ANNENBERG
BROWN UNIVERSITY

EdWorkingPaper No. 23-735

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

Alessandro Castagnetti
LIUC University

Derek Rury
University of Chicago

We administer a survey to study students' preferences for relative performance feedback in an introductory economics class. To do so, we elicit students' willingness to pay for/avoid learning their rank on a midterm exam. Our results show that 10% of students are willing to pay to avoid learning their rank. We also find that female students are willing to pay $1 more than male students. We also confirm that beliefs about academic performance do not predict preferences for information. After randomizing access to information about rank, students report needing more study hours to achieve their desired grade and being less likely in the top half of the ability distribution in the class. These effects are driven by stronger effects from people who overestimated their midterm rank compared to those who underestimated their performance. We do not find an overall effect of learning about rank performance on final course grade. We also confirm that students' preferences for feedback do not interfere with their belief updating.

VERSION: February 2023

Suggested citation: Castagnetti, Alessandro, and Derek Rury. (2023). Student Demand For Relative Performance Feedback: Evidence from a Field Experiment. (EdWorkingPaper: 23-735). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/7dg2-4d23

Student Demand For Information About Relative Exam Performance: Evidence from a Field Experiment

Alessandro Castagnetti (LIUC University)*

Derek Rury (University of Chicago)†

This version: February 14, 2023

Abstract

We administer a survey to study students' preferences for performance feedback in an introductory economics class. To do so, we elicit students' willingness to pay for/avoid learning their rank on a midterm exam. Our results show that 10% of students are willing to pay to avoid learning their rank and that female students are willing to pay \$1 more than male students. We also confirm that beliefs about academic performance do not predict preferences for information. After randomizing access to information about rank, students are more (less) likely to report needing more hours of study time per week to achieve their desired grade and are more likely to believe they are below (above) average in ability in the class. We do not find an overall effect of learning about rank performance on final course grade. We also confirm that students' preferences for feedback do not interfere with their belief updating.

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

1 Introduction

Economics predicts that people have a positive demand for information, particularly information that is beneficial for future decision making. One important type of information is performance feedback. In education, performance feedback is regularly provided to students across all grades in the form of assignments and exams. This feedback forms the basis for how students learn about their ability and make decisions regarding their academic future. To that end, performance feedback information in education regularly carries what is known as an *instrumental value*; a feature of information that helps students make decisions about their human capital under uncertainty. Research in behavioral economics also shows that this information may also have an *intrinsic value*; a feature of information that is relevant independent of how useful it is for future decision making (Bénabou and Tirole, 2002). This is especially true for information that reveals one's type, such as information on attractiveness (Eil and Rao, 2011), relative effort (Azmat and Iriberri, 2015) or intelligence (Möbius et al., 2022; Zimmerman, 2020).

In this paper, we argue that performance feedback in education potentially carries this intrinsic value of information. Furthermore, while the instrumental value of performance feedback carries a positive effect on demand for information (as it only helps the actor make better decisions), the intrinsic value of performance feedback may carry either a positive *or* negative effect on demand. Because of these potentially opposing effects, we do not know whether students' demand for performance feedback information is either positive or negative. Moreover, it is unclear whether students' desires to avoid information about themselves can override the beneficial features of performance feedback.

To address this, we administered a survey at a public university in the US designed to elicit students' demand 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 an exam. We find that students are willing to pay around \$1.00 on average to learn their rank, primarily driven by nearly 43% of students that have a positive willingness to pay for knowing their actual rank on the exam. This confirms the utility of performance feedback, as a large proportion of students value this information. However, 10% of individuals have a negative willingness to pay, implying they are willing to pay to *avoid learning* their rank. 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 performance on the midterm. This contradicts other work studying preferences for information about oneself.[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 preferences for feedback cannot be explained by performance on the exam or students' beliefs. 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 with students their actual performance rank on the exam. We find that students report requiring more hours of study time and being less likely

[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 which may help counteract that they tend to hold low beliefs about their ability or performance in different tasks (Bordalo et al., 2019; Coffman, 2014). 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., 2022).

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.

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 students' beliefs. Looking at beliefs about performance, we find that students are typically inaccurate when guessing their rank, with a SD from the true rank of nearly 20 rank points. 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 good (bad) news reported needing fewer (more) study hours to achieve their desired grade and were more (less) likely to be above average ability-wise in their class.

Lastly, we show that students' beliefs about performance and their preference for information do not interfere with the 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). When studying whether students' WTP/WTA for information influences our results, we fail to find any evidence that students' preferences influence our treatment effects.

This work contributes to three different strands of literatures. 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 more broadly may lead individuals 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 people avoid ego-relevant performance feedback (Castagnetti and Schmacker, 2022; Eil and Rao,

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

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 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 analysed how achievement influences learning about one's overall performance as well as performance relative to a group of peers (Li, 2018). Most studies find positive effects on achievement for all students (Azmat and Iriberri, 2010; Bandiera et al., 2015; Brade et al., 2018) 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

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

[5]Another paper looking at high school students in Mexico find 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).

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.

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 show the experimental results; section 6 discusses our results and how our results fit within the literature on demand for information and performance feedback; section 7 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 four in the fall and two in the winter. All classes in the experiment were introductory economics courses, which 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

[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 home works, lecture videos and attendance or attention credits. One class determined grades via a single midterm and a final exam.

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. The survey was coded using the software Qualtrics and took about 10 minutes to complete. 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 the 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 scores as being a percentile. This ordering puts scores within a distribution between 1 and 100, placing those who scored higher on the midterm to be ranked lower (e.g. highest scoring student would be at the one percent level).[7] 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 decile, representing the likelihood their performance would be in each decile. This gives us a prior distribution of beliefs about midterm performance.

Participants then moved onto 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

[7] We selected this arrangement of percentiles to not 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.

willing to pay \$10 to avoid learning their own rank to 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 WTA/WTP) 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. The point where students switched between the two represents their WTP/WTA for performance feedback. A screenshot of the The WTP/WTA elicitation is found in Figure 1.[8]

Figure 1: WTP/WTA Elicitation Procedure

	A	B	
1. **Receiving** \$10.00 and learning your true rank.	○	○	**Receiving** \$0 and not learning your true rank.
2. **Receiving** \$9.00 and learning your true rank.	○	○	**Receiving** \$0 and not learning your true rank.
3. **Receiving** \$8.00 and learning your true rank.	○	○	**Receiving** \$0 and not learning your true rank.
4. **Receiving** \$7.00 and learning your true rank.	○	○	**Receiving** \$0 and not learning your true rank.
5. **Receiving** \$6.00 and learning your true rank.	○	○	**Receiving** \$0 and not learning your true rank.
6. **Receiving** \$5.00 and learning your true rank.	○	○	**Receiving** \$0 and not learning your true rank.
7. **Receiving** \$4.00 and learning your true rank.	○	○	**Receiving** \$0 and not learning your true rank.
8. **Receiving** \$3.00 and learning your true rank.	○	○	**Receiving** \$0 and not learning your true rank.
9. **Receiving** \$2.00 and learning your true rank.	○	○	**Receiving** \$0 and not learning your true rank.
10. **Receiving** \$1.00 and learning your true rank.	○	○	**Receiving** \$0 and not learning your true rank.
11. **Receiving** \$0.00 and learning your true rank.	○	○	**Receiving** \$0 and not learning your true rank.
12. **Paying** \$1.00 and learning your true rank.	○	○	**Paying** \$0 and not learning your true rank.
13. **Paying** \$2.00 and learning your true rank.	○	○	**Paying** \$0 and not learning your true rank.
14. **Paying** \$3.00 and learning your true rank.	○	○	**Paying** \$0 and not learning your true rank.
15. **Paying** \$4.00 and learning your true rank.	○	○	**Paying** \$0 and not learning your true rank.
16. **Paying** \$5.00 and learning your true rank.	○	○	**Paying** \$0 and not learning your true rank.
17. **Paying** \$6.00 and learning your true rank.	○	○	**Paying** \$0 and not learning your true rank.
18. **Paying** \$7.00 and learning your true rank.	○	○	**Paying** \$0 and not learning your true rank.
19. **Paying** \$8.00 and learning your true rank.	○	○	**Paying** \$0 and not learning your true rank.
20. **Paying** \$9.00 and learning your true rank.	○	○	**Paying** \$0 and not learning your true rank.
21. **Paying** \$10.00 and learning your true rank.	○	○	**Paying** \$0 and not learning your true rank.

[8] The full description of the WTP/WTA elicitation procedure, along with the entire survey, can be found in Appendix XXX.

2.3 Subsequent Survey Questions

Students were then asked what grade (e.g. A+ to A-, B+ to B-, etc.) they expected to get 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 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: Performance Feedback

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 WTA/WTP question would be binding for that student. 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 similar a similar WTP are randomly assigned to be in the treatment group (those that receive relative feedback information) or in the control group (those that do not receive relative feedback information).[9]

Students assigned to the treatment group were provided with their actual midterm one percentile rank. This information ranged between 1-100 with 1 being the top percentile and 100 being the lowest. We were able to link students to their actual midterm rank using students university ID numbers collected as part of the survey.[10,11] A students rank on midterm one provides useful information as course grades are often curved. 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 worst, where $X = rank - 1$ and $Y = 100 - rank$. An example of the information treatment can be found in Figure 2. Students randomized into the control group received no information about their rank on the midterm.

Figure 2: Information Treatment

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 100.

Your rank on the midterm was 30.

This implies that 29 percent of students performed better than you and 70 performed worse than you.

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

[9] The reason this procedure is not fully random is that students with extreme preferences will either never or will always receive performance feedback. In the analysis we will take this into count and run appropriate robustness checks.

[10] 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.

[11] 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.

ability distribution.[12]

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.[13]

Secondly, students could be selected to have their WTP/WTA measure determine their payoff. Under this scenario, students would be alotted \$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 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.

[12]Students in the control group were not asked 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.

[13]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.

3 Data and Descriptive Results

To study students' beliefs about performance, WTA/WTP 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. There were a total of 1,429 students enrolled in all five 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.[14]

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 a 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 impacted performance, we analyse performance on either midterm two or the final exam. As several classes contained a policy of allowing students to drop their lower exam score, we take the maximum of the two exams to measure the effect on class performance.

We next study how well our randomization procedure balance pre-treatment variables across treatment groups. Table 1 displays results that studies a regression of all of our pre-treatment variables on treatment status. Only WTP/WTA is statistically significant (p <

[14] 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.

0.01). This is not surprising given that whether students were randomized to treatment or control depended to some extent on their WTP measure. A joint significance test rejects the hypothesis that variables were unbalanced between treatment groups. We therefore conclude that our treatment was administered successfully.

Table 2 shows descriptive statistics for our study participants. 57% of students in our sample are female.[15] 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 they believed to be in the top half of the ability distribution, 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.[16] A quick analysis studying how our sample compares to the 1,429 students 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 or on students' potential outcomes of midterm one performance feedback.

We next study how well students know where their actual performance fell within the

[15] 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%.

[16] Mean and standard deviation measures vary between the six classes. The total possible points for each midterm were not standardized across the classes. To make scores comparable, we divided the number of points students received by the number of points possible for each exam. This makes all student scores range between 0 and 1, with each score presenting a percentage of points they received.

midterm one distribution. As mentioned above, information about the mean and variance of midterm one performance were not communicated with students via the professor or Canvas.[17] 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 each performance decile on midterm one. Figure 1 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.[18]

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, the resulting mean prior belief would be $[1 * (.20) + 5 * (.30) + 10 * (.50)] * 10$ which in this case would equal 67.[19] Creating this for the entire sample, we observe a mean prior belief of 39.8, which is similar to the 42nd percentile of actual performance.

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

[17]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. Second, knowing what proportion of students performed better/worse than them is likely to help them calibrate how much to study or whether to complete the class.

[18]There are non-monotonicaly changes for decile X and Y, but they are only slight deviations from deciles on above or below these deciles.

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

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, 2018). 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 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.[20]

It is important to note that up until the time students took midterm one, students had very few chances to learn about their relative performance in economic courses.[21] 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 overestiamted (underestimated) would report needing more (less) time studying 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

[20]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.

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

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 due to the instrumental value of relative performance feedback. For example, relative performance feedback may reveal information about their returns to study effort and therefore help students calibrate how much to study (Rury and Carrell, 2022). 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 instrument 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. Furthermore, we do not know whether students use different tactics (ie. self-deception) to avoid processing this information. 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 2 presents the results graphically. First, the mean WTP/WTA is of \$1.02. Students, therefore, have a positive WTP 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.[22] Third, the distribution of WTP/WTA is not symmetric. Almost 43% of students with non-zero WTP/WTA are willing to spend money to learn their rank. On average, they are willing to pay \$4.30. However, and importantly, we also find that a significant fraction of students are indeed willing to pay to avoid relative performance feedback. 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' WTA/WTP correlates with *beliefs* about their midterm one performance. It is important to underline here that theoretical work make no clear cut predictions on the relationship between beliefs and the demand of performance feedback. For example, Köszegi (2006) predicts that individuals with higher beliefs will seek less information 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 makes 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 4. Here,

[22] 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.

we find no statistically significant correlation between having a higher mean prior belief and a higher WTP. This finding contradicts other work in the ego-utility literature (cites).

Our results therefore show on aggregate no straightforward correlation between prior beliefs and demand for information. 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 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 4 also shows that there is no statistically significant relationship.

Lastly, we study how students gender influence preferences for feedback. We find that being a female is associated with a \$1 increase in students' WTP measure. This result cannot be explained either by differences in rank nor 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 women being 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 your midterm one rank, P_i contains pre-treatment variables, when appropriate, and ϵ_i is a random disturbance term.[23]

Table X 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 ($p < 0.01$). To examine the effect on acheivement, we study the max score between midterm two and the final exam.[24] We see in table X that treatment has a null effect on exam performance.

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

[23] We estimate robust standard errors.

[24] There was one class which did not have a midterm two. In this case, we only study the effect on the final exam.

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 ($p < 0.01$) compared to those who underestimated their performance who report needing 0.46 more hours ($p < 0.05$). 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 over being 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, just as in the overall sample, we fail to detect effects for over/under estimators.

5.3 Results by WTP/WTA

Our results studying students’ WTP/WTA show that 10% of students would pay to avoid learning their midterm rank. One concern is that students with negative demand for information might process performance feedback information differently than those with positive demand. For example, students exhibiting information avoidance may deceive themselves to preserve their self-esteem upon receiving negative news. Table 4 shows results that study whether treatment effects depend on students’ WTP/WTA measures. Here we restrict our analysis to students who received negative news from our treatment. We also create an indicator that equals one if students’ WTP/WTA is negative and zero otherwise. In order for these estimates to demonstrate a self-serving information processing bias, we would expect treatment effects from the interaction between WTP/WTA and treatment to be positive for both of our belief questions. This would indicate that those exhibiting information avoidance are ignoring the information about their rank performance and failing to update their beliefs. Here we see that treatment effects are negative, yet imprecise for both belief questions. Results on performance are strongly null. The 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.

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 exhibiting motivated reasoning. We also show that female students are willing to pay $1 more to learn their rank than male students.

We also find that students hold inaccurate beliefs about their rank on the midterm exam, highlighting the instrumental value of providing relative performance feedback to students. We find support for this when studying how providing information about midterm rank affects student beliefs. We find that upon learning their rank, students update beliefs about the number of hours they need to study per week as well as how likely they are in the top half of the class ability distribution. More importantly, the effects of the information disclosure depends 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 report needing less study hours to achieve their desired grade and being more likely to be above average. We also fail to detect any effects on achievement.

Results from this experiment help us confirm two further important facts about performance feedback in classes. First, beliefs about performances do not predict preferences for information on feedback. This contradicts several papers in the experimental literature studying motivated reasoning. Second, 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 making decisions about their human capital.

Our results highlight that students do not hold uniform preferences over learning about their relative performance. In light of this, one policy implication might be that rank information is made available on request only to those students who want to learn their rank. In this way, students who do not want to receive it are not forced to receive it. On the other hand, future work should address whether sharing this information to those students that attach a negative value to performance feedback is ultimately beneficial or not. 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 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. We expect this negative preference for information to be exacerbated within populations with a large fraction of struggling students. 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 which come with performance feedback such as taking the SAT or ACT, which serve as important proxies for going to college.

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. 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 (2022): *When to Apply?*

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.

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.

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 (2018): "Knowing Who You Are; The Effect of Feedback on Short and Long-Term Outcomes," *Working Paper*.

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.

KÖSZEGI, 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*.

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," *Annu. Rev. Econ.*, 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

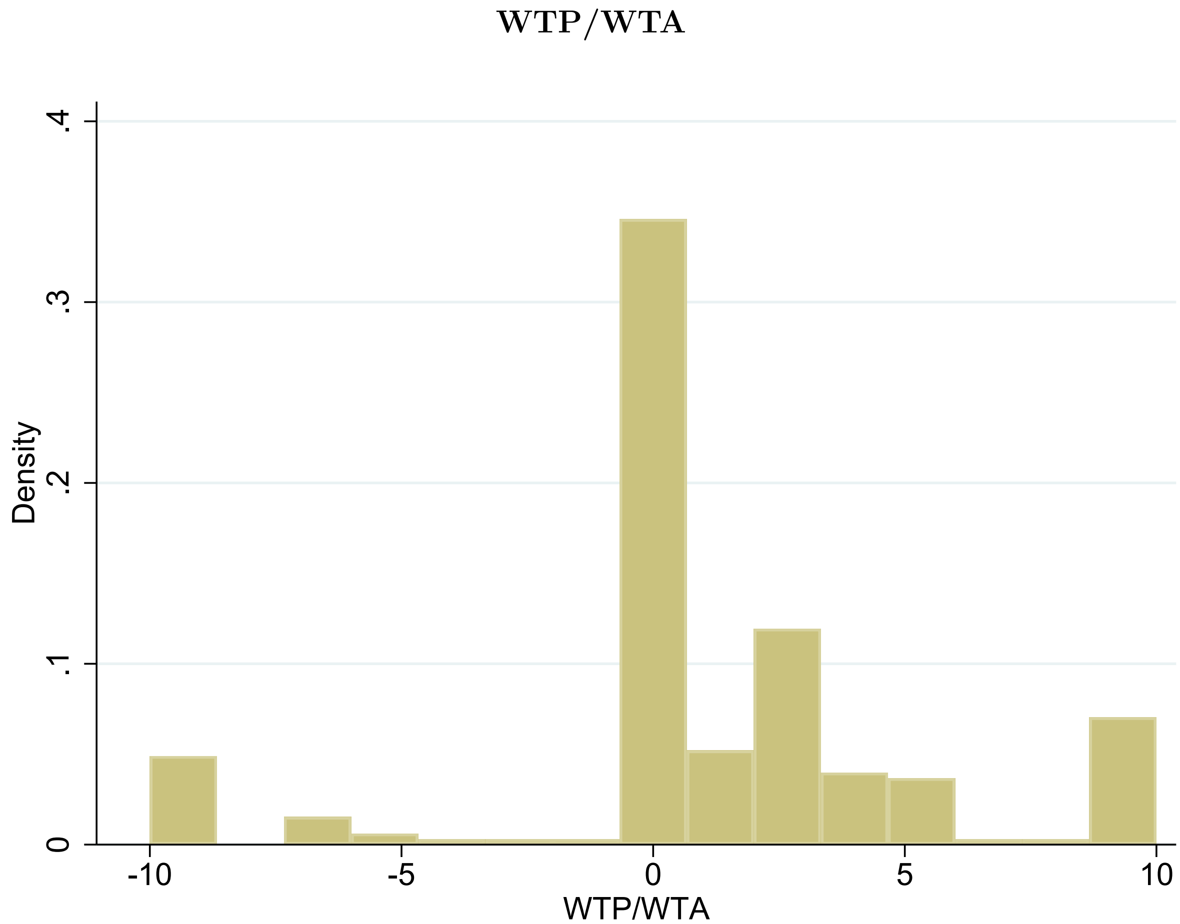

Figure 5
Willingness to Pay/Avoid for Midterm 1 Rank

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

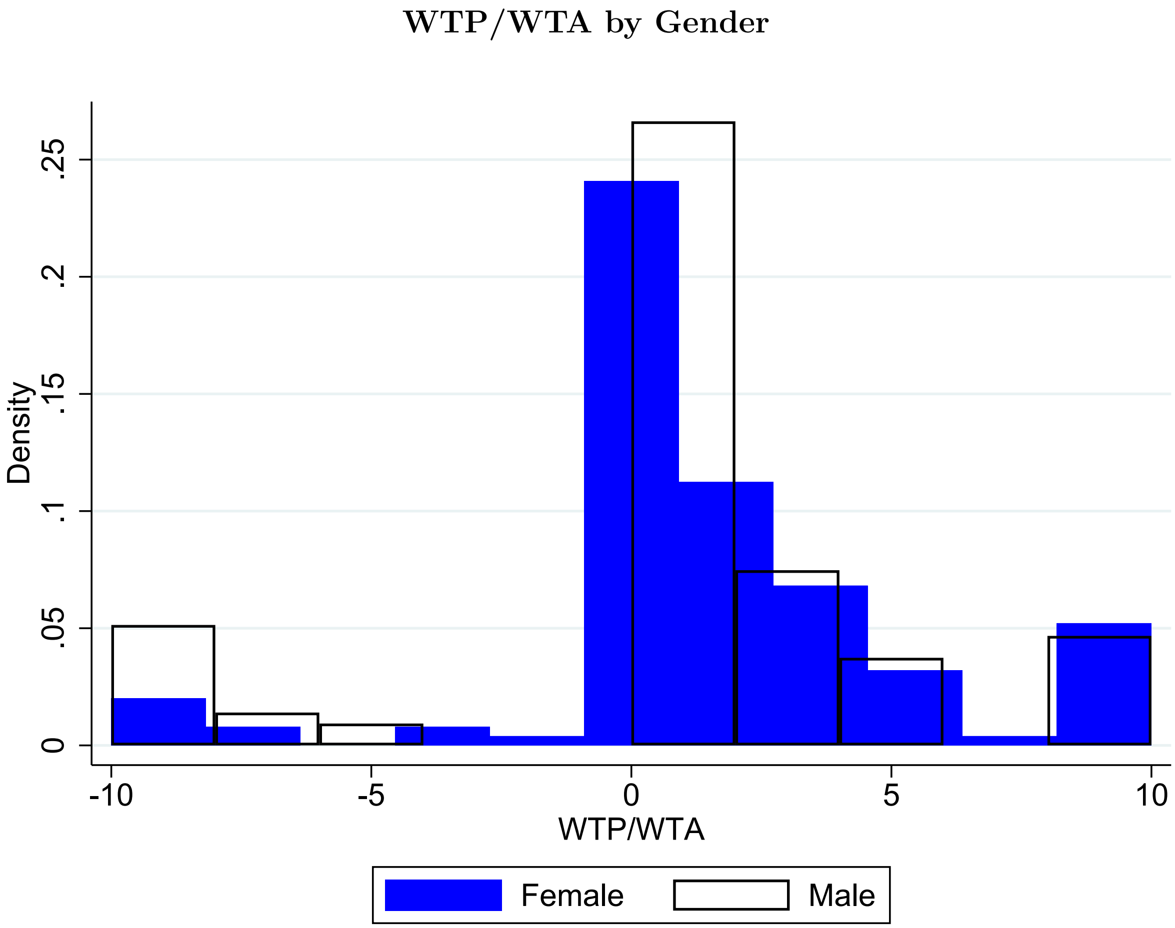

Figure 5
Willingness to Pay/Avoid for Midterm 1 Rank

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

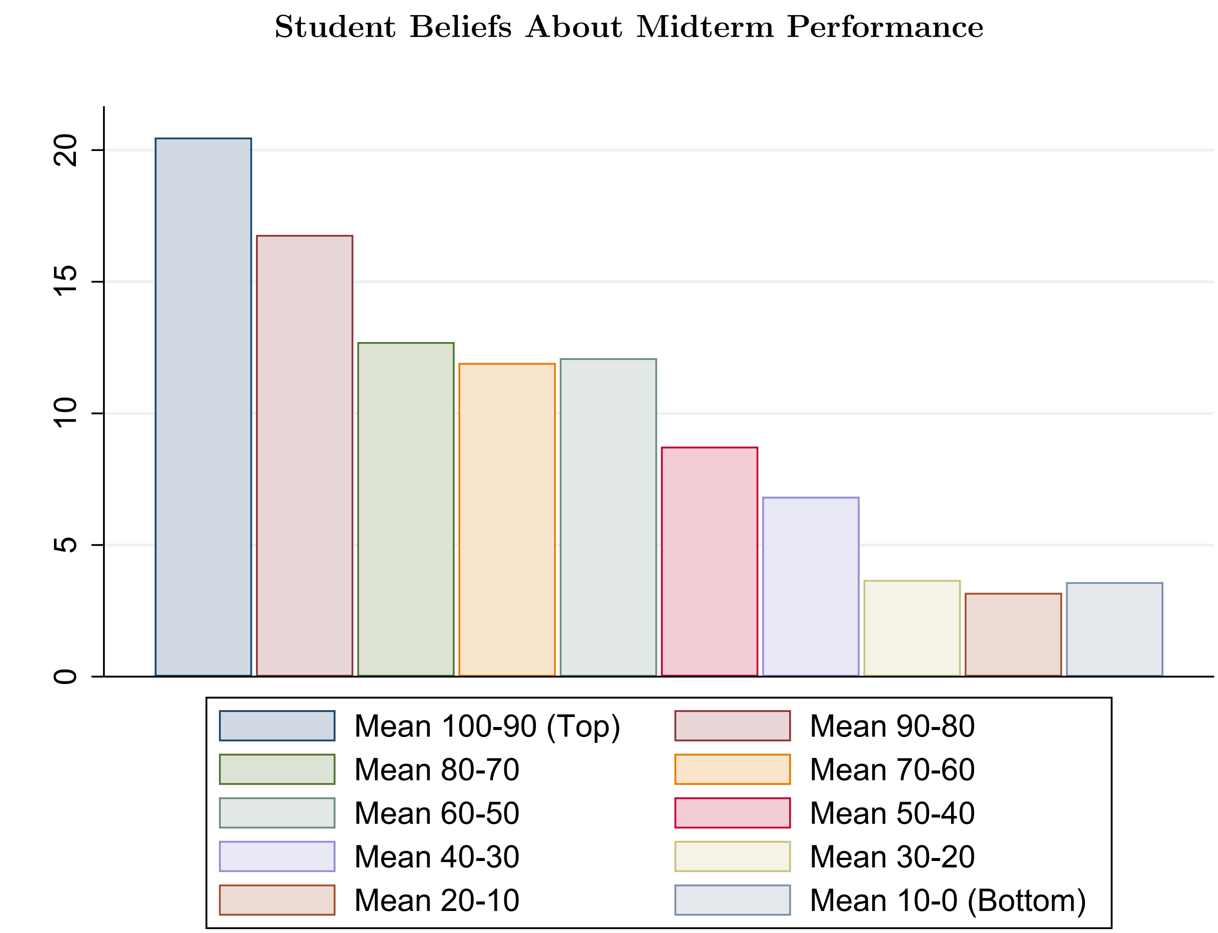

Figure 5
Student Beliefs About Midterm Performance

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.

Table 1
Balance Tests

	(1) Treat
WTP/WTA	0.045***
	(0.007)
Midterm 1 Beliefs	-0.002
	(0.002)
A grade	-0.140
	(0.171)
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: The three columns in this table represent three different samples; (1) is the students in the 8th grade sample; (2) represents those in the HS sample; and (3) represent those in the SAT/ACT outcomes sample. Each column represent the average for the respective variable in that sample. These variables constitute those used in my main estimation sample for each set of outcomes. While the number of observations does not equal the bottom row for each sample, it offers a good approximation of the data used in the sample. star(* 0.10 ** 0.05 *** 0.01)

Table 2
Descriptive Statistics

	Count	Mean	SD	Min	Max
Female	253	.5612648	.497216	0	1
Midterm 1 Rank Beliefs	247	39.16682	23.37458	10	100
Midterm 1 Rank	253	40.92885	29.87067	1	100
Midterm 1 Performance	242	.7893471	.1632486	.24	1
A grade	244	.6721311	.4704017	0	1
B grade	244	.2786885	.4492755	0	1
C grade	244	.0491803	.2166889	0	1
Hrs/Week Needed	242	8.960744	6.718519	0	40
Prob(Top Half)	244	75.42623	25.54828	0	100
WTP/WTA	244	1	4.307138	-10	10

Notes: The three columns in this table represent three different samples; (1) is the students in the 8th grade sample; (2) represents those in the HS sample; and (3) represent those in the SAT/ACT outcomes sample. Each column represent the average for the respective variable in that sample. These variables constitute those used in my main estimation sample for each set of outcomes. While the number of observations does not equal the bottom row for each sample, it offers a good approximation of the data used in the sample. star(* 0.10 ** 0.05 *** 0.01)

Table 3
WTP/WTA Mediators

	(1) WTP/WTA	(2) WTP/WTA
Mean Prior Belief	-0.010	-0.015
	(0.019)	(0.019)
Average Performance	-3.874	-4.291
	(2.834)	(2.819)
Female		1.186**
		(0.552)

Notes: This table reports results from two separate regressions that study how beliefs about midterm one performance and actual midterm 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 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) Exam 2/Final Max
Average Treatment Effects			
Treat	1.496***	-4.872**	-0.005
	(5.92)	(-2.76)	(-0.42)
Good News/Bad News			
Bad News (Treat)	2.861***	-13.751***	0.008
	(0.517)	(2.981)	(0.014)
Good News (Treat)	0.458**	3.971***	-0.010
	(0.192)	(1.456)	(0.014)
WTP/WTA			
WTP/WTA*Treat	0.130	0.631	-0.001
	(1.32)	(1.57)	(-0.43)

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. The model studying whether students can remember their rank on midterm one includes an indicator for whether they received "bad" or negative news from the treatment. Heterogeneous treatment effect models also include a variable that captures the effect of that variable on the outcome. We estimate robust standard errors. star(* 0.10 ** 0.05 *** 0.01)