Student Demand For Relative Performance Feedback: Evidence from a Field Experiment

Alessandro Castagnetti (LIUC University)*

Derek Rury (University of Chicago)†

This version: February 1, 2024

Abstract

We argue that performance feedback carries two distinct values; an instrumental value and an intrinsic value. While the instrumental value helps with future decisions about human capital, the intrinsic value can impact a students' utility directly. To study these concepts, we administer a survey in an introductory economics class in which we elicit students' willingness to pay for or avoid learning their rank on a midterm exam. Our results show that 10% of students are willing to pay to avoid learning their rank. Upon learning their rank, students report needing more study hours to achieve their desired grade and being less likely to be in the top half of the ability distribution in the class. These effects depend on students' ex-ante beliefs about their performance in an intuitive way. We do not find an overall effect of learning about rank performance on final course grade, although for students who received "good news", we see a negative effect. We also confirm that students' preferences for feedback do not interfere with their belief updating when receiving negative news.

Key Words: Rank, Education, Beliefs, Performance, Human Capital
JEL Codes: I26, I29, D83.

*Address: LIUC University, email: scastagnetti@liuc.it

†Address: Harris School of Public Policy, University of Chicago, Keller Center, email: rury@uchicago.edu

The AEA registry for this field experiment can be found using the number AEARCTR-0006457 under the title "Motivated Beliefs in Higher Education".

1 Introduction

Students frequently receive performance feedback (e.g. exam grades) as part of their education. This feedback forms the basis for how students learn about their ability and make decisions regarding their academic future. Importantly, this performance feedback carries an *instrumental value*, or a feature of information that helps students make decisions about their human capital under uncertainty. As an example, learning about their performance in Econ 101 will help students' learn whether taking Econ 102 is utility-maximizing. Because of this feature, we expect students to have a non-negative demand for the instrumental value of performance feedback.

In this paper, we argue that performance feedback can also have an *intrinsic value*, or a feature of information that impacts an individual's psychology, or utility, directly. Previous work shows that the intrinsic value of performance feedback may carry either a positive or negative effect on demand for feedback as individuals may prefer not to receive information that might hurt their ego (Bénabou and Tirole, 2002; Eil and Rao, 2011; Möbius et al., 2022). When considering both the instrumental and the intrinsic values, we do not know whether students' demand for performance feedback information is either positive or negative as they may have opposing effects on demand. Moreover, it is unclear whether students' desires to avoid processing negative information about themselves (i.e., self-deception) can override the useful instrumental features of performance feedback, ultimately affecting their beliefs about their own ability, classroom performance and future take-up of educational opportunities.

To address this, we administered a survey at a public university in the US designed to elicit students' preferences for information about their relative exam performance. Students were asked to report their willingness to *pay for* (WTP) and their willingness to *avoid* (WTA) learning their relative rank on their first midterm. We argue that learning *relative* performance feedback is informative as grades are often based explicitly on comparisons

between classmates (i.e., curved). It also gives students objective information about how they compare to others. We find that students are willing to pay around $1.00 on average to learn their rank, primarily driven by nearly 43% of students, who have a strictly positive willingness to pay to know their actual rank on the exam. However, 10% of individuals have a negative willingness to pay, implying they are willing to pay to *avoid learning* their rank. We characterize these students as suffering from *information avoidance*. Because of the potential opposing effects on demand for information (instrumental = positive, intrinsic = positive/negative), we view this ten percent number as an underestimate of the proportion of students with negative *intrinsic* value of information for performance feedback.

One contribution of our paper is that we show our WTP/WTA measures are not related to either performance or beliefs about performance on the midterm. This contradicts other work studying individuals' preferences for information about themselves.[1] We also find that female students are willing to pay an additional $1 for information about their rank ($p<0.05$). This difference in WTP/WTA cannot be explained by performance on the exam or students' beliefs (and any other relevant control), implying that there are gender differences in preferences for feedback. This result represents a novel finding in the literature of demand for information about performance as previous work finds that there are either no differences in demand for information by gender (Eil and Rao, 2011) or that women are more likely to avoid information (Sharma and Castagnetti, 2023).[2]

To study the effect of performance feedback on beliefs and achievement, we also conducted an experiment where we randomly shared students' actual performance rank on the exam.

[1] Although our results are not consistent with some theoretical and experimental literature (e.g., Burks et al. (2013); Eil and Rao (2011); Köszegi (2006); Möbius et al. (2022)), they are consistent with findings in other studies (e.g., Sharma and Castagnetti (2023)) and might be driven also by the specific educational setting we are analyzing.

[2] This finding shows that women are more willing to receive relative performance feedback which may help counteract the notion that they tend to hold low beliefs about their ability or performance in different tasks (Bordalo et al., 2019; Coffman, 2014; Coffman et al., 2021). These beliefs can ultimately have adverse impact on gender gaps in self-promotion (Exley and Kessler, 2022), contribution of ideas (Coffman, 2014), and job applications (Coffman et al., 2020).

We find that students report requiring more hours of study time and believing they are less likely to be above average in the class as a result of learning their rank. These findings demonstrate the *instrumental* value of performance feedback as students update their beliefs in response to new information. When studying how our treatment impacted performance in the class, we fail to detect any significant effects on exam scores for the sample overall.

A further contribution of this paper is that we collect prior beliefs about performance rank. Previous work studying how students respond to performance feedback consists of papers studying natural experiments, which typically do not collect information on students' beliefs. Looking at beliefs about performance, we find that students are typically inaccurate when guessing their rank. The standard deviation between their beliefs and actual rank is nearly 19 rank points (in terms of midterm scores). When studying whether our treatment effects depend on whether students received "good" or "bad" news (Bobba and Frisancho, Forthcoming), we find intuitive treatment effects on beliefs: students who received bad news reported needing more study hours to achieve their desired grade and that they were less likely to be above average ability-wise in their class. Along those same lines, those who received good news reported being more likely to be above average in the class, further demonstrating the instrumental value of performance feedback. When looking at effects on overall learning, we find a negative treatment effect for students who received good news regarding their rank on their first midterm.

Lastly, we show that students' beliefs about performance and their preference for information do not interfere with how they process information about their human capital. Previous work has found that agents often exhibit self-deception and asymmetric memory when recalling unflattering information (Bénabou and Tirole, 2002; Zimmerman, 2020; Chew et al., 2020). When studying whether students' WTP/WTA for information influences our results, we fail to find any evidence that students' preferences impact our treatment effects. Moreover, contrary to results found in Zimmerman (2020), preferences for information do

not influence whether students remember their rank one month later, ruling out any dynamic effects of memory influencing beliefs (Coffman et al., 2021).

This work contributes to three different strands of literature. First, it contributes to the theoretical and empirical literature on motivated beliefs and the demand for information. Theoretical work shows that ego-utility and motivated beliefs may lead individuals more broadly to avoid information (e.g., Bénabou and Tirole (2002); Brunnermeier and Parker (2005); Köszegi (2006)).[3] This is supported by experimental findings which show that individuals avoid ego-relevant performance feedback (Castagnetti and Schmacker, 2022; Eil and Rao, 2011; Möbius et al., 2022). Empirical studies on information avoidance so far have studied health (Oster et al., 2013; Ganguly and Tasoff, 2016) and financial settings (Karlsson et al., 2009; Sicherman et al., 2015). We add to this literature by being the first paper to study demand for information in an education setting, where the intrinsic value of information is likely to be high.

This paper also contributes to the experimental literature studying how individuals process information that carries an ego-relevant dimension (i.e., information about own ability).[4] Different mechanisms of biased information processing have been tested in the lab. These include: asymmetric updating (Castagnetti and Schmacker, 2022; Coutts et al., 2020; Eil and Rao, 2011; Möbius et al., 2022), selective recall (Chew et al., 2020; Zimmerman, 2020), and motivated errors (Exley and Kessler, 2019). These papers find that individuals process ego-relevant information self-servingly. That is, subjects are more likely to update more strongly to positive than negative signals about their ability, they are more likely to remember positive than negative performance feedback, and they are more likely to make mistakes to reach higher beliefs about themselves. To this literature we contribute the analysis of ego-relevant information processing in terms of asymmetric updating and selective recall in

[3] For a review, see Goldman et al. (2017).

[4] This literature is grounded on theoretical work that underlines the ways in which ego motives can affect information processing (Bénabou and Tirole, 2002, 2016).

an education context, in which the stakes are admittedly higher than those found in the lab.

Lastly, this paper adds to the literature on performance feedback. Several studies have analyzed how achievement influences learning about one's overall performance as well as performance relative to a group of peers (Li, 2018; Dobrescu et al., 2021). Most studies find positive effects on achievement for all students (Azmat and Iriberri, 2010; Bandiera et al., 2015; Brade et al., 2018; Dobrescu et al., 2021) although one study in a college setting finds that low-performing students perform worse as a result of feedback (Azmat et al., 2019).[5] Research has also studied how students respond to learning about the performance of different cohorts (Owen, 2022; Rury, 2022). In a paper similar to our own, Li (2018) provides information to female students about career prospects and where they rank within their economics classes. Unfortunately, the author does not provide this information separately, so it remains unclear what role performance feedback plays on behavior. A contribution of our paper is that we can isolate the role of performance feedback and measure the effect of our information on important beliefs about students' human capital decisions.

The rest of the paper is structured as follows: section 2 outlines the experiment; section 3 describes the data and presents descriptive results; section 4 discusses our willingness to pay/avoid measure and results; section 5 shows the experimental results; section 6 discusses our results and concludes.

2 Design of Experiment

2.1 The Setting

This study was conducted at a selective public research university in central California during the fall 2020 and winter 2021 quarters. Six classes participated in this experiment, including

[5]Another paper looking at high school students in Mexico finds an asymmetric response where high performing students choose more rigorous academic tracts and lower performing students choose less rigorous tracts after receiving feedback (Bobba and Frisancho, Forthcoming).

four in the fall and two in the winter. All classes in the experiment were introductory economics courses; this serves as an important part of the experiment as students enrolled in introductory courses are least likely to know their academic ability in economics. The majority of points used to decide the final grade were determined by performance on the two course midterms and the final exam.[6]

Within 24 hours of taking their first midterm exam, students in each class were invited to take a survey about their experience in the class. Students learned about the survey through a combination of professor advertisements as well as notifications through the course website (i.e., Canvas). The survey was shared with every student in each class through a message on Canvas. Importantly, information about what the survey was about was not disclosed in order to minimize selection bias in completing the questionnaire.[7] The survey was coded using the software Qualtrics and took about 10 minutes to complete.[8] As part of the survey, students were asked to release their academic and demographic information from the university's registrar. All students who began the survey completed this step.

2.2 Prior Beliefs and Demand for Information

One main motivation of our survey was to elicit students' beliefs about performance and their willingness to pay to learn their rank on their first midterm. To study student beliefs on performance, we asked students to think of their class midterm performance as being a percentile among all other students in the class. This ordering puts scores within a distribution between 1 and 100, placing those who scored higher on the midterm lower in the

[6] All six classes assigned more than 60% of course points to exams, while one class assigned 100% of points to exams. The remaining points for the other classes consisted of a combination of homework, lecture videos and attendance or attention credits. One class determined grades via a single midterm and a final exam.

[7] Information that revealed statistics about the distribution of test scores (other than the raw score for each individual) were deactivated as part of participating in this survey. Furthermore, students at this university typically do not have much information above learning the variance or standard deviation of scores and listing a student's rank is not done in most if not all courses.

[8] Appendix A provides the experimental instructions.

ranking (e.g., highest scoring student would be at the one percent level).[9] This distribution was then broken into ten deciles, ranging between 0-10 and 90-100. To elicit beliefs about performance, students were then asked to place a probability distribution over each performance decile, representing the likelihood of their performance being in each decile. This gives us a prior distribution of beliefs about midterm performance.

Participants then moved on to our willingness to pay (WTP) or willingness to avoid (WTA) elicitation procedure. In this procedure, students were given a series of 21 choices where each choice point represented a decision to choose between paying to learn (not learn) their rank and not paying to learn (not learn) their rank. These choices ranged between willing to *receive* \$10 to learn their own rank and willing to pay \$10 to learn their rank. In the procedure, which was represented as a choice between two columns, all of the options that required students to *pay* (either WTP/WTA) were in the left column, where decisions to *not pay* to learn (not learn) were to the right. The structure of the choice questions was such that students only switched between the two columns at most once. Within the willing to receive range, whenever a student moved from willing to receive money to receiving \$0 and not learning their rank, we interpret this as expressing having a willingness to pay so that they *do not* learn their rank. Therefore, the point where students switched between the two represents their WTP/WTA for performance feedback. A screenshot of the WTP/WTA elicitation is found in Figure 1.[10]

[9] We selected this arrangement of percentiles so as not to confuse the students. In our percentile system, performing at a level with only 10% of students performing higher would place students at the 10% level instead of the 90%. We wanted high ranks to correspond to high percentile numbers. We included several examples in the survey to describe what we meant by percentiles.

[10] Our WTP/WTA procedure was piloted before implementing it within our classroom samples. In our pilot procedure, 29 students completed our survey. When discussing the WTP/WTA avoid procedure, no students expressed confusion about what was being asked of them. Furthermore, among the students we discussed the survey with, each student understood that the WTP/WTA question allowed them to express a willingness to avoid information.

2.3 Subsequent Survey Questions

Students were then asked what grade (e.g., A+ to A-, B+ to B-, etc.) they expected to receive in the course. Next, students were asked how many hours per week they needed to study in order to achieve the grade they selected in the previous question. Students were then asked to consider the distribution of ability of students in their economics class. They were then asked to select how likely they thought they were to be in the "top half" of this ability distribution. This question was designed to elicit how well-suited they believed they were for study in economics, despite their beliefs about midterm performance, which may capture idiosyncrasies, such as how students felt on exam day. Lastly, we also elicited students' time preferences using a qualitative question following Falk et al. (2018). The question asked how willing they were to give up an item that would give them immediate benefit but would provide them with even more benefit in the future.

2.4 Treatment and Control Groups: Midterm Rank Information

Another primary motivation for our survey was to measure how students respond to learning about their performance relative to their classmates. As we discuss below, receiving performance feedback information is tightly connected to the WTP/WTA elicitation method. In fact, students were made aware that this procedure is relevant and meaningful regarding both their payoffs and whether they receive relative feedback information or not.

At the beginning of the experiment, the survey generated a random draw between 1 and 21 which dictated what level of the WTP/WTA question would be binding for that student. When determining who is to receive information about midterm rank, this allows us to assign participants into "treatment" and "control" groups. This assignment is almost as good as random as it is to a large extent unrelated to WTP/WTA choices. In fact, students with a similar WTP/WTA are randomly assigned to the treatment group (those

that receive relative feedback information) or to the control group (those that do not receive relative feedback information).[11] When studying our experimental outcomes and to account for eventual differences in WTP/WTA across conditions, we also report several robustness checks, which we include in the appendix.

Students assigned to the treatment group were provided with their actual midterm one percentile rank. We were able to link students to their actual midterm rank using students university ID numbers collected as part of the survey.[12,13] A students' rank on midterm one provides useful information as course grades are often, and in the case of the classes we studied, curved.[14] Along with the rank information, students were also presented with a note that stated that X percent of students performed better, while Y percent performed worse, where $X = rank - 1$ and $Y = 100 - rank$. Students randomized into the control group received no information about their rank on the midterm.

2.5 Follow-up Questions

After students learned their rank, we asked again about the number of hours they needed to study to achieve their intended grade and the probability they were in the top half of the ability distribution.[15] We study these questions as our main belief outcomes.

[11] The reason this procedure is not fully random is that students with extreme preferences will either never or always receive performance feedback. Thus, in the analysis section we will take this into account and perform robustness analyses to address whether differences in WTP/WTA across the two groups affect the experimental results.

[12] Midterm scores were shared with the authors directly after students completed the exam. Scores were then ranked amongst all test takers and percentile rank measures were created for each student. Percentile rank measures were rounded to the nearest percentile before being uploaded to the survey.

[13] Importantly, professors were instructed not to disclose (and did not disclose) any relative performance feedback information to students while the survey was live to guarantee the research design and results.

[14] The economics department studied in this experiment typically employs a B- average for each introductory economics course, although we do not have data on whether these particularly classes were curved.

[15] Students in the control group were not asked for this information again. Our assumption was that the time that would have elapsed for control students would have been less than a minute. Asking students questions again, such as hours they believed they needed to study to achieve their expected grade, would be redundant and possibly confusing.

2.6 Incentive Scheme and Payments

To ensure students answered our survey questions truthfully, the survey was incentivized using the following procedure. At the beginning of the survey, participants were informed that there were two different ways that their responses would affect their total payout. Each of these two possibilities would have been randomly chosen after the payoff relevant decisions were made. Firstly, students could have their beliefs about their actual midterm percentile rank determine their final payoff. Students were told this meant that more accurate beliefs came with greater payoffs. They were also informed that answering truthfully would be a dominant strategy.[16]

Secondly, students could be selected to have their WTP/WTA measure determine their payoff. Under this scenario, students would be allotted $10.00 (of real money) from which they could spend to learn (or avoid learning) their midterm rank. From this amount, they could decide how much to pay to receive or avoid information about their performance on their midterm. In this scenario, one decision (out of the 21 questions) would be randomly drawn and the decision in that question would be binding and determine their final payout. Payments were calculated after completion of the survey and students were paid in Amazon gift cards.

2.7 Follow-up Survey

A brief follow-up survey was administered to all students in our sample one month after the initial survey. As part of the follow-up, students were asked to provide their belief of

[16]The incentive mechanism for the elicitation of prior beliefs consisted in the Binarized Scoring Rule proposed by Hossain and Okui (2013) with a fixed price of $10.00. Under this method, truthful reporting is orthogonal to subjects' risk preferences and it does not rely on expected utility theory. For a detailed explanation of this elicitation procedure see also Schotter and Trevino (2014). We did not explain to subjects how the procedure worked, as withholding the description of the mechanism increases truthful reporting (see Danz et al. (2020)). The interested participant, however, could click on a button to read a detailed description of the elicitation method.

their performance rank on midterm one. This was done primarily to test whether students' memories of their original rank were subject to asymmetry in feedback provision recall (Zimmerman, 2020). Students were incentivized to complete this follow up survey via a raffle for an Amazon gift card of the value of $10.00.

3 Data and Descriptive Results

To study students' beliefs about performance, WTP/WTA measures and how they responded to our intervention, we use responses from students who participated in the survey and exam performance from each class in the study. We also use students' listed genders from administrative data collected through the university registrar for all students who completed their Family Education Rights and Privacy Act (FERPA) release. A total of 1,429 students were enrolled in all six classes combined. Of this group, 285 students began the survey, and 235 students completed it, for a completion rate of 16.4%. Comparing midterm one scores between those who started and those who completed the survey reveals that there is no statistical difference between the two. A total of 152 students completed the follow-up survey, although only 88 of these students had completed the original survey.[17]

For our outcomes of interest, we study students' responses to two belief questions; how many hours students need to study to achieve their desired grade and how likely they are to be in the top half of the ability distribution. We view these two questions as capturing the usefulness, or the instrumental value, of performance feedback as they represent changes to students' beliefs about themselves as a result of new information. Responses for these questions are found in our survey and are described in Section 2. To measure how treatment

[17]For analyses that study willingness to pay and course outcomes, we leverage all the data available. This implies that the total sample in each analysis may change, depending on which questions we are studying. The number presented here comes from a completion of the survey. There were 235 students that completed the survey, but there were 243 students who learned their rank on the exam, for example. Therefore, when studying how students performed in class after learning their rank, we use the 243 number. The results do not change significantly when we consider only the 235 responses in all analyses.

impacted performance, we analyze student learning across the entire course. To do this, we look at a percentage score measure of course performance. We normalize course performance to have a mean of 0 and a standard deviation of 1.

We next study how well our randomization procedure balances pre-treatment variables across treatment groups. Table 1 displays regression analyses of all of our pre-treatment variables on treatment status. Only WTP/WTA is statistically significant. This is not surprising given that whether students were randomized to treatment or control depended to some extent on their WTP/WTA measure.[18] When excluding our WTP/WTA measure, a joint significance test rejects the hypothesis that variables were unbalanced between treatment groups. As mentioned above, we also report robustness checks to test whether our main results are sensitive to including our WTP/WTA measure in our regressions and find that our results are unchanged.

Table 2 shows descriptive statistics for our study participants. 57% of students in our sample are female.[19] Just over two-thirds of the sample indicated that they expected to receive an "A" (A-, A, or A+) in the class, 28% reported they expected a "B", and less than 5% expected a C or lower. Students also reported needing nearly 9 hours of study time per week to achieve their desired grade. When asked about how likely to be in the top half of the ability distribution they believed they were, they reported there was a 75% likelihood, on average.

Looking at performance on midterm one, study participants scored an average of 78 points out of 100.[20] A quick analysis studying how our sample compares to the 1,429 students in

[18] As noted above in the text, in the Appendix we address whether this difference drives our experimental results. In short, we do not find any evidence that differences in WTP/WTA measures across the experimental conditions affect both students' beliefs and class performance.

[19] Comparing the percentage of students in our sample who are female to the population of the university overall, our sample is 57% female while the student body is 61%.

[20] Raw mean and standard deviation measures vary between the six classes. The total possible points for each midterm were not standardized across the classes. Students' scores were normalized so that all scores range between 0 and 1, with each score presenting a percentage of points they received.

our sampling frame who did not take the survey reveals that our students are somewhat positively selected. Students who completed our survey perform about 7 points, or a third of a standard deviation higher, than non-participant students on midterm 1. While this difference in midterm one performance may raise concerns about external validity, we would like to point out that students did not know the nature of the experiment through survey advertisements. While we acknowledge that there may be selection into taking our survey along some dimension that we cannot observe, we can confidently rule out selection based *explicitly* on students' willingness to pay about midterm one performance.

We next study how well students know where their actual performance fell within the midterm one distribution. As mentioned above, information about the mean and variance of midterm one performance was not communicated to students via the professor or Canvas.[21] On average, students ranked near the 42nd percentile on midterm one, indicating that 41 percent of students performed better and 57 percent performed worse. As part of the survey, students were asked to indicate how likely they were to be in each performance decile on midterm one. Figure 2 plots students' beliefs about their performance on midterm one by each decile. Two facts emerge from this figure. First, students placed the highest probability that they were in the top performance decile on average (20%). Secondly, students placed less probability on each successive decile, nearly monotonically.

The first moment of the prior belief distribution is a key piece of information. We therefore create a *mean prior belief* (MPB) variable by multiplying the probability place for each decile probability by its corresponding rank number, and then multiplying this sum by 10. For example, for a student who placed a 20% chance on being in the top decile, a 30% chance in being in the 40-50% decile, and a 50% chance in being in the lowest decile,

[21] While students may not learn their actual rank on a midterm, we believe information about their rank to be relevant for two reasons. First, as student scores are often curved, their position within the class distribution is likely to be very important. The economics department at this university is known to curve course grades. Second, knowing what proportion of students performed better/worse than they did is likely to help them calibrate how much to study or whether to complete the class.

the resulting mean prior belief would be $[1 * (.20) + 5 * (.30) + 10 * (.50)] * 10$ which in this case would equal 67.[22] Creating this for the entire sample, we observe a mean prior belief of 39.45, which is close to the 42nd percentile of actual performance. Moreover, we find that mean prior beliefs are strongly predictive of actual rank in midterm 1 performance.[23]

However, when studying how well calibrated students' beliefs about their own performance are, we find a standard deviation for the difference between actual performance and MPB to be 19 rank points.[24] This indicates a strong potential role for performance feedback in changing students' beliefs about their own ability. We next explore the proportion of students who overestimated or underestimated their performance on midterm one. This can be computed since we know both students' actual rank performance and their prior beliefs. Furthermore, for those students who receive performance feedback we can study how they react to such information *depending* on whether they have received "good" news or "bad" news. In particular, following the experimental literature on belief updating (e.g., Castagnetti and Schmacker (2022); Eil and Rao (2011); Möbius et al. (2022)) we define "good" ("bad") news when a student's disclosed actual rank is lower (higher) than her prior belief (in our setting the MPB). Instead, previous research studying natural experiments typically do not contain information about students' beliefs prior to receiving feedback (Azmat et al., 2019; Azmat and Iriberri, 2010; Bandiera et al., 2015; Goulas and Megalokonomou, 2021). We view this feature of our experiment as an important contribution as we test whether students who over(under)estimated their performance respond differently to feedback later in the paper.

We therefore tabulate how many students had a MPB value that was above their actual midterm rank, meaning that they had overestimated their midterm one performance. Nearly

[22] As a reminder about ranks, higher ranks mean worse performance and lower ranks mean better performance.

[23] A regression of actual rank on mean prior belief estimates a coefficient of 0.988 (p-value < 0.00).

[24] It is worth noting that all students *did* learn their own score on the exam.

half our sample (46.9%) believed they performed better than they actually did. This also implies that students in this group, who received performance feedback information, received "bad" news about their midterm performance.[25]

It is important to note that up until the time students took midterm one, they had very few chances to learn about their relative performance in economic courses.[26] Their midterm one rank, therefore, would provide a significant shock to their beliefs about ability in this field. We therefore hypothesize that the treatment would induce changes in beliefs such that those who overestimated (underestimated) would report needing more (less) time to study and are less (more) likely to be in the top half of the ability distribution. When studying how our treatment might impact final exam performance, our hypotheses are less clear. Previous research has shown that student performance increases as a result of feedback (Azmat et al., 2019; Bandiera et al., 2015), although there is work that shows feedback can induce lower performance (Azmat and Iriberri, 2010). Therefore, we remain agnostic about effects of our treatment on performance.

4 Willingness to Pay/Avoid Information

We next explore students' willingness to pay to either learn their rank or avoid learning their rank on the first midterm. From a purely rational perspective, students should have a non-negative WTP/WTA due to the instrumental value of relative performance feedback. For example, relative performance feedback may reveal information about their returns to

[25] We can also take a more conservative approach when studying prior beliefs. In particular, we can study prior beliefs in terms of mean prior decile. That is, we only take into account what *decile* their MPB is in. We call this measure the *mean decile belief* (MDB). When doing this, we only label those who have a higher (lower) MDB than their actual performance as being underconfident (overconfident). We follow the same method as above to define as "good" and "bad" news the rank information disclosed to them. Results are robust to this alternative approach. For instance, we see that students have a MDB of 46.6 and that our results on under(over)confidence look similar to our MPB.

[26] In fact, as mentioned above, this field experiment was purposely implemented in introductory economic courses taken mainly by first year university students.

study effort and therefore help students calibrate how much to study (Rury and Carrell, 2022; Ersoy, 2023).

Research in behavioral economics, however, finds that information may also contain an *intrinsic* value. In fact, a vivid theoretical and experimental literature suggests that information may directly enter the agent's utility function (Bénabou and Tirole (2002); Brunnermeier and Parker (2005); Möbius et al. (2022); Zimmerman (2020)). For example, information about performance may also increase students' utility (directly) if the student performs well or ranks higher than other students. Conversely, students may prefer to avoid learning about their performance as it may come at a cost to the students' ego. We contend therefore that information about relative performance will carry both an *instrumental* value, which influences students' valuation of this information, and an *intrinsic*, or pure "ego-utility" value.

When considering students' demand for performance feedback information, we posit that the instrumental value will exert a positive effect, while the intrinsic value may exert either a positive *or* negative effect. While previous work has studied the effects of different forms of performance feedback on achievement, we still do not know exactly what students' demand for this information is. To the best of our knowledge, this is the first attempt to measure students' valuation of such information in an actual education context. These preferences for feedback, and how they might influence information processing, have implications for welfare and efficiency.

We next present results from our WTP/WTA procedure described above. Figure 3 presents the results graphically. First, the mean WTP/WTA is of \$1.02. Students, therefore, have a positive WTP/WTA to receive such information. The second result shows that 46% of students were not willing to pay any amount to learn/avoid information about their rank.[27]

[27]It may be the case that students may have a positive instrumental value for this information, but their negative intrinsic value may have canceled that effect out.

Third, the distribution of WTP/WTA is not symmetric. Almost 43% of students are willing to spend a positive amount to learn their rank. On average, they are willing to pay $4.30. However, and importantly, we also find that a statistically significant fraction of students are indeed willing to pay to avoid relative performance feedback.[28] Nearly 10% of students are in this category (with a mean WTA of $7.96). As explained above we see this result of information avoidance as being driven by pure ego-utility.

We next analyze how students' WTP/WTA correlates with *beliefs* about their midterm one performance. It is important to underline here that theoretical work makes no clear-cut predictions on the relationship between beliefs and the demand for performance feedback. For example, Köszegi (2006) predicts that individuals with higher beliefs will seek less information in order not to hurt the ego-utility deriving from such high beliefs, whereas other models predict that individuals enjoy acquiring evidence confirming a positive belief (e.g., Burks et al. 2013). Moreover, Bénabou and Tirole (2002) also make no prediction as the authors write that it will depend on both the instrumental and intrinsic values of information. It is therefore a question to be answered empirically. Our results are presented in Table 3. Here, we find no statistically significant correlation between having a higher mean prior belief and a higher WTP/WTA. This finding contradicts other work in the experimental ego-utility literature (e.g., Burks et al. 2013).[29]

Our results therefore show no straightforward correlation between prior beliefs and demand for information on aggregate. This may be because the instrumental and intrinsic values of information may be higher or lower depending on beliefs about midterm one performance and could be influencing students WTP/WTA responses differently. In the case of high beliefs, the instrumental value of information may be low, as students may be quite

[28]Statistically significance is assessed by running a test of equality in means that WPT/WTA (for those with WTP/WTA measures equal or less than 0) is equal to 0. The test rejects the null hypothesis in equality of means (p-value < 0.00).

[29]We also do not find any significant correlation between rank performance and WTP/WTA preferences.

confident about their performance, while the intrinsic value may be high, for those students who are motivated by learning about their rank. On the other hand, for low beliefs, the instrumental value of information may be quite high, as one can re-calibrate how study effort maps onto exam performance. On the other hand, the news of a low rank may be difficult to digest from an intrinsic value perspective, and it may be worth protecting oneself from unflattering, bad news (Castagnetti and Schmacker, 2022; Eil and Rao, 2011; Möbius et al., 2022). For completeness, we also present the results of the correlation between elicited WTP/WTA measures and *actual* performance on midterm one. Table 3 also shows that there is no statistically significant relationship.

Lastly, we study how students' gender influences preferences for feedback. We find that being a female is associated with a $1 increase in students' WTP/WTA measure (see Figure 4). This result cannot be explained either by differences in rank or by differences in prior beliefs. We thus see this result as differences in preferences for information by gender. Results in laboratory experiments show either no differences by gender in information seeking behavior in ego relevant tasks such as IQ tasks and a beauty ranking task (Castagnetti and Schmacker, 2022; Eil and Rao, 2011) or that women are more averse to feedback in a male-stereotypical task (Sharma and Castagnetti, 2023). Our results are therefore at odds with previous literature. It is however important to note that our setting is different from the studies mentioned above, as here we are capturing preferences in an educational context in which the instrumental value of information goes beyond the experiment. It is therefore plausible that this feature explains the difference between our results and those in the literature.

5 Experimental Results

5.1 Average Effects

To estimate treatment effects from our performance feedback intervention, we estimate the following statistical model:

$$y_i = \alpha + \beta T_i + \gamma P_i + \epsilon_i$$

where y_i is an outcome of interest, T_i is an indicator for receiving information on midterm one rank, P_i contains pre-treatment variables. In particular, when the outcome variable is Hrs/week needed to study, the pre-treatment variable corresponds to the number of hours students selected before treatment; when the outcome variable is the probability of students believing they are in the top half, the pre-treatment variable is the same probability students selected before treatment; finally, when the outcome variable is overall course performance (as measured in normalized percentage points earned in the course), the control variable is midterm one performance. ϵ_i is a random disturbance term. We estimate robust standard errors.[30]

Table 4 presents results on beliefs and performance. First, we see that treatment induced students to report needing more hours to study to achieve their desired grade. The estimated treatment effect is 1.49 hours and is significant at the 1% level. Students also downgraded how likely they were to be in the top half of the ability distribution by -4.87 points ($p < 0.01$). When studying the effect on learning in the course, we see in Table 4 that treatment has a null effect on exam performance across the entire sample.[31]

[30] In Section 5.4 we show that adding control variables to our estimations do not change the results.
[31] We also study whether the treatment changes the propensity to take future economics courses.

5.2 Results by Type of Information Received

We next explore effects on students' beliefs by focusing on students who over- or underestimated their midterm one rank. To do so, we estimate two separate regressions, where each regression conditions on whether students received "good" or "bad" news. Table 4 shows that students who overestimated their performance (received negative news) report needing 2.86 more hours to achieve their grade ($p < 0.01$). Instead, those who underestimated their performance report needing 0.46 more hours ($p < 0.05$).[32] A t-test confirms that these two responses are significantly different from each other. More pronounced differences emerge studying likelihood of being above average in the class. Those who overestimated their performance reported being 13.7 pp less likely to be above average in the class ($p < 0.01$), while those who underestimated their performance report being 3.97 pp more likely ($p<0.01$). When studying the effect on exam scores, we do not find significant changes when looking at students who received bad news, although the effect is positive. When looking at students who received good news, however, we find a significant negative impact on overall course score. Learning your relative performance was better than expected on your first midterm causes a -0.182 SD decrease in overall learning in the course.

5.3 Results by WTP/WTA

Finally, it is also relevant to study whether treatment effects depend on students' WTP/WTA measures. One reason we might see differential effects based on students' WTP/WTA is that students with a negative intrinsic value for information may protect themselves from feedback that may come with a negative shock to their utility. To cleanly analyze this effect, we run the same model as above with the addition of two variables: student's WTP/WTA measure and the interaction between the treatment indicator and the WTP/WTA. The latter

[32] It is important to note that this result specifically is not robust since it disappears when adding control variables or when using alternative measures for the prior belief and rank as can be seen in Appendix B.

allows us to cleanly capture the causal effect of providing information on information processing by students' WTP/WTA. Table 5 shows the results. While our results by treatment status remain the same, we do not find any evidence that the treatment effects depend on students' WTP/WTA. We therefore conclude that students do not exhibit (instantaneous) self-deception after learning about their relative performance.

5.4 Robustness of our Results

In the Appendix, we perform several robustness checks to analyze the stability of our main findings. In particular, in Appendix B.1 we add relevant control variables to our econometric estimations. In Appendix B.2 we perform sub-sample analyses based on WTP/WTA measures to address the identified differences across the experimental conditions at the baseline, as described in Section 3. Whereas, in Appendix B.3, we use a different approach to define the mean prior belief and the student's rank. Overall, our results are robust to these alternative estimations. Finally, in Appendix B.4 we run further analyses to also rule out that the treatment effect does depend on students' WTP/WTA or that it affects the main outcome variables depending on whether the students received "bad" or "good" news.

5.5 Follow-Up Survey

We also conducted a brief follow-up survey a month after the initial intervention was complete. This survey asked students to recall what their rank was on midterm one. This approach is similar to tests performed in Zimmerman (2020), who studies whether biased beliefs persist because of asymmetric memory. In our case, students may resist performance feedback information by selectively forgetting results that might be interpreted as being unflattering. We therefore estimate the following model:

$$remember_i = \alpha + \beta_1 negative_i + \beta_2 T_i + \beta_3 negative_i * T_i + \gamma P_i + \epsilon_i$$

where $remember_i$ equals the students' belief of the rank one month after the intervention minus their beliefs of the rank elicited during our original intervention. Here we study potential differential effects based on receiving negative news (as in Zimmerman (2020)), as we have few observations that have a negative WTP in our follow-up survey sample. To ensure that belief updating between students who receive good and bad news is comparable, we multiply our outcome $remember_i$ by -1 for students who receive bad news. The most common case of asymmetric memory is students forgetting the negative feedback they received earlier in the course. Therefore, if students suffered from asymmetric memory, we would expect to see β_3 to be positive and significant. Results are presented in table 5. We see that the interaction between treatment and receiving negative news is positive, which potentially captures inflated memories of midterm one performance, but is not statistically significant. As we face severe power issues due to a small sample size, we advocate future research study on whether students face biases from asymmetric memory when provided with performance feedback information.

6 Discussion and Conclusion

In this paper, we design and administer a survey to study students' preferences over relative performance feedback. We find that on average students are willing to pay around $1.00 to learn their rank. Interestingly, we also show that 10% of students are willing to pay to *avoid* learning their rank. Because information has potentially opposing effects on demand (instrumental and intrinsic), we conservatively characterize students with a willingness to avoid value as being an underestimate of having a negative intrinsic value. We also show that female students are willing to pay $1 more to learn their rank than male students. This

finding underlines important gender differences for performance feedback which may hold important implications on how different genders seek relative performance feedback and, consequently, how much they learn thanks to it.

We also find that students hold inaccurate beliefs about their rank on the midterm exam. When studying how they respond to learning their midterm rank, we find students update beliefs about the number of hours they need to study per week as well as how likely they are to be in the top half of the class ability distribution. This underscores the benefit of performance feedback. More importantly, the effects of the information disclosure depend on the type of information conveyed. That is, students who receive bad news report needing more study hours and being less likely to be above average in the class. Students who receive good news believe they are more likely to be above average, although the effect of needing more (or less) hours to achieve their desired grade is not clear cut.[33] We also fail to detect any effects on achievement.

Results from this experiment help us confirm further important facts about performance feedback in classes. First, beliefs about performance do not predict preferences for information on feedback. This contradicts several papers in the experimental literature studying motivated reasoning. Second, relative performance feedback provides information to students about their own ability, which allows them to incorporate such information into their beliefs regarding their effort and overall ability. Third, we show that WTP/WTA measures do not mediate treatment effects on beliefs. This demonstrates that students do not deceive themselves or exhibit other information processing biases when updating their beliefs and/or making decisions about their human capital.

Our results highlight the fact that students do not hold uniform preferences over learning about their relative performance. These preferences, together with the positive effects

[33]While we find a positive statistically significant effect in the main regression, this result does not hold once we perform robustness analyses.

of providing relative performance feedback on beliefs that we document (and that others document on performance outcomes too (Azmat and Iriberri, 2015; Azmat et al., 2019), raise the question on whether educational institutions should provide such information to students, whether on demand or by default.[34] While further work is needed to draw policy implications, our paper highlights that one cannot derive conclusions only on the effects of information on relevant outcomes, but should also consider students' preferences. In fact, future work should address whether sharing this information to those students that attach a negative value to performance feedback is ultimately beneficial or not, taking into account its consequences, including on students' (ego) utility. It remains an open question whether the effect on utility from the intrinsic value of information (if it is negative) is greater than the instrumental value students receive upon learning about their performance.

We also recommend that future work study other features of relative performance feedback. For instance, future research should study how preferences for information change as students' knowledge of their own ability changes. We also conducted this experiment at a selective research university where the average academic preparation was quite high. Despite this, we found that 10% are willing to pay to avoid information about their performance. Even though we found preferences are unrelated to beliefs or performance, it is possible that negative preferences for information might be higher within populations with a large fraction of students who are struggling academically. In these cases, it would be informative to learn whether preferences for feedback influence students' decisions to take on educational investments. This is particularly true for those that come with performance feedback such as taking the SAT or ACT, which serve as important proxies for college attendance.

[34]It is important to note that across institutions, education levels, and countries the provision of relative performance feedback varies significantly, although the cases in which this information is disclosed are less frequent.

References

Azmat, G., M. Bagues, A. Cabrales, and N. Iriberri (2019): "What You Don't Know... Can't Hurt You? A Natural Field Experiment on Relative Performance Feedback in Higher Education," *Management Science*, 65, 3714–3736.

Azmat, G. and N. Iriberri (2010): "The Importance of Relative Performane Feedback Information: Evidence from a Natural Experiment Using High School Students," *Journal of Public Economics*, 94, 435–452.

——— (2015): "The Provision of Relative Performance Feedback: An Analysis of Performance and Satisfaction," *Journal of Economics and Management Strategy*, 25, 77–110.

Bandiera, O., V. Larcinese, and I. Rasul (2015): "Blissful Ignorance? A Natural Experiment on the Effect of Feedback on Students' Performance," *Journal of Labour Economics*, 34.

Bénabou, R. and J. Tirole (2002): "Self-confidence and Personal Motivation," *The Quarterly Journal of Economics*, 117, 871–915.

——— (2016): "Mindful Economics: The Production, Consumption, and Value of Beliefs," *Journal of Economic Perspectives*, 30, 141–64.

Bobba, M. and V. Frisancho (Forthcoming): "Self-Perceptions about Academic Achievement," *Journal of Econometrics*.

Bordalo, P., K. Coffman, N. Gennaioli, and A. Shleifer (2019): "Beliefs about Gender," *American Economic Review*, 109, 739–73.

Brade, R., O. Himmler, and R. Jäckle (2018): "Normatively Framed Relative Performance Feedback - Field Experiment and Replication," *MRPA Working Paper*.

Brunnermeier, M. K. and J. A. Parker (2005): "Optimal expectations," *American Economic Review*, 95, 1092–1118.

Burks, S. S., J. P. Carpenter, and L. Goette (2013): "Overconfidence and Social Signalling," *Review of Economic Studies*, 80, 949–983.

Castagnetti, A. and R. Schmacker (2022): "Protecting the Ego: Motivated Information Selection and Updating," *European Economic Review*, 142, 104007.

Chew, S. H., W. Huang, and X. Zhao (2020): "Motivated False Memory," *Journal of Political Economy*, 128, 3913–3939.

Coffman, K., M. Paola Ugalde Araya, and B. Zafar (2021): "A (Dynamic) Investigation of Stereotypes, Belief-Updating, and Behavior," *Working Paper*.

Coffman, K. B. (2014): "Evidence on Self-stereotyping and the Contribution of Ideas," *The Quarterly Journal of Economics*, 129, 1625–1660.

Coffman, K. B., M. Collis, and L. Kulkarni (2020): *When to Apply?*, Harvard Business School.

Coutts, A., L. Gerhards, and Z. Murad (2020): "Who to blame? Self-serving Attribution Bias with Multi-dimensional Uncertainty," Tech. rep.

Danz, D., L. Vesterlund, and A. J. Wilson (2020): "Belief elicitation: Limiting Truth Telling with Information on Incentives," Tech. rep., National Bureau of Economic Research.

Dobrescu, I., M. Faravelli, R. Megalokonomou, and A. Motta (2021): "Feedback, Learning and Performance: Evidence from a Randomized Controlled Trial," *Economic Journal*.

Eil, D. and J. M. Rao (2011): "The Good News- Bad news Effect: Asymmetric Processing of Objective Information about Yourself," *American Economic Journal: Microeconomics*, 3, 114–38.

Ersoy, F. (2023): "Effects of Perceived Productivity on Study Effort: Evidence from a Field Experiment," *Journal of Economic Behavior and Organization*.

Exley, C. L. and J. B. Kessler (2019): "Motivated errors," Tech. rep., National Bureau of Economic Research.

——— (2022): "The Gender Gap in Self-promotion," *The Quarterly Journal of Economics*, 137, 1345–1381.

Falk, A., A. Becker, T. Dohmen, B. Enke, D. Huffman, and U. Sunde (2018): "Global Evidence on Economic Preferences," *The Quarterly Journal of Economics*, 133, 1645–1692.

Ganguly, A. and J. Tasoff (2016): "Fantasy and Dread: the Demand for Information and the Consumption Utility of the Future," *Management Science*, 63, 4037–4060.

Goldman, R., D. Hagman, and G. Loewenstein (2017): "Information Avoidance," *Journal of Economic Literature*, 55, 96–135.

Goulas, S. and R. Megalokonomou (2021): "Knowing who you Actually are: The Effect of Feedback on Short-and Longer-term Outcomes," *Journal of Economic Behavior & Organization*, 183, 589–615.

Hossain, T. and R. Okui (2013): "The Binarized Scoring Rule," *Review of Economic Studies*, 80, 984–1001.

Karlsson, N., G. Loewenstein, and D. Seppi (2009): "The Ostrich Effect: Selective Attention to Information," *Journal of Risk and Uncertainty*, 38, 95–115.

Kőszegi, B. (2006): "Ego Utility, Overconfidence, and Task Choice," *Journal of the European Economic Association*, 4, 673–707.

Li, H.-H. (2018): "Do Mentoring, Information and Nudge Reduce the Gender Gap in Economics," *Economics of Education Review*, 64, 165–183.

Möbius, M. M., M. Niederle, P. Niehaus, and T. S. Rosenblat (2022): "Managing Self-confidence: Theory and Experimental Evidence," *Management Science*, 68, 7793–7817.

Oster, E., I. Shoulson, and E. Dorsey (2013): "Limited Life Expectancy, Human Capital and Health Investments," *American Economic Review*, 103, 1977–2002.

Owen, S. (2022): "College Field Specialization and Beliefs about Relative Performance: An Experimental Intervention to Understand Gender Gaps in STEM," *Working Paper*.

Rury, D. (2022): "Tightening the Leaky Pipeline(s): The Role of Beliefs About Ability in STEM Major Choice," *Working Paper*.

Rury, D. and S. Carrell (2022): "Knowing What it Takes: The Effect of Information About Returns To Studying on Study Effort and Achievement," *Working Paper*.

Schotter, A. and I. Trevino (2014): "Belief Elicitation in the Laboratory," *Annual Review of Economics*, 6, 103–128.

Sharma, K. and A. Castagnetti (2023): "Demand for Information by Gender: Evidence from a Laboratory Experiment," *Journal of Economic Behavior & Organization*, 207, 172–202.

Sicherman, N., G. Loewenstein, D. J. Seppi, and S. P. Utkus (2015): "Financial Attention," *The Review of Financial Studies*, 29, 863–897.

Zimmerman, F. (2020): "The Dynamics of Motivated Beliefs," *American Economics Review*, 110, 337–363.

7 Figures and Tables

WTP/WTA Elicitation Procedure

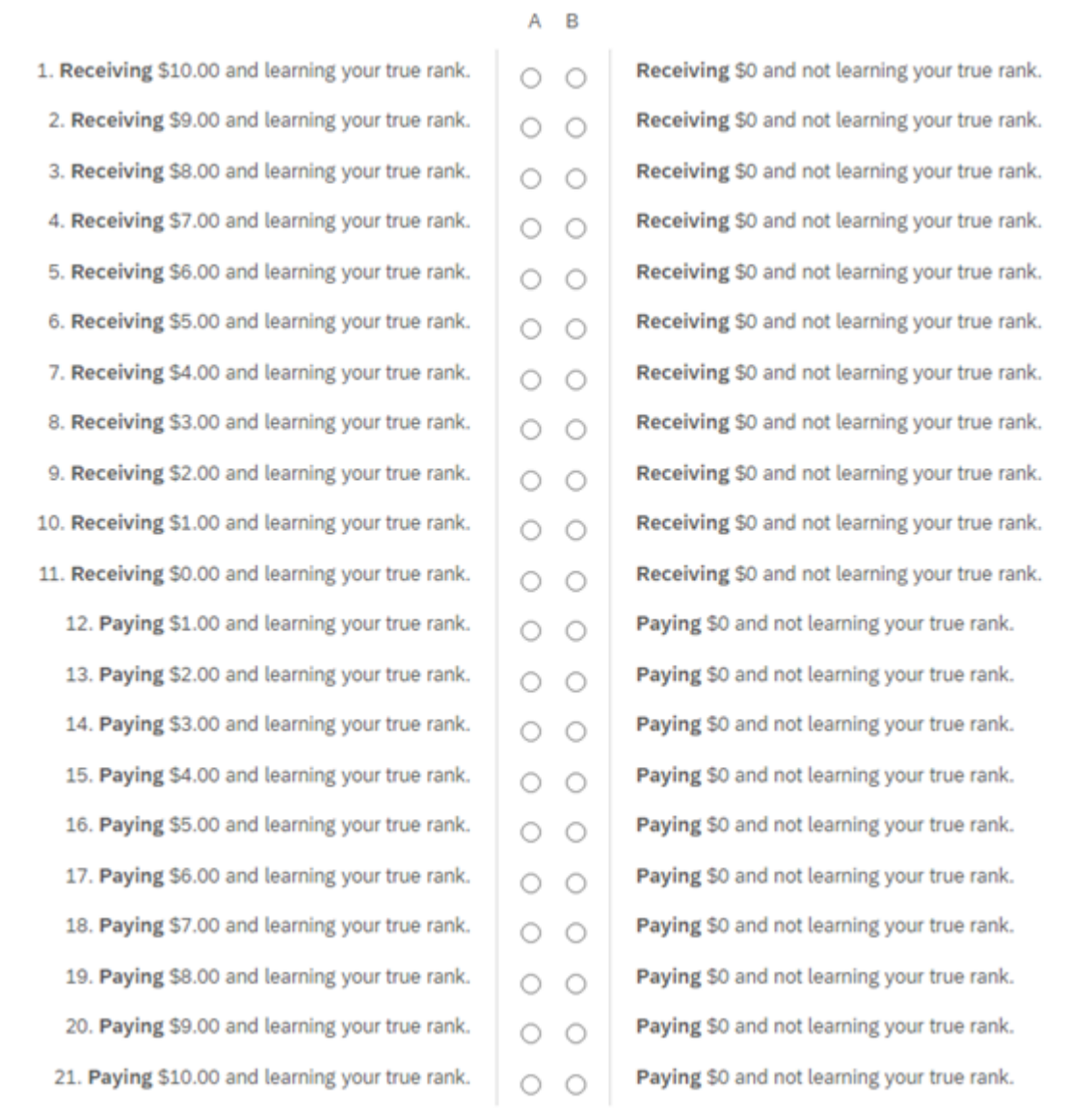

Figure 1
WTP/WTA Elicitation Procedure

Note: This figure shows the willingness to pay and willingness to avoid elicitation procedure.

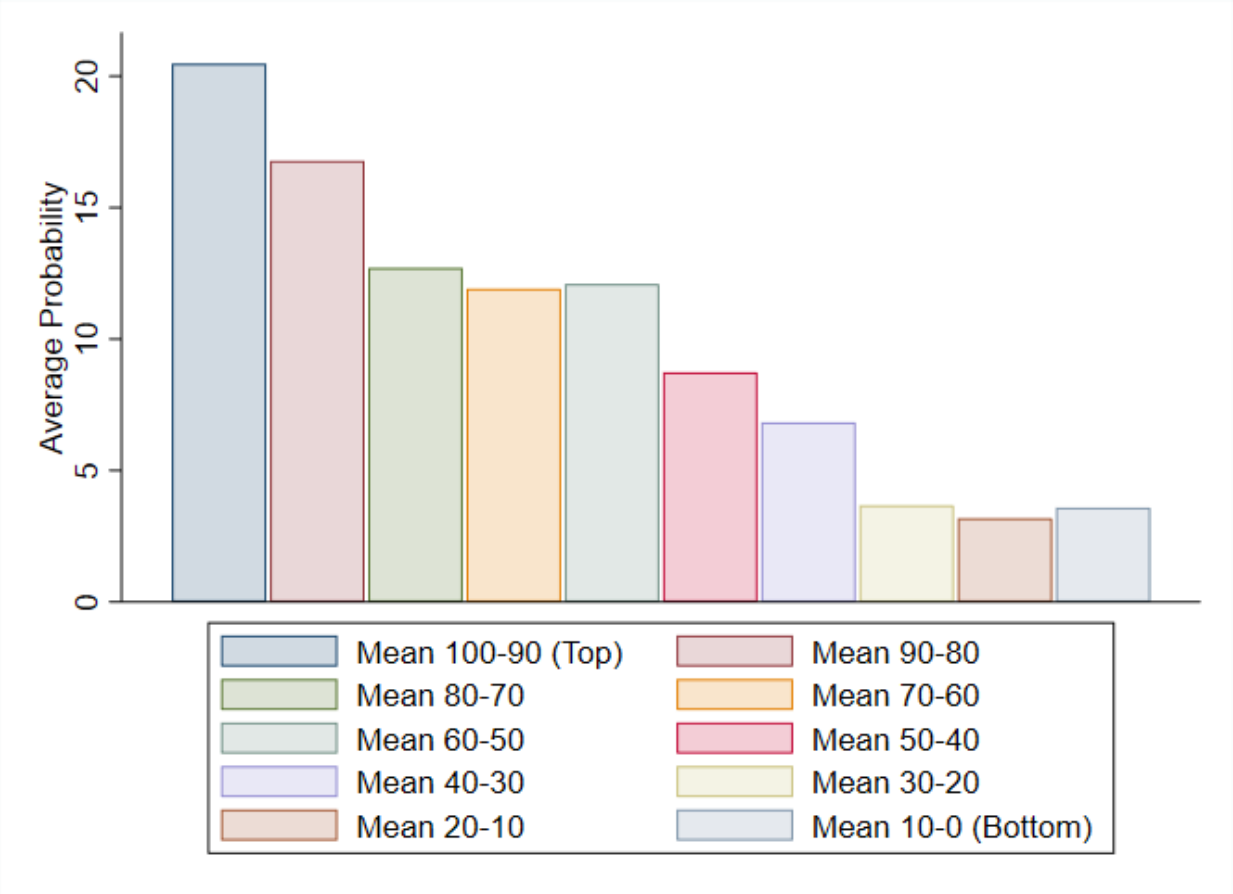

Figure 2
Student Beliefs About Midterm Performance

Note: This figure plots the average probability that students placed for each performance decile for their midterm one performance. Students were constrained so that their probabilities would sum to one.

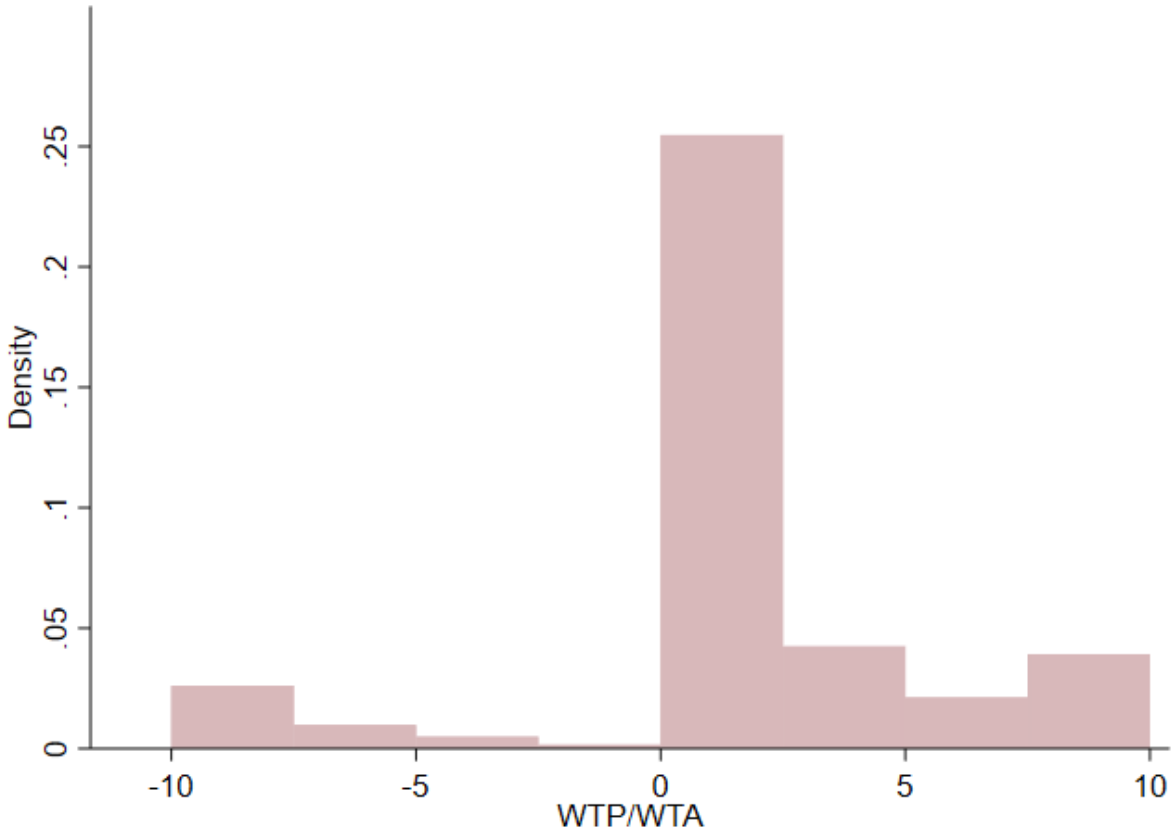

Figure 3
Willingness to Pay/Avoid for Midterm 1 Rank

Note: This figure plots the willingness to pay and willingness to avoid midterm one rank information for the entire sample.

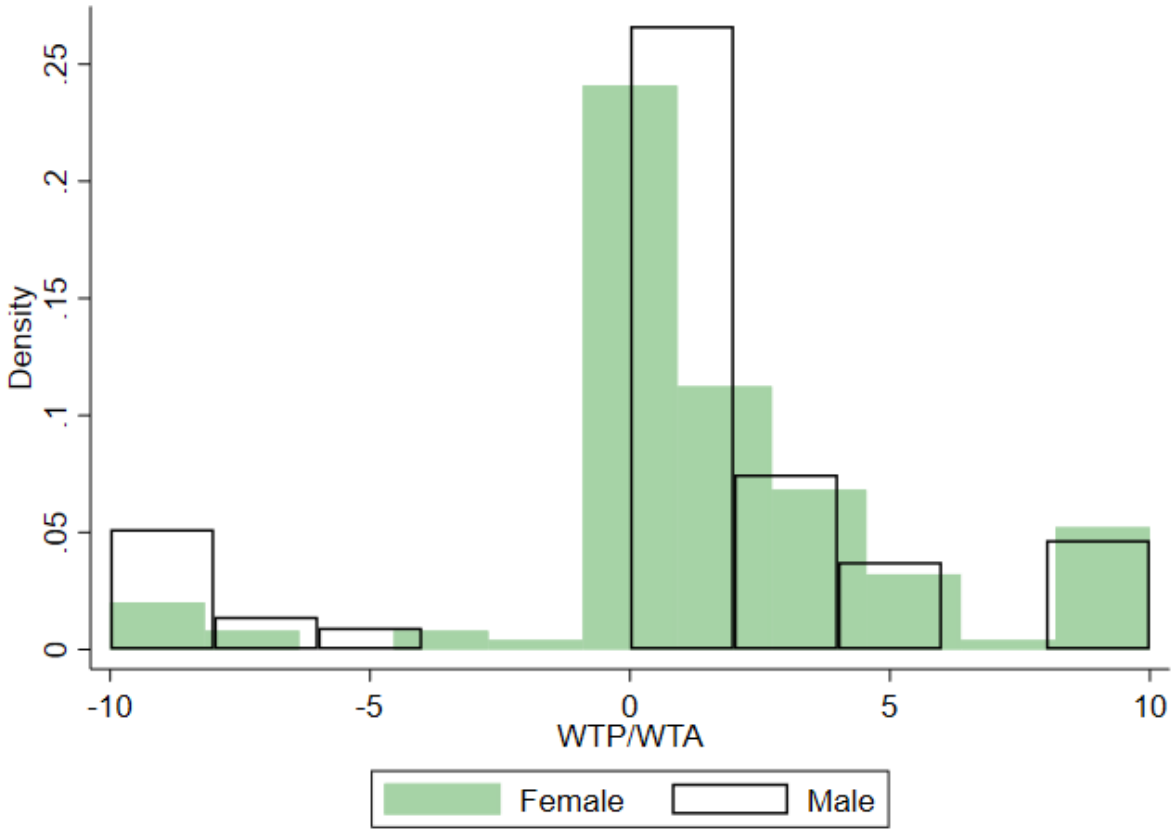

Figure 4
Willingness to Pay/Avoid for Midterm 1 Rank

Note: This figure plots the willingness to pay and willingness to avoid midterm one rank information measures for male and female students.

Table 1
Balance Tests

	(1) Treat
WTP/WTA	0.045***
	(0.007)
Midterm 1 Beliefs	-0.002
	(0.002)
Expected A grade	-0.140
	(0.171)
Expected B grade	-0.226
	(0.158)
Female	-0.036
	(0.061)
Hrs/Week Needed	0.000
	(0.005)
Prob(Top Half)	0.001
	(0.002)
Midterm 1 Performance	0.622
	(0.471)
Observations	231

Notes: This table reports results from a single regression that studies the relationship between the variables listed in the columns and likelihood of being in the treatment group. Expected A and B variables represent students reported beliefs about their expected grade. Midterm beliefs represent students' mean prior beliefs variable which takes a weighted sum of how well students believed they performed on midterm one. Hours per week needed represents students beliefs about how many hours students believe they need to study to achieve the grade they reported earlier in the survey. The probability students believe they are in the top half of the ability distribution also comes from the survey. Midterm 1 performance is a variable in [0,1] that captures how well students performed on that assessment. star(* 0.10 ** 0.05 *** 0.01)

Table 2
Descriptive Statistics

	Count	Mean	SD	Min	Max
Female	253	0.561	0.497	0	1
Midterm 1 Rank	253	40.929	29.871	1	100
Midterm 1 Beliefs	247	39.167	23.375	10	100
Diff Rank - Beliefs	246	1.682	19.248	-60.450	63
Midterm 1 Performance	242	0.789	0.163	0.240	1
Expected A grade	244	0.672	0.470	0	1
Expected B grade	244	0.279	0.449	0	1
Expected C grade	244	0.049	0.217	0	1
Hrs/Week Needed	242	8.961	6.719	0	40
Prob(Top Half)	244	75.426	25.548	0	100
WTP/WTA	244	1.000	4.307	-10	10

Notes: This table reports descriptive statistics for the sample used in our analysis. Expected A, B and C variables represent students reported beliefs about their beliefs about their expected grade. Midterm beliefs represent students' mean prior beliefs variable which takes a weighted sum of how well students believed they performed on midterm one. Hours per week needed represents students' beliefs about how many hours students believe they need to study to achieve the grade they reported earlier in the survey. The probability students believe they are in the top half of the ability distribution also comes from the survey. Midterm 1 performance is a variable in [0,1] that captures how well students performed on that assessment. star(* 0.10 ** 0.05 *** 0.01)

Table 3
WTP/WTA Mediators

	(1) WTP/WTA	(2) WTP/WTA
Mean Prior Belief	-0.000	-0.002
	(0.020)	(0.020)
Midterm Rank Performance	0.009	0.008
	(0.017)	(0.017)
Female		1.220**
		(0.509)

Notes: This table reports results from two separate regressions that study how beliefs about midterm one performance and actual midterm rank performance and whether or not the student is female impact students' willingness to pay/avoid information about their rank. Column one only includes students' beliefs and rank performance, while column two adds whether the student is female. star(* 0.10 ** 0.05 *** 0.01)

Table 4
Treatment Effects

	(1) Hrs/Week Needed	(2) Prob(Top Half)	(3) Course Performance
Average Treatment Effects			
Treat	1.496***	-4.872**	-0.066
	(0.253)	(1.766)	(0.078)
Bad News/Good News			
Bad News (Treat)	2.721***	-13.212***	0.135
	(0.500)	(2.83)	(0.126)
Good News (Treat)	0.407**	4.054***	-0.182**
	(0.185)	(1.449)	(0.091)

Notes: This table reports results from a regression that study the outcomes listed in the column headers. Regressions include an indicator for treatment status and the corresponding pre-treatment variable for each outcome. For overall course performance, we include students' midterm one performance. Robust standard errors are in parenthesis To study regressions looking at students who receive bad or good news, we condition on the students' mean prior belief variable and then compare that to the students actual rank. star(* 0.10 ** 0.05 *** 0.01)

Table 5
Treatment Effects (Follow-Up Survey)

	(1) Remember Rank
Treat	-4.313
	(9.554)
Negative	-9.485
	(10.991)
Negative*Treat	3.787
	(13.119)

Notes: This table reports results from a series of regressions that study the outcomes listed in the column headers. Regressions include an indicator for treatment status, type of information received, and their interaction. We estimate robust standard errors. star(* 0.10 ** 0.05 *** 0.01)

A Experimental Instructions

A.1 Welcome Page

Welcome to the research study!

We are interested in understanding students beliefs about their performance and their preferences for information. In particular, you will be asked to answer some questions about one of the courses you are studying this term: ECN XX. Please be assured that your responses will be kept completely confidential.

The study should take you around 10-15 minutes to complete. Your participation in this research is voluntary. You have the right to withdraw at any point during the study, for any reason, and without any prejudice. Participating (or not participating) will not affect student's grades or standing in the class. If you would like to contact the Principal Investigator in the study to discuss this research, please e-mail Derek Rury at drury@ucdavis.edu.

You will be compensated for your participation in this survey as some of the questions in this survey will be monetarily incentivized. More specifically, you may earn between $0 and $20 as part of taking this survey. Compensation will be calculated only after the survey is completed and payments will be processed within 3 weeks. The amount you earn will depend on two factors; 1) your responses to the two incentivized questions and; 2) which of these two questions is selected to calculate your payout, which will be randomly determined. We provide more details on payments later in the survey.

On the next page you will be asked whether or not you'd like to allow the UC Davis registrar to release your academic records for the purposes of this study. Under the Family Education Rights and Privacy Act (FERPA) your academic records cannot be released without your consent. Please read the page carefully. We would also like to clarify that no personal identifying information will be used as a part of this study.

Please click the arrow below if you wish to proceed to the survey.

A.2 Informed Consent

[Informed consent where students had to read the policy and then added name and the university's ID number.]

A.3 Prior Beliefs Elicitation Instructions (1)

Let's begin the survey by considering your own performance on midterm one in ECN XXX.

What do you believe your rank is in terms of scores on midterm one compared to everyone else in the course?

Here are some examples:

- You could be the top performer in the exam. That is, your performance is in the top 1%.
- You could be in the middle of the distribution. That is, your performance is in the top 50%.
- You could be the worst performer in the exam. That is, your performance is in the top 100%.

In short, the better you believe you performed relative to the class, the closer your rank should be to the top 1%. The worse you believe you performed, the closer your rank should be to the top 100%.

In this survey, we will ask you to answer some questions about what you believe your rank in midterm one is.

A.4 Prior Beliefs Elicitation Instructions (2)

Now, we are going to form ten groups of students based on their relative performance in midterm one:

1. Group 1: those whose score is in the top 10% (the top 10% of performers)
2. Group 2: those whose score is between the top 20% to top 10%.
3. Group 3: those whose score is between the top 30% to top 20%.
4. Group 4: those whose score is between the top 40% to top 30%.
5. Group 5: those whose score is between the top 50% to top 40%.
6. Group 6: those whose score is between the top 60% to top 50%.
7. Group 7: those whose score is between the top 70% to top 60%.
8. Group 8: those whose score is between the top 80% to top 70%.
9. Group 9: those whose score is between the top 90% to top 80%.
10. Group 10: those whose score is between the top 90% and 100% (or, equivalently, worst 10% performers).

If the score falls into more than 1 group (because of ties in scores), it will be randomly determined to which one of these groups it corresponds.

Your Belief about Your Group Rank in Midterm one

Now we want to know to which group you think your performance in midterm one belongs. That is, what do you believe is your group rank in terms of scores on midterm one compared to everyone else in the course.

In particular, we are interested in your estimate of the likelihood of being in each of these 10 groups mentioned above. We are going to ask you to state the probability with which you think you belong to each group. Please keep in mind that the sum of these 10 probabilities need to be exactly 100%.

A note about your payments

You may be able to earn up to $10 as part of this question. The probabilities you state will affect your payments if this part of the survey is randomly selected to count for payments (there are two questions that can be randomly selected to determine your payments).

In short, the more accurate your answers, the more likely it is that you earn $10 instead of $0. That is, the more percentage points you assign to your actual rank (the group you belong based on your actual performance in midterm one) and the fewer percentage points you allocate to the groups you do not belong, the more likely it is that you earn $10. This implies that you have the highest chance of earning the $10 by stating what you really believe your rank is.

If you would like to read a detailed description of this payment procedure, please click "Yes, I'd like to see how my payments are calculated". If you'd prefer to continue with the survey, please click "No, I'd like to continue with the survey".

A.5 Calculation Payments Belief Elicitation

Details on the calculation of your additional payments

It is not necessary that you read and fully understand the following section on the calculation of your additional payments. In fact, you can skip this part. What it is important, instead, is that you understand that it is in your best interest to state what you really believe your true rank is.

After you state the probabilities of being in each of the ten groups, the computer will randomly draw another number k. This number is between 0 and 100,000. (More precisely, this number is drawn from a discrete uniform distribution on the interval 0 to 100,000. You will receive the $10 if the sum S is smaller or equal to k. S is the sum of the following elements:

1) The squared deviation between the number of points that you allocated to your actual group rank, and 100 points.

2) For each group rank that you do not belong: the squared deviation between 0 points and the number of points that you allocated to this group rank. If the sum S is smaller or equal than k you will receive \$10, otherwise \$0. The payoff rule is therefore as follows: \$10 if $S \leq k$, \$0 if $S > k$ This means the following: If the sum of the squared deviations exceeds the drawn k number, you will receive 0. If, however, the sum of the squared deviations is smaller (or equal) than k, you will receive \$10. You can notice here that it is in your best interest to: Keep the difference between the points allocated to your actual rank and 100 points as low as possible. That is, to allocate as many points as possible to your actual rank Allocate as few points as possible to every other rank that does not correspond to your actual group rank.

A.6 Prior Belief Elicitation

Please state your beliefs of belonging to each group in terms of your midterm performance in ECN XX.

(each entry should be an integer between 0 and 100 with the sum of all entries equaling 100).

- Probability that you belong to Group 1 (top 10% performers - highest performers):
- Probability that you belong to Group 2 (between top 20% to top 10% performers):
- Probability that you belong to Group 3 (between top 30% to top 20% performers):
- Probability that you belong to Group 4 (between top 40% to top 30% performers):
- Probability that you belong to Group 5 (between top 50% to top 40% performers):
- Probability that you belong to Group 6 (between top 60% to top 50% performers):
- Probability that you belong to Group 7 (between top 70% to top 60% performers):

- Probability that you belong to Group 8 (between top 80% to top 70% performers):
- Probability that you belong to Group 9 (between top 90% to top 80% performers):
- Probability that you belong to Group 10 (top 90% performers, or equivalently worst 10% performers):

A.7 WTP/WTA Instructions (1)

You will now have the opportunity to learn to which group you belong based on your midterm one performance. In other words, you can learn your performance rank on midterm one within your class.

Whether we reveal this information to you will be determined by your choices in the following question. Your choices will may also determine your payment, so please consider these choices carefully.

The decision you have to make will proceed as follows. You will have to make 21 choices indicating whether your prefer option A or option B. In each you will have to decide to either pay a certain amount of money to learn your rank or not pay anything and not learn your rank. Some options ask you how much money you would be willing to receive to learn your rank as opposed to not receiving any money and not learning your rank.

If this question was selected to determine your payments, you will begin the question with $10 (in real money!) Once you make your choices, we will randomly choose a number between 1 and 21 and execute your decision for that number (this includes both payments and learning your rank).

Please select if you would like to see an example how this question will work or if you would like to proceed to making your selections.

A.8 WTP/WTA Instructions (2)

Here is an example:

Example: Let's say you are willing to receive any amount of money ($10 to $0) offered to learn you rank, and also that you're willing to pay up to $3 to learn your rank, but not $4 or greater.

A random number from 1 to 21 will be picked to be determined if you see your rank and your payment.

Let's say that number 13 is randomly chosen for this question to determine your payment and whether you learn your rank. This number corresponds to the comparison between willing to pay $2 to learn your rank or pay $0 and not learn your rank. In this example, you indicated that you indeed would. In this scenario, you would have $2 deducted from your $10 initial sum and learn your rank later in the survey.

If number 17 would have been selected, which corresponds to paying $6 to learn your rank, you would neither pay $6 nor learn your rank as you indicated you would rather pay $0 and not learn your rank for that pair.

Please click the arrow below to proceed.

A.9 WTP/WTA Price List

Before you make your selections, please note that if your preferences are sensible, you will only switch between columns A and B no more than once. For example, if you are willing to receive $8 to learn your rank, not willing to pay $2 to learn your rank, but are willing to pay $5 to learn your rank, this would not be sensible as paying $5 to learn your rank costs you more than $2 to do so.

To ensure that your responses are sensible, we will restrict your responses so that you can only switch between columns A and B no more than once.

Would you prefer:

[Price List Here]

A.10 Second Part

Now we will ask you a few questions about your expectations for this course.

A.11 Elicitation Grade Expectation

What grade do you expect to get in ECN XX?

A.12 Elicitation Hours Needed to Study to Achieve Desired Grade

How many hours per week do you believe you need to study for ECN XX to achieve the grade you selected in the previous question?

A.13 Elicitation Ability

Now consider your inherent ability to do well in ECN XX.

Similar to a midterm score, there is a distribution of ability to do well in this class that exists when considering all of your classmates' abilities.

What do you think is the probability you are in the top half of the ability distribution in this class? (answer should be between 0% and 100%, where 0% means you are certain that you are NOT in top half, while 100% means that you are certain that you are in the top half of the ability distribution).

[Slider Here]

A.14 Elicitation Time Preferences

How willing are you to give up something that is beneficial for you today in order to benefit more from that in the future? (Please indicate your answer on a scale from 0 to 10, where 0 means you are "completely unwilling to do so" and a 10 means you are "very willing to do so").

A.15 Relative Rank Information Provision

Based on your responses to the price-list question earlier, you've been selected to receive information about your rank on the midterm for ECN XX.

Your percentile rank on the midterm was [rank here].

This implies that YYY percent of students performed better than you and ZZZ performed worse than you.

A.16 Elicitation Ability Post

Now consider again your inherent "ability" to do well in ECN XXX.

What do you think is the probability you are in the top half of the ability distribution in this class? (answer should be between 0% and 100%, where 0% means you are certain that you are NOT in top half, while 100% means that you are certain that you are in the top half of the ability distribution).

The initial position of the slider corresponds to your previous answer.

A.17 Elicitation Hours Needed to Study to Achieve Desired Grade Post

Now that you've learned your rank on the course midterm, how many hours per week do believe you need to study for ECN XXX now to achieve the grade you selected in the question

asked previously?

A.18 End of Survey

Thank you for taking the survey! If you have any questions please contact Derek Rury at drury@ucdavis.edu.

B Robustness of Results

B.1 Adding Control Variables

Here we perform the same regressions as above but by including relevant observable characteristics as control variables to confirm that our results are not being driven by any observable characteristic. These include student's gender, mean prior rank belief, the specific course she is in, and the student's WTP/WTA. Moreover, for the beliefs about the number of hours required to achieve the desired grade we add the student's desired grade as control. Moreover, we run the same regressions by type of information received. That is, whether the student received "bad" or "good" news about her performance. In Table A1 we show the results. Results are consistent with the main results shown in the paper, highlighting the robustness of our findings. The only significant difference consists in the hours needed to achieve the students' desired grade for those that underestimated their performance. In fact, now the coefficient is smaller in magnitude and statistically not significant. This result is not surprising since it implies that those that performed better than expected do not believe they need to increase the number of hours of study.

Table A1

Treatment Effects - Including Control Variables

	(1) Hrs/Week Needed	(2) Prob(Top Half)	(3) Exam 2/Final Max
Average Treatment Effects			
Treat	1.140***	-5.496***	0.020
	(0.269)	(2.023)	(0.018)
Good News/Bad News			
Bad News (Treat)	2.106 ***	-14.276***	0.049
	(0.601)	(3.622)	(0.033)
Good News (Treat)	0.256	4.335***	-0.007
	(0.229)	(1.602)	(0.016)
Controls	✓	✓	✓

Notes: This table reports results from a series of regressions that study the outcomes listed in the column headers. Regressions include an indicator for treatment status and pre-outcomes variables, where appropriate. Hrs/week needed includes the number of hours students selected before treatment as a control. Probability students are in the top half includes the probability students selected before treatment as a control. The model studying exam two and final exam performance includes midterm one performance as a control. All models include the following control variables: student's gender, mean prior rank belief, the specific course she is in, and the student's WTP/WTA. We estimate robust standard errors. star(* 0.10 ** 0.05 *** 0.01)

B.2 Excluding Observations based on WTP/WTA

As mentioned in the paper, WTP/WTA measures are not orthogonal to the treatment status. Indeed, participants with higher (lower) WTP/WTA are more (less) likely to receive performance feedback. This implies that our results might be driven by these differences across conditions. To explore this possibility, we perform the same regressions by excluding those individuals with extreme WTP/WTA preferences. In Tables A2, A3, A4, and A5 we show the results for the average effects on both beliefs and performance. In Table A2 (A3) we include only those students with WTP/WTA preferences in the interval $[-3; +3]$ ($[-5; +5]$). The analysis allows us to have two comparable groups. In particular, these exclusions ensure that in both groups we have students that receive and do not receive performance feedback information. In particular, for A2 we also have two groups with identical average WTP/WTA preferences.[35] For this analysis, and by construction, assignment to the treatment group is as good as random. Finally, in Tables A4 and A5 we exclude either those with high WTP/WTA measures (above \$5.00) or low WTP/WTA (below \$-5.00).

All in all, these analyses show that the estimated coefficients are stable and the results are not driven by specific subgroups of students in terms of their WTP/WTA preferences. In other words, treatment effects are not driven by students' preferences for relative performance feedback.

[35] That is, there are no statistically significant differences in WTP/WTP across the experimental groups comprising of a total of 171 students $\delta = 0.077, p - value = 0.642$).

Table A2

Treatment Effects - WTP/WTA in the interval [-3;+3]

	(1) Hrs/Week Needed	(2) Prob(Top Half)	(3) Exam 2/Final Max
Average Treatment Effects			
Treat	1.302***	-5.326**	0.017
	(0.261)	(2.150)	(0.020)
Good News/Bad News			
Bad News (Treat)	2.696***	-10.351***	0.050
	(0.525)	(3.499)	(0.038)
Good News (Treat)	0.197	4.195**	-0.015
	(0.242)	(1.706)	(0.017)
Controls	✓	✓	✓

Notes: This table reports results from a series of regressions that study the outcomes listed in the column headers. Regressions include an indicator for treatment status and pre-outcomes variables, where appropriate. Hrs/week needed includes the number of hours students selected before treatment as a control. Probability students are in the top half includes the probability students selected before treatment as a control. The model studying exam two and final exam performance includes midterm one performance as a control. All models include the following control variables: student's gender, mean prior rank belief, the specific course she is in, and the student's WTP/WTA measure. We estimate robust standard errors. star(* 0.10 ** 0.05 *** 0.01)

Table A3

Treatment Effects - Excluding Relatively High and Low WTP/WTA

	(1) Hrs/Week Needed	(2) Prob(Top Half)	(3) Exam 2/Final Max
Average Treatment Effects			
Treat	1.323***	-5.097**	0.019
	(0.243)	(2.144)	(0.020)
Good News/Bad News			
Bad News (Treat)	2.694***	-11.194***	0.044
	(0.469)	(3.398)	(0.035)
Good News (Treat)	0.189	4.403**	-0.011
	(0.226)	(1.734)	(0.017)
Controls	✓	✓	✓

Notes: This table reports results from a series of regressions that study the outcomes listed in the column headers. Regressions include an indicator for treatment status and pre-outcomes variables, where appropriate. Hrs/week needed includes the number of hours students selected before treatment as a control. Probability students are in the top half includes the probability students selected before treatment as a control. The model studying exam two and final exam performance includes midterm one performance as a control. All models include the following control variables: student's gender, mean prior rank belief, the specific course she is in, and the student's WTP/WTA measure. We estimate robust standard errors. star(* 0.10 ** 0.05 *** 0.01)

Table A4

Treatment Effects - Excluding Relatively High WTP/WTA (above $5.00)

	(1) Hrs/Week Needed	(2) Prob(Top Half)	(3) Exam 2/Final Max
Average Treatment Effects			
Treat	1.348***	-5.159**	0.019
	(0.229)	(2.018)	(0.019)
Good News/Bad News			
Bad News (Treat)	2.606***	-11.864***	0.041
	(0.451)	(3.515)	(0.032)
Good News (Treat)	0.305	4.130**	-0.008
	(0.220)	(1.594)	(0.017)
Controls	✓	✓	✓

Notes: This table reports results from a series of regressions that study the outcomes listed in the column headers. Regressions include an indicator for treatment status and pre-outcomes variables, where appropriate. Hrs/week needed includes the number of hours students selected before treatment as a control. Probability students are in the top half includes the probability students selected before treatment as a control. The model studying exam two and final exam performance includes midterm one performance as a control. All models include the following control variables: student's gender, mean prior rank belief, the specific course she is in, and the student's WTP/WTA measure. We estimate robust standard errors. star(* 0.10 ** 0.05 *** 0.01)

Table A5

Treatment Effects - Excluding Relatively Low WTP/WTA (below -$5.00)

	(1) Hrs/Week Needed	(2) Prob(Top Half)	(3) Exam 2/Final Max
Average Treatment Effects			
Treat	1.143***	-5.172**	0.021
	(0.275)	(2.114)	(0.019)
Good News/Bad News			
Bad News (Treat)	2.305***	-14.102***	0.048
	(0.648)	(3.610)	(0.035)
Good News (Treat)	0.163	4.836***	-0.008
	(0.220)	(1.668)	(0.017)
Controls	✓	✓	✓

Notes: This table reports results from a series of regressions that study the outcomes listed in the column headers. Regressions include an indicator for treatment status and pre-outcomes variables, where appropriate. Hrs/week needed includes the number of hours students selected before treatment as a control. Probability students are in the top half includes the probability students selected before treatment as a control. The model studying exam two and final exam performance includes midterm one performance as a control. All models include the following control variables: student's gender, mean prior rank belief, the specific course she is in, and the student's WTP/WTA measure. We estimate robust standard errors. star(* 0.10 ** 0.05 *** 0.01)

B.3 Alternative Definition of Prior Belief and Rank

In the main analyses we have computed the prior belief as the first moment of prior belief distribution. Given that the belief elicitation asked students to report their prior beliefs in terms of deciles, we now compute prior belief as the *mean* prior decile belief (MPD). Coherently, the rank is now defined as the rank decile. Similarly, we compute good (bad) news as the difference between the MPD and the rank decile. A6 shows the results. These results highlight no significant difference compared to those in Table A1.

Table A6

Treatment Effects - Including Control Variables

	(1) Hrs/Week Needed	(2) Prob(Top Half)	(3) Exam 2/Final Max
Average Treatment Effects			
Treat	1.144***	-5.512***	0.020
	(0.271)	(2.033)	(0.018)
Good News/Bad News			
Bad News (Treat)	1.657**	-13.007***	0.050
	(0.801)	(4.272)	(0.040)
Good News (Treat)	0.068	5.102***	-0.012
	(0.283)	(1.735)	(0.021)
Controls	✓	✓	✓

Notes: This table reports results from a series of regressions that study the outcomes listed in the column headers. Regressions include an indicator for treatment status and pre-outcomes variables, where appropriate. Hrs/week needed includes the number of hours students selected before treatment as a control. Probability students are in the top half includes the probability students selected before treatment as a control. The model studying exam two and final exam performance includes midterm one performance as a control. All models include the following control variables: student's gender, mean prior rank belief, the specific course she is in, and the student's WTP/WTA measure. We estimate robust standard errors. star(* 0.10 ** 0.05 *** 0.01)

B.4 Robustness of WTP/WTA*Treat

We conduct further analyses to ensure that the treatment effect does not depend on students' WTP/WTA. Table A7 shows the results of three different analysis. The first one consists of adding control variables as robustness. The second on running the analysis depending on whether students received positive or negative news about their performance. The third model, building on the previous analysis, specifically focuses on students who received negative news from our treatment. We run this model to study whether students exhibiting information avoidance may deceive themselves to preserve their self-esteem upon receiving negative news. To run this analysis, we create an indicator that equals one if students' WTP/WTA is positive and zero otherwise. Overall, the three sets of results confirm that the usefulness of performance feedback is not mitigated by students' preferences for information, even when it is negative. One caveat to note is that since the performance feedback and the follow-up belief questions were presented in close proximity, the ability of students to exhibit motivated reasoning or self-deception was limited.

Table A7
Treatment Effects

	(1) Hrs/Week Needed	(2) Prob(Top Half)	(3) Exam 2/Final Max
Treatment Effect by WTP/WTA			
WTP/WTA*Treat	0.142	0.549	0.001
	(0.095)	(0.423)	(0.004)
Bad News/Good News			
Bad News (WTP/WTA*Treat)	0.293*	0.850	0.001
	(0.157)	(0.767)	(0.006)
Good News (WTP/WTA*Treat)	0.013	0.242	0.004
	(0.071)	(0.419)	(0.004)
Treatment Effect by WTP/WTA Restricted			
WTP/WTA*Treat (dummy)	0.858	-3.531	-0.023
	(1.488)	(10.261)	(0.058)
Controls	✓	✓	✓

Notes: This table reports results from a series of regressions that study the outcomes listed in the column headers. Regressions include an indicator for treatment status, WTP/WTA, their interaction, and pre-outcomes variables, where appropriate. Hrs/week needed includes the number of hours students selected before treatment as a control. Probability students are in the top half includes the probability students selected before treatment as a control. The model studying exam two and final exam performance includes midterm one performance as a control. The models with controls include the following variables: student's gender, mean prior rank belief, and the specific course she is in. We estimate robust standard errors. star(* 0.10 ** 0.05 *** 0.01)