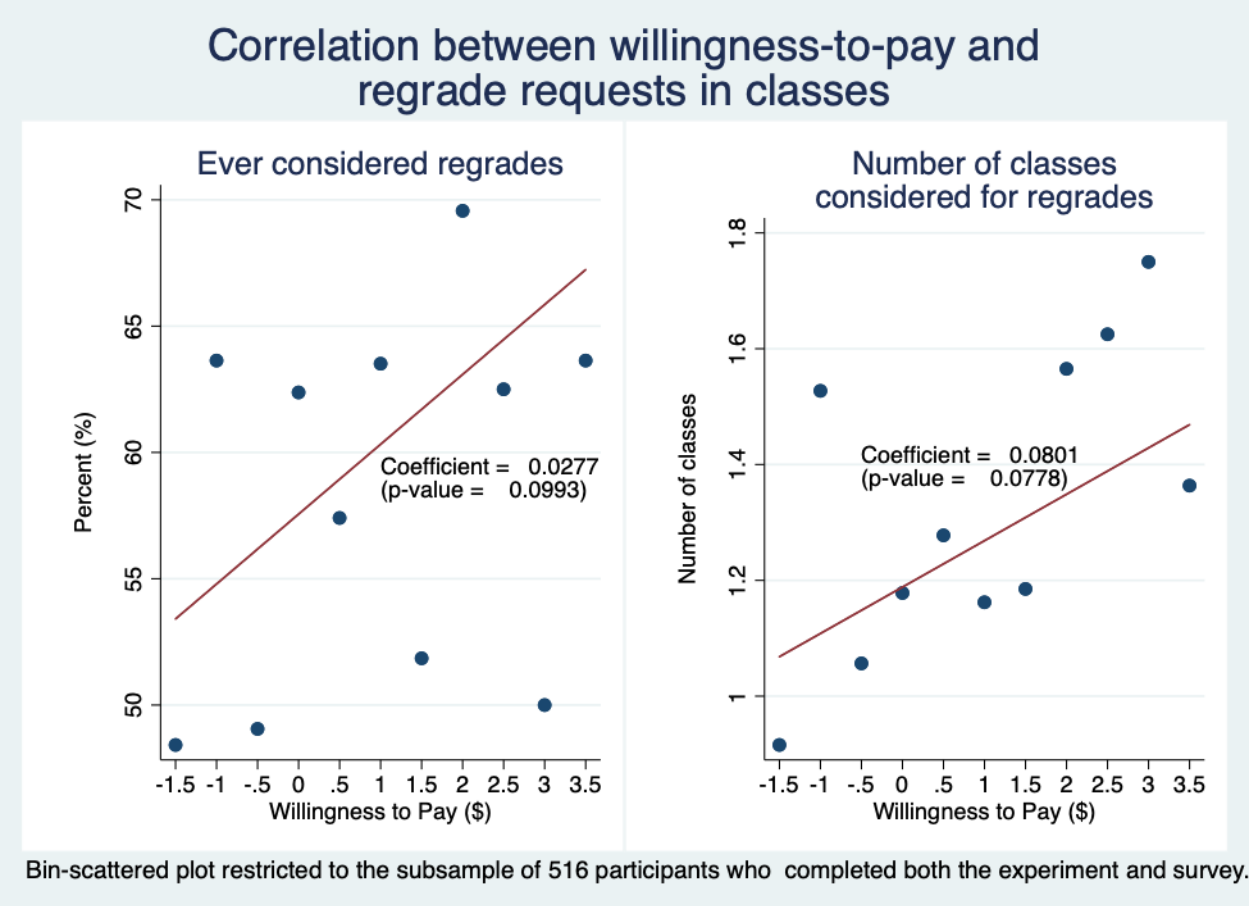

Figure 6: Correlation between WTP and regrade experience in class

Table 1: Summary statistics of transcript data by students' gender

	Female students	Male students	Difference
Percent received a grade change	0.424	0.509	-0.085***
	(6.501)	(7.119)	(0.012)
Percent received a positive grade change	0.404	0.479	-0.075***
	(6.342)	(6.902)	(0.011)
Percent received a negative grade change	0.021	0.031	-0.010***
	(1.432)	(1.753)	(0.003)
Percent took class from female instructors	50.0	38.9	11.1***
	(50.0)	(48.8)	(0.088)
Average term GPA	3.15	2.95	0.201***
	(0.723)	(0.786)	(0.001)
Percent in College of Agriculture	4.87	4.00	0.872***
	(21.5)	(19.6)	(0.036)
Percent in College of Business	8.63	13.2	-4.58***
	(28.1)	(33.9)	(0.053)
Percent in College of Engineering	1.86	8.21	-6.35***
	(13.5)	(27.5)	(0.037)
Percent in College of Human Services	15.5	9.13	6.40***
	(36.2)	(28.8)	(0.057)
Percent in College of Liberal Arts	33.6	30.6	3.07***
	(47.2)	(46.1)	(0.081)
Percent in College of Natural Resources	3.53	5.85	-2.32***
	(18.5)	(23.5)	(0.036)
Percent in College of Natural Sciences	23.2	23.4	-0.251***
	(42.2)	(42.3)	(0.073)
Percent in College of Veterinary Sciences	4.40	2.32	2.07***
	(20.5)	(15.1)	(0.031)
Percent in Intra-University	4.38	3.30	1.07***
	(20.5)	(17.9)	(0.033)
Observations	716,772	624,780	

$*p < 0.10, **p < 0.05, ***p < 0.01$. Standard errors in parentheses.
a. 85,250 transcript records have missing values for instructors' gender. The sample size is 671,276 and 585,026 for female and male student records, respectively.

Table 2: Probit regression analysis of transcript data

			Probit Regression		
	[1]	[2]	[3]	[4]	[5]
	A. Dependent variable: Upward grade change $\in \{0, 1\}$				
Male student	0.000748***	0.000705***	0.000675***	0.000742***	0.000808***
	(0.000118)	(0.000124)	(0.000123)	(0.000118)	(0.000121)
Observations	1,341,552	1,341,160	1,256,302	1,341,552	1,294,820
	B. Dependent variable: Downward grade change $\in \{0, 1\}$				
Male student	0.000102***	0.000029	0.000102***	0.000098***	0.000092***
	(0.000029)	(0.000034)	(0.000030)	(0.000028)	(0.000029)
Observations	1,341,552	1,120,274	1,256,302	1,341,552	1,294,073
Control for Panels A and B	None	Colleges Departments	instructors' gender and rank	Student class standing	GPA and initial grade
			Sensitivity Analysis		
	Baseline	Excl. students w/2+ changes	Excl. instructors w/10%+ changes	Excl. A+ and A	Excl. F
	C. Dependent variable: Upward grade changes $\in \{0, 1\}^a$				
Male student	0.000780***	0.000642***	0.000651***	0.001184***	0.000830***
	(0.000132)	(0.000116)	(0.000126)	(0.000195)	(0.000133)
Observations	1,212,471	1,200,284	1,209,127	819,533	1,168,199
	D. Dependent variable: Downward grade changes $\in \{0, 1\}^a$				
Male student	-0.000006	-0.000020	-0.000018	0.000046	-0.000006
	(0.000038)	(0.000038)	(0.000032)	(0.000052)	(0.000038)
Observations	990,617	967,413	988,165	560,603	990,617

$*p < 0.1; **p < 0.05; ***p < 0.01$. Marginal effects are reported in the table. Standard errors are clustered at the student level and reported in parentheses. *Male* is an indicator for male students.

a. Panels C and D control for college, department, instructors' gender and ranking, student's class standing, student's term GPA, and student's initial grade.

Table 3: Summary statistics of the instructor survey

	End of semester	During semester	Difference
A. Regrade requests			
Percent requested regrades (%)	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			
Male 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)
Female more aggressive (%)	11.05	13.90	2.85
	(31.48)	(34.74)	(2.95)
C.Results of regrade requests by students' sex			
Grade increased, female students (%)	17.94	37.20	19.26***
	(26.38)	(29.50)	(3.52)
Grade unchanged, female students (%)	81.19	61.07	-20.12***
	(26.72)	(29.88)	(3.70)
Grade decreased, female students (%)	0.87	1.73	0.86
	(4.61)	(6.42)	(0.65)
Grade increased, male students (%)	17.01	32.83vv	15.82***
	(25.17)	(28.40)	(2.73)
Grade unchanged, male students (%)	81.98	65.46vv	-16.53***
	(25.49)	(28.91)	(3.00)
Grade decreased, male students (%)	1.00	1.71vv	0.71
	(4.97)	(6.31)	(0.70)
D.Instructor overall impression on success of regrade requests			
Male students more successful (%)	2.27		
Males and females similarly successful (%)	75.95		
Female students more successful (%)	21.78		

* p<0.10; ** p<0.05, *** p<¡0.01 indicate the significance level of 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 in the class.
v p<0.10; vv p<0.05, vvv p<0.01 indicate the significance level of gender differences at the end of/during the semester.
Standard errors in parentheses. All statistics are weighted by class size.

Table 4: Summary statistics of the student survey

A. All students	Female	Male	Difference
Fraction ever considered regrades	0.59	0.61	-0.02
	(0.49)	(0.49)	(0.03)
Fraction ever asked for regrades	0.39	0.41	-0.02
	(0.49)	(0.49)	(0.03)
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	0.59	0.73	-0.14**
	(0.92)	(1.12)	(0.06)
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 aggressive, 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)
Risk aversion coefficient	1.16	0.88	0.27***
	(0.77)	(0.74)	(0.04)
Big 5 trait: Extroversion[d]	3.14	3.06	0.08
	(0.98)	(0.97)	(0.06)
Big 5 trait: Agreeableness[d]	3.66	3.51	0.15***
	(0.82)	(0.84)	(0.05)
Big 5 trait: Conscientiousness[d]	3.82	3.66	0.16***
	(0.75)	(0.79)	(0.04)
Big 5 trait: Neuroticism[d]	3.43	2.85	0.58***
	(0.96)	(0.90)	(0.06)
Big 5 trait: Openness[d]	3.55	3.49	0.06
	(0.88)	(0.81)	(0.05)
Observations	868	427	
B. Sub-sample of students who asked for regrades	Female	Male	Difference
Percent grade component increased (%)	61.04	66.88	-5.84
	(48.85)	(47.22)	(4.77)
Percent grade component unchanged (%)	37.34	33.12	4.22
	(48.45)	(47.22)	(4.74)
Percent grade component decreased (%)	1.62	0.00	1.62
	(12.66)	(0.00)	(1.02)
Percent final grade increased (%)	30.52	33.77	-3.25
	(46.12)	(47.45)	(4.60)
Percent final grade unchanged (%)	66.23	59.74	6.49
	(47.37)	(49.20)	(4.74)
Percent final grade decreased (%)	3.25	6.49	-3.25
	(17.75)	(24.72)	(2.01)
Observations	338	175	

$*p < 0.10; **p < 0.05, ***p < 0.01$ indicate the significance level of gender differences. Standard errors in parentheses. a. Scale ranges between 1 (never true to me) to 7 (always true to me). b. Scale ranges between 1 (I don't have any control) to 7 (I have full control). c. Scale ranges between 1 (not stressed at all) to 7 (extremely stressed). d. Scale ranges between 1 (lowest) and 5 (highest).

Table 5: 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)
B.	Big 5 trait: Extroversion[e]	3.03	2.96	0.07
		(0.97)	(0.94)	(0.08)
	Big 5 trait: Agreeableness[e]	3.62	3.48	0.13*
		(0.85)	(0.82)	(0.07)
	Big 5 trait: Conscientiousness[e]	3.74	3.64	0.10
		(0.77)	(0.80)	(0.07)
	Big 5 trait: Neuroticism[e]	3.52	2.89	0.63***
		(1.03)	(0.94)	(0.09)
	Big 5 trait: Openness[e]	3.71	3.47	0.24***
		(0.90)	(0.83)	(0.08)
C.	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)
	True score of the quiz	11.57	12.98	-1.41***
		(3.19)	(2.87)	(0.27)
	Percent asking for regrades if paying $3.50 (%)	2.47	6.44	-3.96**
		(15.56)	(24.60)	(1.78)
	Percent asking for regrades if paying $3.00 (%)	3.53	6.87	-3.33*
		(18.50)	(25.34)	(1.93)
	Percent asking for regrades if paying $2.50 (%)	6.01	7.30	-1.29
		(23.80)	(26.06)	(2.20)
	Percent asking for regrades if paying $2.00 (%)	8.83	13.73	-4.90*
		(28.43)	(34.49)	(2.77)
	Percent asking for regrades if paying $1.50 (%)	14.49	18.45	-3.97
		(35.26)	(38.88)	(3.27)
	Percent asking for regrades if paying $1.00 (%)	26.86	35.19	-8.34**
		(44.40)	(47.86)	(4.07)
	Percent asking for regrades if paying $0.50 (%)	36.40	46.78	-10.39**
		(48.20)	(50.00)	(4.34)
	Percent asking for regrades if paying $0.00 (%)	60.07	61.37	-1.30
		(49.06)	(48.79)	(4.33)
	Percent asking for regrades if getting paid $0.50 (%)	71.02	70.82	0.21
		(45.45)	(45.56)	(4.02)
	Percent asking for regrades if getting paid $1.00 (%)	80.92	82.40	-1.48
		(39.36)	(38.16)	(3.43)
	Percent willing to pay a positive cost for regrades (%)	36.40	46.78	-10.39**
		(48.20)	(50.00)	(4.34)
	WTP ($)	0.05	0.25	-0.19*
		(1.22)	(1.38)	(0.11)
	Observations	283	233	

* $p < 0.10$; ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses. a. *Over-confidence* measures participants' average absolute gap (between 0 and 1) between probability assignments and 0 for wrong answers in the quiz. b. *Under-confidence* measures participants average absolute gap (between 0 and 1) between probability assignments and 1 for correct answers in the quiz. 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. Scale ranges between 1 (lowest) and 5 (highest).

Table 6: Porbit 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]	[10]
Male student	0.103**	0.099**	0.095**	0.084**	0.105**	0.089**	0.080*	0.083*	0.067	0.068
	(0.042)	(0.043)	(0.042)	(0.043)	(0.042)	(0.042)	(0.043)	(0.046)	(0.045)	(0.046)
Risk aversion[a]		-0.018					-0.015			-0.015
		(0.029)					(0.029)			(0.029)
Over-confidence[b]			0.344***				0.233*			0.253*
			(0.113)				(0.133)			(0.134)
Under-confidence[c]				-0.438***			-0.090		-0.272*	-0.080
				(0.146)			(0.178)		(0.156)	(0.179)
Over-optimism[d]					-0.008		-0.005			-0.005
					(0.007)		(0.007)			(0.007)
Uncertainty[e]						-0.390***	-0.332***		-0.323***	-0.326***
						(0.094)	(0.100)		(0.101)	(0.101)
Extroversion[f]								-0.008	-0.002	-0.004
								(0.024)	(0.024)	(0.024)
Agreeableness[f]								-0.026	-0.024	-0.030
								(0.026)	(0.026)	(0.026)
Conscientiousness[f]								-0.007	-0.013	-0.011
								(0.028)	(0.028)	(0.028)
Neuroticism[f]								-0.020	-0.009	-0.008
								(0.023)	(0.023)	(0.023)
Openness[f]								-0.014	-0.012	-0.015
								(0.025)	(0.024)	(0.024)
Observations	516	516	515	516	516	516	515	516	516	515
F-test (p-value): all coefficients (other than Male student and Constant) = 0							.0005	.833	.006	.008

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. 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' average absolute gap (between 0 and 1) between probability assignments and 0 for wrong answers in the quiz. c. *Under-confidence* measures participants average absolute gap (between 0 and 1) between probability assignments and 1 for correct answers in the quiz. 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. *Conscientiousness*, *Neuroticism*, *Openness* are the Big Five personal traits with the value ranging between 1 and 5.