The Value of a Signal: Information Processing among Students Outside the Lab

Catalina Franco*

January 29, 2026

Abstract

Correctly processing informational signals and assessing relative performance are key for decision-making in education. While laboratory studies have found that individuals process information in biased ways, little is known about how students update beliefs in real-life education settings. Leveraging elicitation methods from lab studies, this paper provides the first evidence of information processing in a high-stakes, real-life education situation: students preparing for a college entrance exam. I study belief updating in response to information and assess the value of a relative performance signal vis-à-vis absolute scores, the mainstream form of information in education. I provide three main findings. First, students are conservative and update about 70% of the Bayesian benchmark implied by the relative performance signal. Second, the relative performance signal is more valuable for students whose absolute score is in the middle of the distribution, where it is harder to assess whether one's score is above or below the median. Third, larger belief updates induced by receiving the signal foster higher confidence in gaining university admission, in particular for female students, suggesting that belief updating can help build non-cognitive skills that are invaluable for student success.

JEL classification: C91, C93, D83, D91, I21, I25, N36

Keywords: belief updating, informational signals, education, non-cognitive skills, external validity

*Centre for Applied Research (SNF) and FAIR, NHH Norwegian School of Economics, catalina.franco@snf.no. Helleveien 30, 5045 Bergen, Norway. I am grateful to Tanya Rosenblat, Hoyt Bleakley, Dean Yang, and Martha Bailey for advice and support over the course of this project. I thank Erik Sørensen, Alexander Cappelen, Kjetil Bjorvatn, Bertil Tungodden, Molly Hawkins, Gonzalo Vázquez-Baré), Nicolás Morales, Francesco Bogliacino, César Mantilla, Leonardo Garzón, and participants at the Rosario Experimental Economics Conference for useful feedback. Funding for this project was generously provided by the Michigan Institute for Teaching and Research in Economics (MITRE) and a Rackham Graduate Student Research Grant. This project received approval from the University of Michigan IRB (HUM00124049).

1 Introduction

Forming unbiased beliefs and correctly incorporating information about one's academic ability are critical inputs for educational decisions and academic success. The traditional economic model predicts that students update beliefs about their ability in an unbiased way. However, laboratory evidence shows that people typically deviate from the Bayesian model of rational updating. The growing literature on motivated beliefs and reasoning highlights that individuals may choose to hold certain beliefs (e.g., overconfidence) that trade off accuracy and desirability (Bénabou & Tirole, 2016; Zimmermann, 2020). The literature on information processing about own ability finds that lab subjects behave as conservative Bayesians when updating beliefs (Möbius et al., 2022; Buser et al., 2018; Coutts, 2019), and that the updating differs when the information is ego-relevant (Ertac, 2011; Grossman & Owens, 2012), or provides good rather than bad news (Eil & Rao, 2011).

The current evidence on information processing relies on laboratory experiments conducted on convenience subject samples.[1] However, little is known about how individuals outside the lab update beliefs about own ability in tasks that have real-life rewards, and that may have large and long-term consequences. To my knowledge, this study is the first to provide such evidence in the important context of education. I elicit beliefs about relative performance and study belief updating among students preparing for a high-stakes exam. Relative performance is a key determinant of many educational processes. The norm, however, is to provide absolute scores to inform students about their performance in primary and secondary education. By comparing with absolute scores, I assess how students' updating responds to a relative performance signal and provide evidence on whose updating benefits the most from that signal.

Students in my sample are enrolled in a preparation course for a college entrance exam to a prestigious public university in Colombia. Relative performance is all that matters

[1]Specifically, white, educated, industrialized, rich and democratic (WEIRD) samples of university students (Henrich, Heine, & Norenzayan, 2010).

for admission at this university as students' scores are ranked, and college slots are solely assigned based on this ranking. All students in my sample take a series of practice tests and receive information about their absolute scores in math and reading. I randomly vary who receives or does not receive a relative performance signal indicating whether one's score in each practice test is above or below the median.[2]

My first research question is to what extent students' information processing corresponds to the Bayesian benchmark. I elicit prior and posterior beliefs about the probabilities that students' scores in math and reading lie in the different quartiles of a distribution containing over 1,000 students taking the same practice test. The first main result is that students update conservatively. Students' updating is about 70% of what a Bayesian agent with the same priors would update. Furthermore, I do not find that updating differs statistically between when the scores lie above and lie below the median (i.e., no asymmetry) or by test subject.

My second research question seeks to assess the value of the relative performance signal vis-à-vis absolute scores. The comparison with the Bayesian benchmark is an average across all students and is identified off changes in the probability allocation from the prior to the posterior stage. If students do not assign prior probability to the section of the distribution where their score lies, it is not possible to apply the Bayesian model. To address this, I study the fraction of probability allocated to above-median quartiles in the posterior relative to the prior stage. This measure of updating allows the inclusion of all students in the sample and increases in statistical power to conduct heterogeneity analyses. I refer to differences in updating when receiving the relative performance signal on top of absolute scores as the *value* of the relative performance signal.

[2] Even though a signal such as above or below the median is too coarse to infer eligibility for a college slot, it still conveys useful information, that is, the students can learn what section of the test they need to work more on, or that they are unlikely to obtain a college slot if their performance in practice tests is below the median. The decision of providing an above/below-median signal mirrors the practice in the lab to make the elicitation mechanism tractable: lab subjects are asked to guess whether their score in a quiz is above or below the median, they are given noisy signals of their true performance, and are incentivized to report truthfully based on an elicitation mechanism.

The second main finding is that the value of the relative performance signal is about 11 pp in math and 16 pp in reading. This means that there is a statistically significant effect of receiving the signal on allocating probability to above-median quartiles. Even without the signal, students with above-median scores can infer where their score lies in the distribution as I observe them assigning more probability to above-median quartiles than students with scores below the median. In contrast, the probability allocation of below-median students who receive and do not receive the signal does not differ, suggesting that relative performance signals have value over and above what students learn from absolute scores only when the information conveyed by the signal is positive (i.e., asymmetric updating).

The value of the signal is highest for students with scores near the middle of the distribution, that is, where uncertainty about whether one's score is above or below the median is highest. This is consistent with the hypothesis that students who obtain very high or very low scores do not learn anything from the signal given that the absolute score already tells them enough about whether they are above or below the median. In addition, I provide suggestive evidence that this effect seems to be driven by female students as they tend to be more responsive to the relative performance signal than are male students. Because females' larger responses are related to starting with a lower allocation of probability to above-median quartiles, my results are consistent with previous findings indicating that women are usually found to be less self-confident than men (Niederle & Vesterlund, 2007), and indicate that the signal helps women better assess their relative performance. Furthermore, higher levels of confidence within the updating task among females correlate with a measure of confidence outside the experiment. Females who receive the relative performance signal and make large updates in the previous practice test are 20 pp more likely to report higher confidence in gaining admission to the university to which they are applying.

This paper contributes to the literature on information processing in experimental economics, specifically on updating in ego-relevant tasks (Ertac, 2011; Eil & Rao, 2011; Grossman & Owens, 2012; Möbius et al., 2022; Berlin & Dargnies, 2016; Buser et al., 2018; Coutts, 2019). Building on this literature, I provide a unique setting in which updating can be observed outside the lab in a real-stakes situation. While some control is relinquished in this setting, this context is fruitful for analyzing information processing for three reasons. First, students care about their performance in practice tests because of the high-stakes nature of the exam for they are preparing. Many of these students will not be able to attend college at all if they do not gain admission to this university. In this sense, performance information from practice tests allows them to assess their preparation progress. Second, it is possible to obtain an objective measure of relative performance that can be used as a relative performance signal. Third, the setting lends itself to allowing the use of elicitation tools and incentives as in the lab.

My paper also complements the literature in education economics investigating how students form and update beliefs about academic performance (Zafar, 2011). Part of this literature focuses on studying how overoptimistic views of abilities at college entrance are associated with academic decisions such as dropping out from college (Stinebrickner & Stinebrickner, 2008). Another part focuses on how interventions eliciting beliefs and providing feedback affect students' academic choices (Bobba & Frisancho, 2022; Gonzalez, 2017; Franco, 2020). To my knowledge, no study has focused on *how* students update beliefs. Even though the state of the art in belief elicitation tools and models for updating has been developed in lab settings, there is a disconnect between those advances and what research in education has implemented in the field. My study aims to bridge that gap and represents a novel effort in assessing the external validity of well-studied behaviors emerging from lab settings (e.g., Levitt & List, 2008; Camerer, 2011; Kessler & Vesterlund, 2015).

2 Experimental design and belief elicitation

2.1 Sample

The sample consists of students enrolled in a preparation course at a private institution that prepares students for standardized exams such as college entrance exams to public universities in Colombia.[3] All students in the sample are preparing for the college entrance exam to Universidad de Antioquia, the largest public university in the region of Antioquia, Colombia. During the first week of classes, students provide consent to participate in a research study in which they could earn cash prizes by filling out weekly surveys. As part of the course, students take a weekly practice test in the subjects covered by the college entrance exam: reading and math. Each practice test and the exam take three hours to complete. The goal of the course in providing weekly practice tests is to familiarize the students with the questions and the time constraints in the exam. In total, students take 11 practice tests during the course. The sample contains 862 observations from 369 students in math, and 904 observations from 386 students in reading.

2.2 Design

To study information processing, I elicit beliefs about the probability of being in each of the four quartiles of the math and reading distributions across eight practice tests. The quartiles are calculated based on the scores of over 1,000 students taking the same practice test as students in the sample. At the start of the course and for its entire duration, one group of participants (control) is randomly assigned to obtain absolute scores only. The other group (treatment) receives, in addition, a signal indicating whether their score is above or below the median. The signal is truthful as in Ertac (2011), Eil and Rao (2011) and Berlin and

[3] The sample is the same as in Franco (2020), in which I analyze the effects of receiving relative performance feedback on academic decisions. Franco (2020) shows that students at this institute are positively selected relative to the global student population in the region and in Colombia.

Dargnies (2016).[4] [5] Appendix Table B.1 presents descriptive statistics and balance tests for students who complete the survey at least once. Appendix Table B.2 reports regression results of a variable indicating whether the student appears in the sample once vs. more times and up to two times vs. more times on baseline characteristics. There are no systematic differences across treatment and control groups or across the number of times a student has participated in the surveys.

2.3 Belief elicitation and incentives

I use a crossover mechanism similar to Möbius et al. (2022) and Berlin and Dargnies (2016). For each practice test, I elicit the probabilities of being in each quartile of the math and reading score distributions.[6] Instead of eliciting probabilities directly, I assign 12 imaginary tokens for each academic subject to distribute among the four quartiles of each score distribution.[7] The framing of this task is a game called the "Quartiles Game" where the students had the chance to win cash prizes by truthfully revealing their beliefs. Prior beliefs were elicited immediately after students took each practice test and were collected either online if the practice test was online, or on paper if the practice test was in person. Posteriors were always elicited online at the same time the students checked their practice test performance report.[8] All students received online training on the definition of quartiles and how the Quartiles Game works. In case of questions, they could contact the researcher by email or phone.

One of the two belief elicitations during a given week was chosen at random to be entered in a raffle (see experimental protocol in Appendix C). After submitting posteriors, students

[4] Other studies provide a signal that is true with some probability (Grossman & Owens, 2012; Möbius et al., 2022; Buser et al., 2018; Coutts, 2019).

[5] After submitting posteriors, students in the treatment group learn their quartile in each subject (see Franco, 2020).

[6] Möbius et al. (2022) and most papers in this literature elicit the probability of being in the top half of the distribution. Histograms of the score distributions are in Appendix Figure A.1.

[7] The reason for this choice is that the students are very young (average age is 17.5) and not all may be familiar with computing probabilities.

[8] Appendix Figure A.2 illustrates when priors and posteriors about relative performance in practice tests are elicited.

were guided through instructions to throw a 12-sided dice that would determine whether they receive zero or a positive amount of cash. Let y be the random draw from the dice and x the number of tokens assigned to the quartile to which the student's score belongs.[9] The specific procedure to determine prizes is similar to Berlin and Dargnies (2016):

1. If $y \leq x$ the student wins COP 20,000 (US\$7).

2. If $y > x$, the student wins COP 20,000 with $y\%$ probability. To implement this, there is a second draw to obtain a new number z. The student wins if $z \leq y$.

To receive a prize, every week a group of students was selected at random through a raffle based on the last digits of their national ID and the last digits of the main prize of the regional lotto. Winners were selected based on proximity of these two numbers until the total of prizes awarded reached COP 300,000 (about US\$100) per week.

3 Conceptual framework and econometric strategy

Let $\hat{b_{jt}}$ and $\hat{b_{kt}}$ be the posterior probabilities assigned to quartiles j and k, both above or below the median. Similarly, $\hat{b_{jt-1}}$ and $\hat{b_{kt-1}}$ are the prior probabilities assigned to the same quartiles. According to Bayes' formula, the posterior probability after receiving an above-median signal ($S_i = 1$) is:

$$\mathbb{P}[Q_j | S_i = 1] = \hat{b_{jt}} = \frac{\hat{b_{jt-1}}}{\hat{b_{jt-1}} + \hat{b_{kt-1}}} \tag{1}$$

where $t-1$ and t correspond to the beliefs before and after receiving the signal in any given round of practice tests. Note that if the signal is above the median ($S_i = 1$), j and k are the top two quartiles of the distribution. The priors concerning the bottom quartiles

[9] Because of lower engagement of students outside the institute when they have to follow these instructions, the mechanism was only applied to the quartile to which the score belongs, and not to each quartile as in Berlin and Dargnies (2016).

disappear from the expression because being in either of the top two quartiles is an event with probability one.

To find an expression that could form the basis of an econometric specification, I take the ratio of posterior beliefs:

$$\frac{\hat{b_{jt}}}{\hat{b_{kt}}} = \frac{\hat{b_{jt-1}}}{\hat{b_{kt-1}}} \tag{2}$$

Adding indicators for the signal that students receive (would have received, in the case of students who only see their absolute scores) and a contemporaneous error term, the econometric specification becomes:

$$\frac{\hat{b_{jt}}}{\hat{b_{kt}}} = \beta_0 \, \mathbb{I}\{S_i = 0\} \times \frac{\hat{b_{jt-1}}}{\hat{b_{kt-1}}} + \beta_1 \, \mathbb{I}\{S_i = 1\} \times \frac{\hat{b_{jt-1}}}{\hat{b_{kt-1}}} + \varepsilon_{it} \tag{3}$$

Bayesian updating corresponds to $\beta_0 = \beta_1 = 1$.[10] Conservatism would imply that $\beta_0 < 1$ and $\beta_1 < 1$, and asymmetry that $\beta_0 \neq \beta_1$.

An important clarification is that the Bayesian benchmark derived in equations 1-3 is defined conditional on the information explicitly modeled in the experiment: prior beliefs over quartiles and the relative performance signal. Although students also observe their absolute score at the posterior stage, incorporating this continuous signal into a structural Bayesian benchmark would require assumptions about students' knowledge of the score distribution, which varies across practice tests and is not observed by students ex ante. The benchmark is therefore intended to capture the correct updating response to the additional relative performance signal, holding constant the information contained in absolute scores, which is common to treatment and control groups.

[10] The reader may recognize that this specification is not identical to other papers in the literature. The main difference is that I cannot write the posterior and prior ratios as logit functions given that I elicit beliefs about quartiles and not about being above or below the median as in previous papers. The other difference is that my signal is truthful so the likelihood ratio is equal to one in equation 2.

3.1 Outcomes of interest

The main outcome in equation 3 is the ratio of posteriors. Note that when one of the indicators for above/below the median equals one, the other is zero by definition. This expression is estimated using OLS pooling treated and control students together.[11] The regression excludes a constant term because the value of the posterior ratio when the prior ratio equals zero has no meaning in the Bayesian model. The standard errors are clustered at the individual level.

Besides the posterior ratio, I study the fraction of tokens allocated to the section of the distribution (above or below the median) where they should be assigning probabilities according to Bayes' formula. This outcome prevents us excluding from the analysis students who have a degenerate prior and receive a signal that is contrary to it. For example, they assign all tokens to the quartiles below the median and receive an above-median signal. In this case, it is not possible to calculate the posterior ratio, but it would be possible to see whether they understand the signal by looking at how they assign tokens in the posterior stage. Another outcome measures whether there is any update at all by creating an indicator equal to one if the posteriors are equal to the priors.

The econometric specification for the fraction of tokens assigned to above-median quartiles is as follows:

$$y_t = \alpha_0 + \alpha_1 \mathbb{I}\{S_i = 1\} + \alpha_2 Treated_i + \alpha_3 \mathbb{I}\{S_i = 1\} \times Treated_i + \alpha_4 y_{t-1} + \varepsilon_{it} \tag{4}$$

where y_t and y_{t-1} are the fraction of tokens (out of 12) allocated to above-median quartiles at the posterior and prior stages, respectively, within a practice test round. Hence, the interpretation of the α coefficients is the additional probability assigned to above-median quartiles, after accounting for the probability assigned in the prior stage. α_0 and α_1 capture the fraction of additional tokens assigned by students not receiving the signal whose score is

[11] Even though control students do not see the signal, it is possible to generate indicators for both groups given that whether a student is above or below the median is known in the dataset.

below and above the median, respectively. α_2 and α_3 measure the effect of receiving the signal on students whose score is below and above the median, respectively. This specification is estimated by OLS separately by test subject. The specification for the outcome measuring any update eliminates the y_{t-1} as it is already part of the outcome measure.

4 Results

4.1 Bayesian updating

I first present results estimating equation 3 for all students regardless of whether they receive the signal, as well as an augmented model estimating separate coefficients for those who receive and do not receive the signal. Table 1 presents the results from both models for math and reading separately, along with p-values of statistical tests of the Bayesian model.

In both exam subjects, students update as conservative Bayesians. In column 1 of Table 1, we see that the ratio of prior beliefs in math predicts about 0.59 and 0.76 of the posterior ratio when the math score is below and above the median, respectively. In the case of reading (column 3), the point estimates are much closer at 0.65 and 0.67 for below and above the median, respectively. In both cases, the hypotheses that the coefficients are equal to one are rejected at the 1% level, and given the magnitude of the coefficients, the results indicate conservatism. These results are well within the range of the findings from lab experiments, which oscillate between around 0.25 and 0.80.[12]

I do not find statistical evidence of asymmetry when testing for $\beta_0 = \beta_1$. In reading, both point estimates are very close to each other. In math, it seems that students update more closely to the Bayesian benchmark when their score is above the median (β_1 is larger than β_0), but the p-value does not allow us to conclude that the two coefficients differ from

[12] A slope below one in the posterior–prior relationship may reflect behavioral underreaction to information, but it can also arise mechanically from noise or discretization in elicited beliefs. The coefficients are therefore best interpreted as reduced-form measures of compression relative to the Bayesian benchmark induced by the signal.

Table 1: Belief updating

	Math		Reading	
	(1)	(2)	(3)	(4)
	All	Augmented	All	Augmented
β_0	0.594***	0.630***	0.648***	0.721***
	(0.077)	(0.098)	(0.050)	(0.072)
β_1	0.760***	0.807***	0.672***	0.626***
	(0.083)	(0.084)	(0.068)	(0.090)
τ_0		0.512***		0.570***
		(0.110)		(0.075)
τ_1		0.719***		0.717***
		(0.135)		(0.097)
p-value β_0=1	0.0000	0.0002	0.0000	0.0001
p-value β_1=1	0.0041	0.0224	0.0000	0.0000
p-value β_0=β_1	0.1441	0.1754	0.7729	0.4096
p-value τ_0=1		0.0000		0.0000
p-value τ_1=1		0.0388		0.0038
p-value τ_0=τ_1		0.2343		0.2436
Observations	704	704	734	734
No. students	328	328	342	342

Notes: The table presents estimates of equation 3 regressing the posterior ratio on the prior ratio, and reports separate coefficients for students with below-median scores (β_0 and τ_0) and above-median scores (β_1 and τ_1). Columns 1 and 3 pool students who receive and who do not receive the signal, while columns 2 and 3 separate these two types of students. The p-values reported at the bottom test whether the predictions of the Bayesian model hold: students update like a Bayesian with the same priors (β_0 = β_1=1 and τ_0 = τ_1=1), and their response to information is symmetric when their score is below or above the median (β_0 = β_1 and τ_0 = τ_1). Standard errors clustered at the student level. *** p<0.01, **p<0.05, * p<0.1.

each other. The finding of no asymmetry is in line with Coutts (2019), (Grossman & Owens, 2012), Buser et al. (2018) and Schwardmann and Van der Weele (2019) but contradicts the finding in the seminal paper by Möbius et al. (2022) that subjects update more closely to the Bayesian benchmark when receiving an above-median signal.

In the augmented model, the coefficients τ_0 and τ_1 are for treated students receiving a below- and above-median signal, respectively, while the β coefficients are for control students who do not receive the signal but whose scores are above or below the median. There is suggestive evidence that treated students update more similarly to a Bayesian when their score is above the median than when it is below, but again, the difference is not statistically significant. Updating by treated students is not substantially different from control students. Overall, the Bayesian model is rejected by the data. There is strong evidence of conservatism in both math and reading, and no statistical evidence of asymmetric updating.[13]

4.2 The value of a signal for updating

Even though Bayesian updating is rejected, it may still be the case that the signal is informative for students to assess where they are in the distribution of scores. I first analyze whether probability allocation varies by whether students receive the signal or not. I then focus on for which students the signal is most useful for updating. Given that providing absolute scores is the most common practice in educational settings, understanding the value of additional (relative) performance information in belief updating may provide useful insights for policymakers and institutions.

Table 2 shows estimates of specification 4, where the variable $Treated$ equals one if the student receives the relative performance signal, and *Above median* indicates that the score in the practice test is above the median. Columns 1 and 3 present results for the fraction of tokens assigned to above-median quartiles, while Columns 2 and 4 present estimates for the variable indicating no updating. In math, students with below-median scores who do *not* receive the signal assign 17% additional tokens to above-median quartiles in the posterior relative to the prior stage (Constant in regression). The coefficient on the treated variable indicates that students who receive the signal and have a score below the median do not allocate fewer tokens to above-median quartiles than the control group. It is as if students with poor performance in math do not understand or disregard the signal.[14] The dynamics

[13]

[14]My design does not allow us to differentiate between motivated beliefs, lack of understanding, or disregard of the signal. Processing information in a biased-optimistic way can have a series of benefits as

in reading are slightly different as students with poor performance who receive the signal (column 3 of Table 2) assign 8 pp less probability to above-median quartiles.

Even without receiving the signal, absolute scores are informative about one's position in the score distribution. Students with above-median scores who do *not* receive the signal allocate 13 pp and 11 pp more tokens to above-median quartiles than students with scores below the median in math and reading, respectively. Nevertheless, the signal is more informative than absolute scores only, given that the interaction coefficient indicates that students who receive the signal that their score is above the median allocate 11 additional pp to above-median quartiles in math and 16 pp in reading.

The regressions of columns 2 and 4 aim to capture the fraction of updates with an outcome indicating whether there is no updating (posteriors are equal to priors). In previous lab experiments, it has been found that substantial fractions of subjects do not update at all: 42% (Möbius et al., 2022), 36% (Coutts, 2019), and 25% (Buser et al., 2018). In my experiment, I find that across all rounds, about 16% and 18% of students do not update in math and in reading, respectively. There are no significant differences between treated and control groups, even though there is a positive coefficient of 10 pp in the likelihood of not updating for students above the median receiving the signal in reading, which is not statistically significant. Recall that posterior elicitation takes place a few days after prior elicitation, so the role of memory could be less prominent than in lab experiments, where updating is done within a single experimental session.

Taken together, my results highlight that students do update beliefs about their relative performance, albeit more so when the signal is positive than when it is negative. The value of the relative performance signal for students with scores above the median, i.e., the incremental effect of treatment for the above-median group relative to the below-median

individuals are more motivated to work and improve performance (Compte & Postlewaite, 2004). However, it can also backfire if it becomes a self-trap blinding individuals and inducing them to make highly inefficient choices (Bénabou & Tirole, 2016). Several models have incorporated the possibility that individuals may use optimistic beliefs as motivation (Bénabou & Tirole, 2002), directly derive utility from maintaining a good image about themselves (Kőszegi, 2006; Brunnermeier & Parker, 2005), and use optimism as a social signal (Burks, Carpenter, Goette, & Rustichini, 2013).

Table 2: Posterior probability allocation to quartiles above median and no updating

	Math		Reading	
	(1)	(2)	(3)	(4)
	Fraction above	No update	Fraction above	No update
Treated	-0.037	0.034	-0.082**	-0.021
	(0.035)	(0.058)	(0.036)	(0.050)
Above median	0.128***	0.057	0.109***	-0.022
	(0.027)	(0.045)	(0.023)	(0.043)
Treated × Above med.	0.107**	-0.030	0.161***	0.101
	(0.043)	(0.069)	(0.043)	(0.063)
Fraction above (prior)	0.553***		0.508***	
	(0.043)		(0.042)	
Constant	0.169***	0.158***	0.215***	0.185***
	(0.028)	(0.037)	(0.031)	(0.036)
Observations	862	862	904	904
No. students	369	369	386	386

Notes: The table presents estimates of equation 4. In columns 1 and 3, I regress the fraction of tokens (probability) assigned to above-median quartiles at the posterior stage on an indicator of whether the score is above the median, a treatment dummy, the interaction between the two, and the fraction of tokens assigned to above-median quartiles at the prior stage. In columns 2 and 4, I regress a variable indicating that the token assignment is exactly the same at the prior and the posterior stages on the treatment and above-median indicators. Standard errors clustered at the student level. *** p<0.01, **p<0.05, * p<0.1.

group, is estimated to be between 11 and 16 pp depending on the subject.

4.3 The value of a signal across the score distribution

Next, I examine who are the students whose updating benefits most from receiving the signal. The effect of the signal is likely to be nonlinear across the distribution of scores. For example, students with a very high or very low score could easily guess their quartile. However, students in the middle of the score distribution may have a harder time inferring this information from the absolute scores only. To test this hypothesis, I modify equation 4 to incorporate indicators of score ranges from 2 to 10 instead of the above median indicator.[15]

Figure 1 plots separate estimates for treated and control students in math (panel (a)) and reading (panel (b)). The y-axis shows the additional fraction of tokens assigned to above-median quartiles in the posterior stage relative to the prior stage. Point estimates are given by the marker and 83.4% confidence bars are plotted to allow for statistical comparisons across groups.[16]

(a) Math (b) Reading

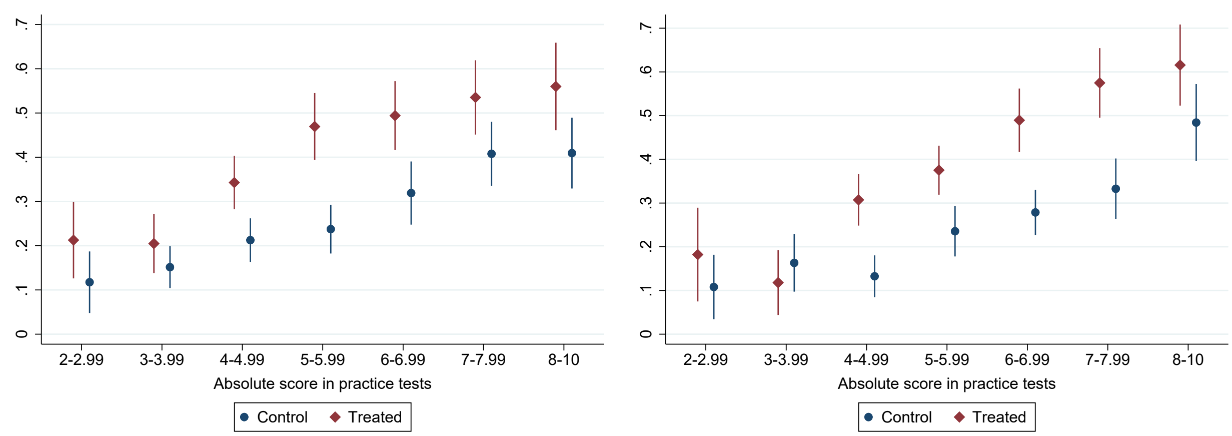


Figure 1: Fraction of tokens assigned to above median by practice test score

Notes: The figure shows the effect of receiving the relative performance signal on above-median token assignment at the posterior stage by practice test score interval. Panel (a) shows the effects for math and Panel (b) for reading. Standard errors clustered at the student level. Coefficients and 83.4% confidence bars obtained from specification 4.

Along the score distribution for both subjects, the point estimates for students receiving the signal tend to be above those of the control group. The differences are, however, only significant around the middle of the distribution, that is, for scores between 4 and 7 in math and 4 and 8 in reading. This is consistent with the signal being most helpful for students whose uncertainty is largest because it is not clear whether their scores would lie in the top

[15] Absolute scores range continuously from 0 to 10. I construct indicators of score ranges such as from 2 to 3 points, 3 to 4 points, etc. Very few students obtain scores below 2, so the graphs omit that category.

[16] 95% confidence intervals would allow us to determine whether the point estimate differs from zero but the goal here is to make comparisons across groups.

or bottom halves of the distribution. For example, a score of 6 in math seems far from the maximum score of 10. However, Appendix Table B.3 shows that a score of 6 is always in the top half of the distribution. Students can learn that they are indeed in the top half with the signal despite the fact that a score of 6 seems rather low given the scale of possible values. In reading (panel (b) of Figure 1), the signal also has value for scores between 7 and 8 points. In this case, higher scores do not always correlate with being above the median as there is more dispersion in scores than in math (see Appendix Table B.3).[17]

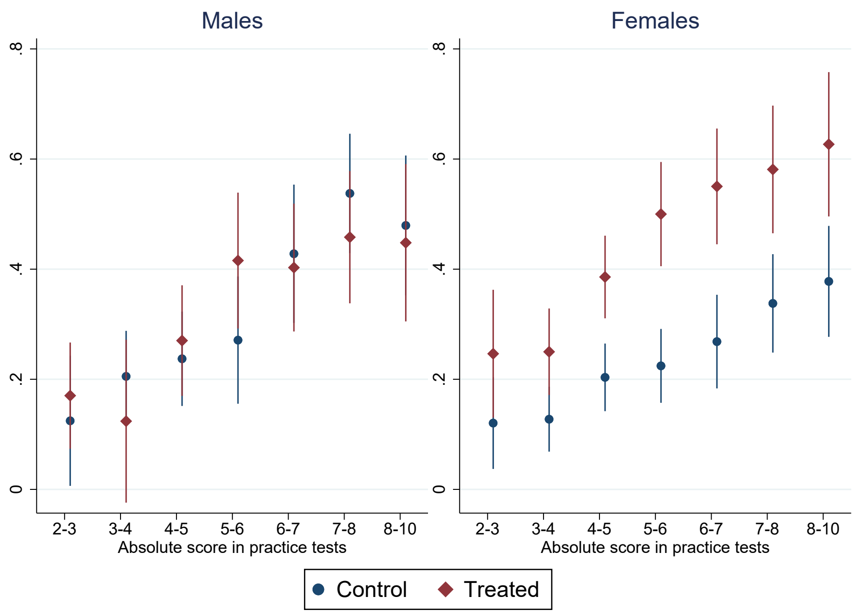
4.4 Heterogeneity by gender

A large body of literature documents gender differences in self-confidence. In the lab, lower levels of confidence among women than among men have been found to be an important driver of decisions such as whether to enter a competition (Niederle & Vesterlund, 2007). Moreover, lower levels of confidence and competitiveness among women are associated with real-world outcomes, such as the gender gap in education choices (Buser et al., 2017) and in labor market earnings (Reuben et al., 2024).

Part of how confidence affects life outcomes may be related to how men and women incorporate informational signals. Previous work in information processing documents heterogeneous responses to signals by gender. The evidence is mixed. While Möbius et al. (2022) and Buser et al. (2018) find that women are more conservative relative to a Bayesian than men, Coutts (2019) does not find gender differences. Ertac (2011) and Berlin and Dargnies (2016) find that women's responses to signals are stronger than men's. Greater conservatism among women implies that they end up with less confidence after receiving a similar set of informational signals than men. Thus, it is relevant to document gender differences in updating in a real-life setting.

[17] The reader may be wondering if this is capturing mistakes in the probability assignment. For example, if students allocate more probability to above-median quartiles when they should not be assigning more probability to those quartiles, it would look like the signal is helping them know better their relative performance when the opposite is true. Appendix Figure A.3 shows that this is not the case. Students with scores below the median who receive the signal are no more likely than control students to assign tokens to

(a) Math



(b) Reading

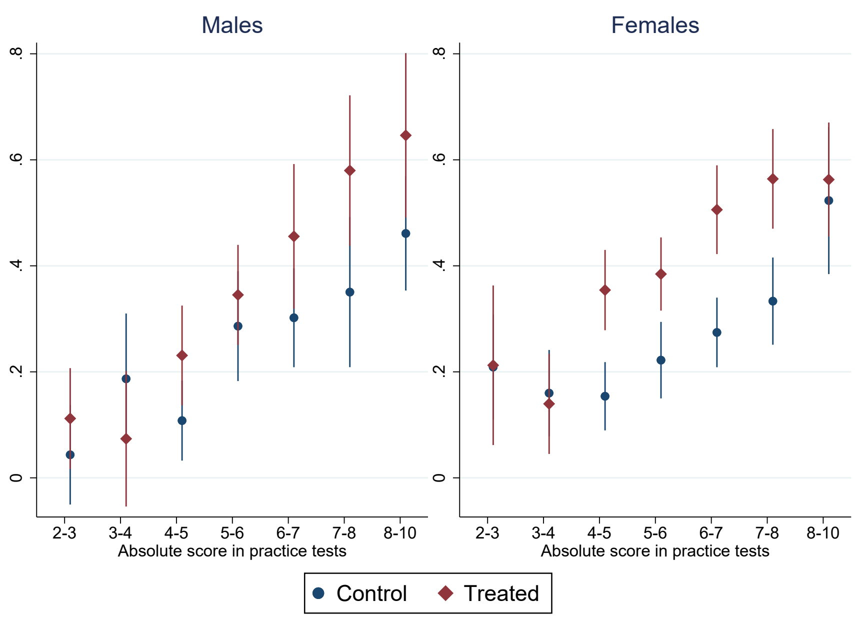


Figure 2: Fraction of tokens assigned to above median by gender

Notes: The figure shows the effect of receiving the relative performance signal on above-median token assignment at the posterior stage by practice test score interval and by gender. Panel (a) shows the effects for math and panel (b) for reading. Standard errors are clustered at the student level. Coefficients and 83.4% confidence bars obtained from specification 4 are estimated by gender. Difference-in-differences (DID) point estimates pooling all scores are in Appendix Table B.4 and by score interval in Appendix Figure A.5.

above-median quartiles.

Figure 2 reports results from specification 4 by gender, and suggests that the value of the signal is higher for female students in both, math and reading. In contrast, the point estimates for men receiving and not receiving the signal are close to each other and not statistically different. Graphically, it seems that receiving the signal helps female students with scores in the middle and upper parts of the distribution make a more accurate assessment of their relative ability. However, difference-in-differences (DID) coefficients point toward relatively large effects across the score distribution, but power is not enough to statistically differentiate most of them from zero (see Appendix Figure A.5). Taken at face value, this finding suggests that the signal helps women build their confidence in their relative performance by updating to a larger extent relative to men.

To further understand why women tend to be more responsive to the signal, I explore whether men and women differ in their priors and how they update in the posterior stage. Appendix Figure A.6 plots how far students are from the optimal allocation of tokens. For example, if a student's score is above the median, she should assign all tokens to above-median quartiles. If she does, the bias in probability assignment is zero (red line in figure panels). A positive value means that she assigns more tokens to above-median quartiles than she should (overconfidence), and a negative value means that she assigns fewer tokens to above-median quartiles than she should (underconfidence). In general, students are overconfident when their scores are low and underconfident when they are high. Women, in particular for reading, start out more overconfident in the lower part of the distribution and more underconfident than men at absolute scores around 5 and higher at the prior stage. Hence, there is more room for them to update their beliefs, which could explain the higher value of the signal for women. In addition, it has been found that women are more sensitive to social signals leading to higher variability in women's behavior relative to men's (Croson & Gneezy, 2009). So, it may well be the case that they are also more sensitive to the informational signals provided in my experiment.

4.5 Effects outside the updating task

Finally, from a policy perspective, it is of interest to know what traits outside of the experiment may be affected by having a more accurate assessment of one's relative academic performance. For top academic performance, it is often the case that besides mastering concepts and study materials, confidence in own abilities plays an important role as it helps develop non-cognitive skills for academic and non-academic attainment (Heckman & Rubinstein, 2001; Valentine et al., 2004; Murphy & Weinhardt, 2020). A better assessment of one's relative performance may also affect perceptions about the difficulty of a practice test, effort and performance itself. For example, a large update of beliefs may change effort by pushing students to be at the top or avoid being at the bottom (Gill et al., 2019).

The updating variable used in this section measures the signed change in belief miscalibration between the prior and posterior stages in the previous' week practice test. For each student, I compute the distance between the token allocation and the optimal allocation implied by the student's realized score, and define updating as the posterior distance minus the prior distance. A positive value therefore reflects a correction from underconfident priors (beliefs move upward toward the optimal allocation), while a negative value reflects a correction from overconfident priors. The magnitude of the variable captures how strongly beliefs adjust toward the optimal allocation, while the sign captures the direction of the initial miscalibration. Zero indicates no updating. This variable is then standardized to ease interpretation.[18]

Table 3 shows the triple and double interactions of a fully saturated model including female and treatment indicators, as well as the standardized t-1 updating variable. The coefficients displayed in the table indicate the additional effect on the outcomes in the col-

[18]The mean update is 0.0006 for math and -0.004 for reading. One standard deviation amounts, respectively, to an update of 0.27 and 0.26 of a total possible update of +1 or -1, which would be the case when all tokens were assigned to the wrong quartiles at the prior stage and all moved to the correct quartiles at the posterior stage. The sign depends on whether the update took place from over- or underconfident prior beliefs. For example, if a student assigns half of the prior tokens and 90% of the posterior tokens to the section of the distribution where her score lies, the size of her update is 0.4.

Table 3: Effects on education outcomes

	(1) Confidence	(2) Score in PT	(3) Study hours	(4) Difficult
Panel A: Updating in math (t-1)				
Treated × Female × Updating	0.198**	0.312	0.240	-0.219**
	(0.096)	(0.325)	(0.830)	(0.087)
Female × Updating	-0.119*	-0.117	-0.558	0.137**
	(0.070)	(0.253)	(0.611)	(0.060)
Treated × Updating	-0.147*	-0.555**	0.055	0.185***
	(0.077)	(0.265)	(0.515)	(0.071)
Constant	0.636***	5.099***	5.944***	0.470***
	(0.075)	(0.239)	(0.758)	(0.068)
Observations	531	597	533	619
No. students	253	280	254	288
Panel B: Updating in reading (t-1)				
Treated × Female × Updating	0.133	-0.186	0.055	-0.037
	(0.096)	(0.179)	(0.766)	(0.067)
Female × Updating	-0.105	-0.010	-0.158	0.009
	(0.069)	(0.134)	(0.514)	(0.054)
Treated × Updating	-0.069	-0.099	-0.187	0.030
	(0.079)	(0.127)	(0.432)	(0.054)
Constant	0.654***	6.364***	4.655***	0.226***
	(0.072)	(0.164)	(0.602)	(0.064)
Observations	547	615	549	636
No. students	260	288	261	295

Notes: Each column presents the interaction coefficients from a triple DID regression of the outcomes in the column headings on a fully saturated model including indicators for female, treatment status, and a standardized version of the size of the update in the previous practice test (t-1) in math (panel A) and reading (panel B). All regressions add practice test fixed effects. The independent variables measure whether the students feel confident or very confident about gaining admission, their score in the following week's practice test (from 0 to 10), self-reported study hours over the following week, and whether they perceived the following practice test to be difficult or very difficult. The score, study hours and difficulty measures refer to the math section of the practice test in panel A and to the reading section in panel B. Standard errors clustered at the student level. *** p<0.01, **p<0.05, * p<0.1. Because these regressions use data from two consecutive beliefs surveys asking retrospective questions such as study hours over the past week, the sample of students is not the same as in previous tables. Appendix Table B.5 compares the characteristics of students in either sample and finds that the two samples do not differ substantially.

umn headings of updating by one standard deviation above the mean update.[19] The triple

[19] These regressions include students who respond to two consecutive beliefs surveys because the outcomes are either measured retrospectively or in between the prior and posterior elicitations. The sample size is smaller as a consequence, but there are no substantial differences between the students in the main sample

interaction in panel A indicates that there is a positive and statistically significant effect in confidence among female students who receive the relative performance signal and have a large belief update regarding their last week's math practice test performance. The point estimate indicates that these female students experience a confidence boost of about 20 pp relative to a base of 63% of students reporting feeling confident or very confident about gaining admission to the university for which they are preparing. Although positive and relatively large, the triple interaction is not statistically significant in the case of updating beliefs in reading (panel B).

There is no effect on the score in the following practice test, or the number of study hours over the next week after the updating takes place. A large update in math seemingly correlates with the perceived difficulty in the next practice test for treated female students, However, this seems to be a reversal of the positive effects found in the double interactions, where the difficulty is perceived to be higher.

While the size of the update is endogenously determined, I find evidence that receiving the signal induces a larger update in math and reading, especially for students whose score is above the median (see Appendix Table B.6). Hence, my result could be understood as the signal inducing a larger update of beliefs which in turn increases confidence in the following practice test.[20]

Overall, my results indicate that updating beliefs in math, the subject that females are thought to feel less confident with, is important for confidence outside the experiment. Females not only improve confidence in assessing their ability within the task in the experiment as shown in previous subsections, but large belief updates also allow them to gain confidence in an important dimension that matters outside the experiment, such as confidence in gaining admission to their intended university.

and in this restricted sample (see Appendix Table B.5).

[20] An instrumental variable analysis would be ideal to establish the causality. Unfortunately, my data is not well positioned for this type of analysis.

5 Conclusion

I study the relative value of different academic performance signals that students receive during their preparation for an important college entrance exam. Students in my sample take weekly practice tests in math and reading and receive absolute scores in both subjects. I elicit prior and posterior beliefs about relative performance in each practice test and induce experimental variation to provide an above-/below-median signal to a randomly selected subgroup of students.

I identify that the signal is most valuable for students with scores in the middle of the distribution, for whom it is harder to assess whether they are above or below the median, and provide suggestive evidence that females' updating is more responsive to the relative performance signal. Large belief updates help build confidence among females regarding their perceived likelihood of gaining a college slot.

To my knowledge, this is the first evidence using the tools and methods in prior laboratory experiments in a real-life, high-stakes setting where beliefs about one's own ability matter for a series of outcomes such as effort, performance in an important test, and college major choices. Lab experiments have focused on eliciting beliefs and studying updating within a constrained task that may not be meaningful for participants once their experimental session is over. I find that students do not differ substantially from the behavior documented in previous lab findings given that they update in a conservative way when compared with the Bayesian benchmark.

The main insight of my paper is that belief updating can help some students build confidence about their likelihood of academic success. In many educational settings, students receive information about their absolute scores in academic evaluations, but rarely receive information about relative performance. I show that top performing students generally have a good understanding of their relative performance from absolute scores, and this knowledge may have already helped develop their confidence. I also show that students who are not performing at the top but rather around the middle range of scores may have a hard time inferring relative performance, which is key in some settings such as college entrance exams or when the distribution of scores is compressed. If learning that one's score is in the top half is interpreted as good news, students who without the signal would not know this information may start building confidence that they can perform well. My findings show that students with scores above the median who receive the signal have larger belief updates, and those with larger belief updates see an increase in their level of confidence, which could be a crucial ingredient for best performance in a high-stakes exam.

One limitation of my study is that the signals and the degree of updating are not randomly assigned but depend on the students' performance and prior beliefs. Also, updating could be correlated with individual observables and unobservables. These features are shared by previous lab experiments studying information processing about ego-relevant information (Benjamin, 2019). Also, observables may not bee too important as Wiswall and Zafar (2015) have found. My paper stresses that besides correcting belief misalignments, addressing informational gaps in education can develop important non-congitive skills such as confidence in academic performance.

References

Bénabou, R., & Tirole, J. (2002). Self-confidence and personal motivation. *The Quarterly Journal of Economics*, *117*(3), 871–915.

Bénabou, R., & Tirole, J. (2016). Mindful economics: The production, consumption, and value of beliefs. *Journal of Economic Perspectives*, *30*(3), 141–64.

Benjamin, D. J. (2019). Errors in probabilistic reasoning and judgment biases. *Handbook of Behavioral Economics: Applications and Foundations 1*, *2*, 69–186.

Berlin, N., & Dargnies, M.-P. (2016). Gender differences in reactions to feedback and willingness to compete. *Journal of Economic Behavior & Organization*, *130*, 320–336.

Bobba, M., & Frisancho, V. (2022). Self-perceptions about academic achievement: Evidence from mexico city. *Journal of Econometrics*, *231*(1), 58–73.

Brunnermeier, M. K., & Parker, J. A. (2005). Optimal expectations. *American Economic Review*, *95*(4), 1092–1118.

Burks, S. V., Carpenter, J. P., Goette, L., & Rustichini, A. (2013). Overconfidence and social signalling. *The Review of Economic Studies*, *80*(3), 949–983.

Buser, T., Gerhards, L., & Van Der Weele, J. (2018). Responsiveness to feedback as a personal trait. *Journal of Risk and Uncertainty*, *56*(2), 165–192.

Buser, T., Peter, N., & Wolter, S. C. (2017). Gender, competitiveness, and study choices in high school: Evidence from switzerland. *American Economic Review*, *107*(5), 125–30.

Camerer, C. (2011). The promise and success of lab-field generalizability in experimental economics: A critical reply to levitt and list. *Available at SSRN 1977749*.

Compte, O., & Postlewaite, A. (2004). Confidence-enhanced performance. *American Economic Review*, *94*(5), 1536–1557.

Coutts, A. (2019). Good news and bad news are still news: Experimental evidence on belief updating. *Experimental Economics*, *22*(2), 369–395.

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

Eil, D., & Rao, J. M. (2011). The good news-bad news effect: asymmetric processing of objective information about yourself. *American Economic Journal: Microeconomics*, *3*(2), 114–38.

Ertac, S. (2011). Does self-relevance affect information processing? experimental evidence on the response to performance and non-performance feedback. *Journal of Economic Behavior & Organization*, *80*(3), 532–545.

Franco, C. (2020). How does relative performance feedback affect beliefs and academic decisions?

Gill, D., Kissová, Z., Lee, J., & Prowse, V. (2019). First-place loving and last-place loathing: How rank in the distribution of performance affects effort provision. *Management Science*, *65*(2), 494–507.

Gonzalez, N. (2017). *How learning about one's ability affects educational investments: Evidence from the advanced placement program* (Tech. Rep.). Mathematica Policy Research.

Grossman, Z., & Owens, D. (2012). An unlucky feeling: Overconfidence and noisy feedback. *Journal of Economic Behavior & Organization*, *84*(2), 510–524.

Heckman, J. J., & Rubinstein, Y. (2001). The importance of noncognitive skills: Lessons from the ged testing program. *American Economic Review*, *91*(2), 145–149.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and brain sciences*, *33*(2-3), 61–83.

Kessler, J., & Vesterlund, L. (2015). *The external validity of laboratory experiments: The misleading emphasis on quantitative effects* (Vol. 18). Oxford University Press Oxford, UK.

Köszegi, B. (2006). Ego utility, overconfidence, and task choice. *Journal of the European Economic Association*, *4*(4), 673–707.

Levitt, S. D., & List, J. A. (2008). Homo economicus evolves. *Science*, *319*(5865), 909–910.

Möbius, M. M., Niederle, M., Niehaus, P., & Rosenblat, T. S. (2022). Managing self-confidence: Theory and experimental evidence. *Management Science*, *68*(11), 7793–7817.

Murphy, R., & Weinhardt, F. (2020). Top of the class: The importance of ordinal rank. *The Review of Economic Studies*, *87*(6), 2777–2826.

Niederle, M., & Vesterlund, L. (2007). Do women shy away from competition? do men compete too much? *The Quarterly Journal of Economics*, *122*(3), 1067–1101.

Reuben, E., Sapienza, P., & Zingales, L. (2024). Overconfidence and preferences for competition. *The Journal of Finance*, *79*(2), 1087–1121.

Schwardmann, P., & Van der Weele, J. (2019). Deception and self-deception. *Nature Human Behaviour*, *3*(10), 1055–1061.

Stinebrickner, R., & Stinebrickner, T. (2008). The effect of credit constraints on the college drop-out decision: A direct approach using a new panel study. *American Economic Review*, *98*(5), 2163–84.

Valentine, J. C., DuBois, D. L., & Cooper, H. (2004). The relation between self-beliefs and academic achievement: A meta-analytic review. *Educational psychologist*, *39*(2), 111–133.

Wiswall, M., & Zafar, B. (2015). How do college students respond to public information about earnings? *Journal of Human Capital*, *9*(2), 117–169.

Zafar, B. (2011). How do college students form expectations? *Journal of Labor Economics*, *29*(2), 301–348.

Zimmermann, F. (2020). The dynamics of motivated beliefs. *American Economic Review*, *110*(2), 337–61.

A Additional Figures

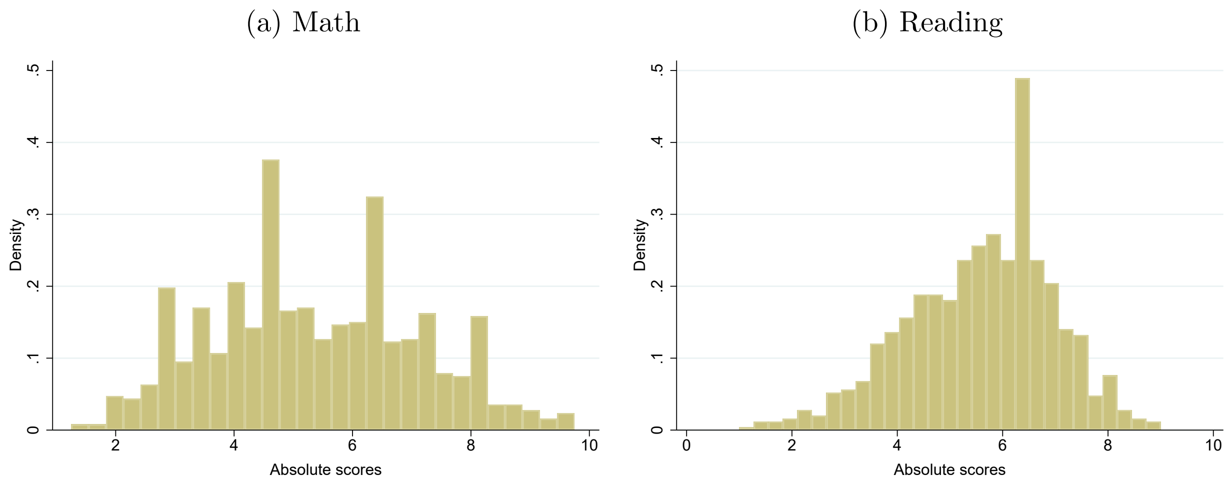


Figure A.1: Density of absolute scores

Notes: Densities of absolute scores using practice tests across all rounds for students in the main sample.

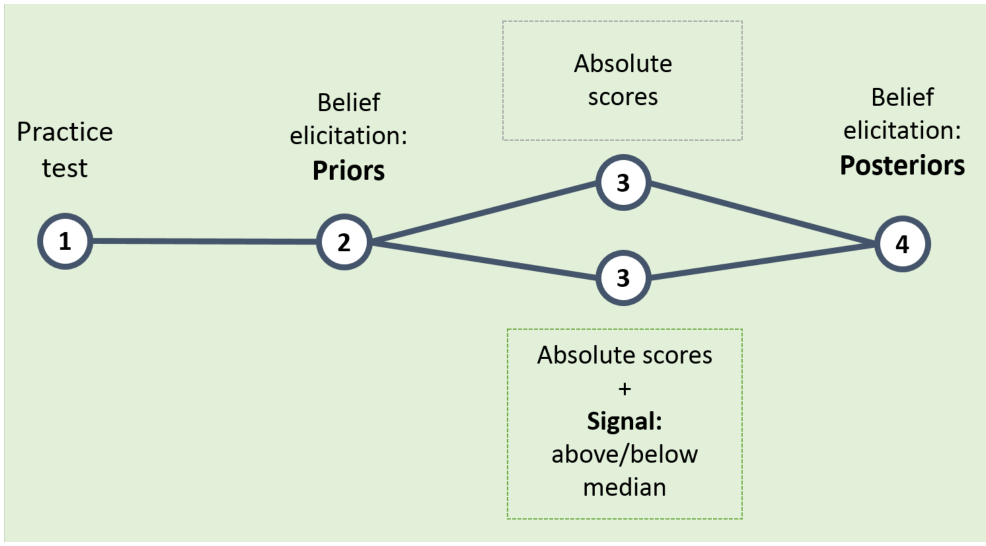


Figure A.2: Prior and posterior belief elicitation

Notes: This chart shows the stages that students follow in a given week. A practice test takes place every Monday, immediately after which prior beliefs are elicited. A few days later students receive a performance report that contains absolute scores only or absolute scores and a relative performance signal, depending on their treatment assignment. After receiving this information, they report posterior beliefs. The treatment was assigned at the beginning of the college entrance exam course and did not change across rounds.

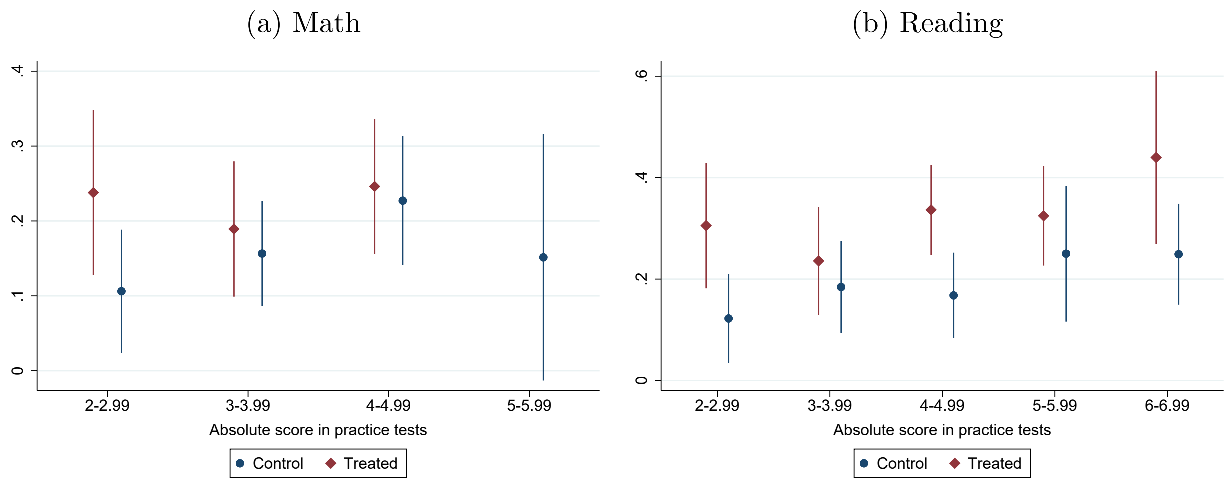


Figure A.3: Probability allocation to above median when score is below median

Notes: The figure shows the effect of receiving the relative performance signal on above-median token assignment at the posterior stage by practice test score interval when students' true score is below the median. Panel (a) shows the effects for math and Panel (b) for reading. Standard errors clustered at the student level. Coefficients and 83.4% confidence bars obtained from specification 4.

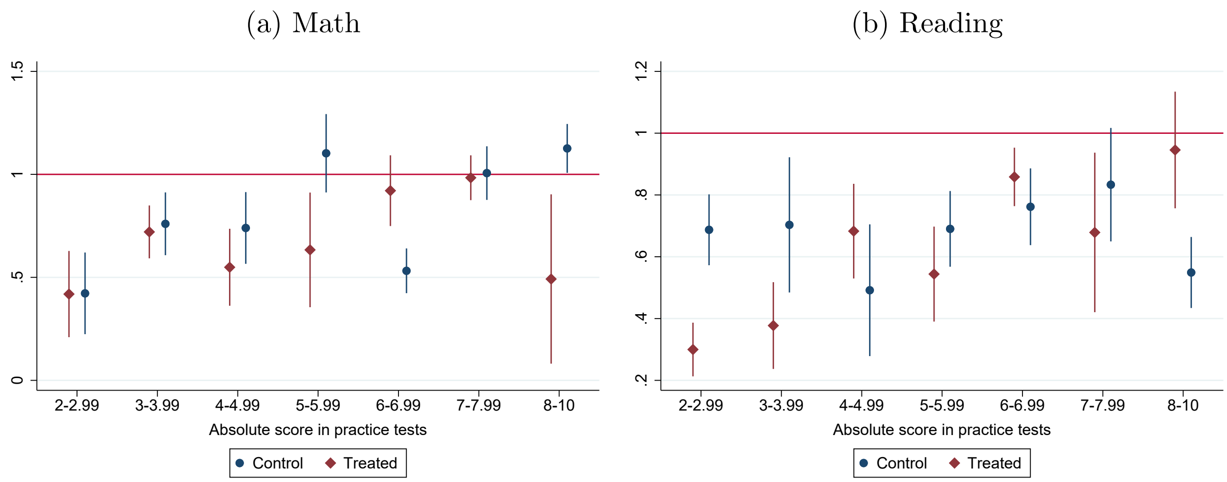


Figure A.4: Additional value of the above-/below- median signal for Bayesian updating

Notes: Each point estimate is the coefficient of the posterior ratio on the prior ratio from specification 3 estimated for each score interval. Error bars reflect 83.4% confidence intervals. Panel (a) shows the effects for math and Panel (b) for reading. Standard errors clustered at the student level.

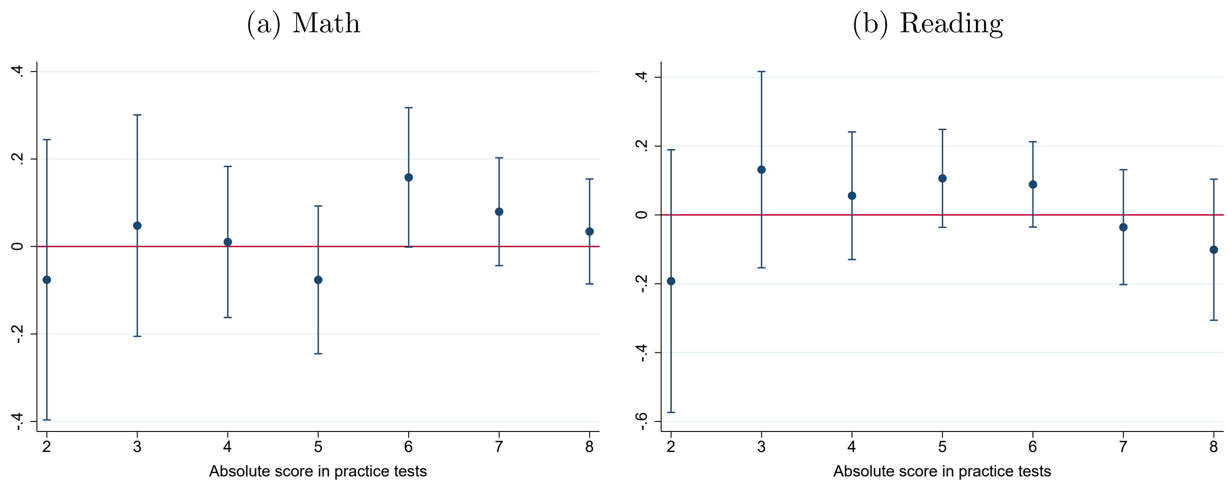
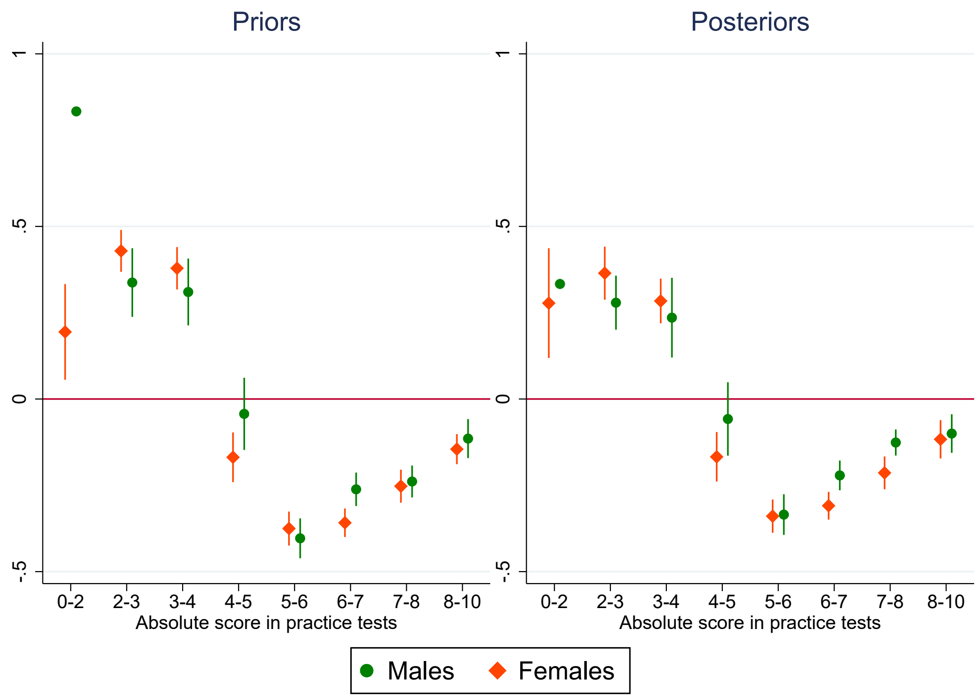


Figure A.5: Differential effects by gender (DiD coefficients)

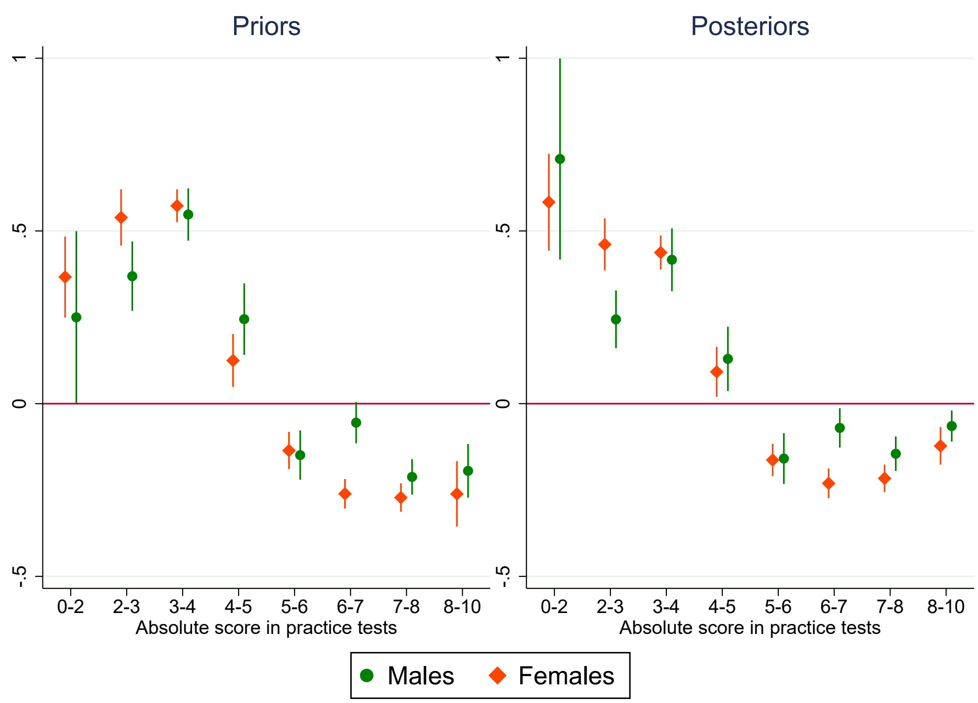
Notes: Coefficients and 95% confidence intervals from a difference-in-differences specification testing differential effects for females relative to males. Panel (a) shows the effects for math and Panel (b) for reading. Standard errors clustered at the student level.

Figure A.6: Accuracy in prior and posterior beliefs by gender

(a) Math



(b) Reading



Notes: The variable plotted is how far students are from allocating tokens to the correct half of the distribution based on their score. For example, if a student's score is above the median, she should assign all tokens to above-median quartiles. If she, however, assigns only half of tokens to above-median quartiles, the variable is -0.5 (she is underconfident). If she assigns all tokens to above-median quartiles, the variable is 0. If her score is below the median and she does not assign all tokens to below median quartiles, the variable would have a positive value (overconfidence). The figures show that students tend to be overconfident when obtaining low scores and underconfident when obtaining middle-range and high scores. Coefficients and 83.4% confidence intervals from specification 4.

B Additional Tables

Table B.1: Balance of characteristics by treatment assignment

Variable	(1) Control	(2) Treated	(3) Difference
Female	0.64	0.62	-0.03
	(0.48)	(0.49)	(0.06)
Took exam before	0.79	0.83	0.04
	(0.41)	(0.37)	(0.05)
Age	17.65	17.23	-0.42**
	(2.19)	(1.27)	(0.20)
Marital status: single	0.98	0.99	0.00
	(0.13)	(0.12)	(0.01)
Main activity: student	0.70	0.75	0.05
	(0.46)	(0.43)	(0.06)
Disabled	0.00	0.02	0.02
	(0.07)	(0.14)	(0.01)
Underrepresented minority	0.12	0.08	-0.04
	(0.32)	(0.27)	(0.03)
Lives in urban area	0.87	0.88	0.01
	(0.34)	(0.32)	(0.04)
Residential strata	2.50	2.56	0.06
	(0.99)	(1.03)	(0.12)
Math score in initial PT	3.23	3.28	0.05
	(1.10)	(1.00)	(0.14)
Reading score in initial PT	5.13	5.32	0.19
	(1.53)	(1.68)	(0.20)
Avg. classroom score in initial PT	37.64	37.75	0.10
	(2.65)	(2.60)	(0.33)
Observations	470	434	904

Notes: Means and standard errors of the variables listed on the left by treatment status in the main sample used in the reading regressions. The difference in characteristics between treatment and control is in column 3. PT stands for practice test. Standard errors clustered at the student level. *** p<0.01, **p<0.05, * p<0.1.

Table B.2: Attrition across rounds

	(1) 1 vs. more	(2) Up to 2 vs. more
Treated	0.039	-0.034
	(0.031)	(0.047)
Female	-0.023	-0.056
	(0.032)	(0.050)
Took exam before	-0.007	-0.064
	(0.041)	(0.065)
Age	0.007	0.004
	(0.010)	(0.014)
Marital status: single	-0.040	0.094
	(0.143)	(0.155)
Main activity: student	0.048	0.050
	(0.037)	(0.055)
Disabled	0.217	0.285
	(0.180)	(0.202)
Underrepresented minority	0.086	0.218**
	(0.058)	(0.096)
Lives in urban area	0.085**	0.050
	(0.038)	(0.069)
Residential strata	0.024	0.003
	(0.017)	(0.024)
Math score in initial PT	0.007	0.001
	(0.017)	(0.026)
Reading score in initial PT	-0.009	-0.018
	(0.011)	(0.018)
Avg. classroom score in initial PT	-0.003	-0.006
	(0.006)	(0.010)
Constant	0.048	0.469
	(0.343)	(0.499)
Observations	880	880
No. students	374	374

Notes: This table show regression coefficients of the outcomes in the columns on baseline characteristics to analyze whether students participating in one survey only (column 1) or up to two surveys (column 2) differ from students participating in more surveys. PT stands for practice test. Standard errors clustered at the student level. *** $p<0.01$, **$p<0.05$, * $p<0.1$.

Table B.3: Above-median classification by score category

Score interval	**Math**	**Reading**
[0-2)	0.00	0.00
[2-3)	0.04	0.00
[3-4)	0.14	0.00
[4-5)	0.64	0.35
[5-6)	0.93	0.77
[6-7)	1.00	0.87
[7-8)	1.00	0.99
[8-10]	1.00	1.00

Notes: This table shows the fraction of students with above-median scores for each score category across the eight rounds of practice tests. For example, scores between 4 and 5 points would be classified as above median about 64% of the time in math and 35% in reading.

Table B.4: DID: Posterior probability allocation to quartiles above median and no updating

	Math		Reading	
	(1)	(2)	(3)	(4)
	Fraction above	No update	Fraction above	No update
Treated	-0.030	0.054	-0.097	-0.081
	(0.055)	(0.115)	(0.065)	(0.084)
Above median	0.197***	0.065	0.142***	-0.044
	(0.047)	(0.080)	(0.045)	(0.064)
Treated × Above med.	0.064	-0.006	0.153**	0.190*
	(0.068)	(0.132)	(0.077)	(0.101)
Female	0.032	-0.046	0.012	-0.106
	(0.044)	(0.079)	(0.041)	(0.076)
Treated × Female	-0.014	-0.022	0.027	0.096
	(0.070)	(0.133)	(0.076)	(0.105)
Above median × Female	-0.104*	-0.011	-0.050	0.041
	(0.056)	(0.097)	(0.051)	(0.085)
Treated × Above med. × Female	0.069	-0.059	0.008	-0.143
	(0.087)	(0.154)	(0.090)	(0.129)
Fraction above (prior)	0.541***		0.504***	
	(0.044)		(0.043)	
Constant	0.154***	0.188***	0.210***	0.250***
	(0.039)	(0.065)	(0.042)	(0.061)
Observations	862	862	904	904
No. students	369	369	386	386

Notes: The table presents estimates of equation 4. In columns 1 and 3, I regress the fraction of tokens (probability) assigned to above-median quartiles at the posterior stage on and indicator for whether the score is above the median, a treatment dummy, the interaction between the two, and the fraction of tokens assigned to above-median quartiles at the prior stage. In columns 2 and 4, I regress a variable indicating that the token assignment is exactly the same at the prior and the posterior stages on the treatment and above-median indicators. Standard errors clustered at the student level. *** p<0.01, **p<0.05, * p<0.1.

Table B.5: Comparison of characteristics between restricted and unrestricted samples

Variable	(1) Unrestricted	(2) Restricted	(3) Difference
Treated	0.48	0.49	0.01
	(0.50)	(0.50)	(0.04)
Female	0.62	0.66	0.04
	(0.49)	(0.47)	(0.04)
Took exam before	0.80	0.84	0.04
	(0.40)	(0.37)	(0.03)
Age	17.47	17.39	-0.08
	(1.95)	(1.33)	(0.13)
Marital status: single	0.99	0.98	-0.00
	(0.12)	(0.13)	(0.01)
Main activity: student	0.73	0.72	-0.01
	(0.45)	(0.45)	(0.04)
Disabled	0.01	0.00	-0.01*
	(0.12)	(0.07)	(0.01)
Underrepresented minority	0.11	0.06	-0.05**
	(0.32)	(0.23)	(0.02)
Lives in urban area	0.88	0.86	-0.02
	(0.32)	(0.35)	(0.03)
Residential strata	2.54	2.51	-0.03
	(1.02)	(0.97)	(0.08)
Math score in initial PT	3.23	3.30	0.07
	(1.03)	(1.12)	(0.10)
Reading score in initial PT	5.19	5.32	0.13
	(1.59)	(1.66)	(0.14)
Avg. classroom score in initial PT	37.66	37.79	0.13
	(2.62)	(2.64)	(0.24)
Observations	678	226	11,341

Notes: The unrestricted sample contains all students who completed at least one of the belief elicitation rounds (reported priors and posteriors about performance in a practice test). The restricted sample contains students who answer two subsequent beliefs surveys. The reason for this is that some of the outcomes are asked retrospectively (e.g., study hours in the past week), or are asked in between prior and posterior belief elicitation (e.g., confidence in gaining admission) and hence do not account for the potential belief updating from receiving information in the performance report. PT stands for practice test. Standard errors clustered at the student level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table B.6: Effect of the signal on the size of the update

	Math		Reading	
	(1)	(2)	(3)	(4)
	Main	Restricted	Main	Restricted
Treated	-0.088	-0.525	-0.224	-0.659**
	(0.146)	(0.333)	(0.154)	(0.323)
Above median	0.244**	0.168	0.279***	0.238
	(0.097)	(0.154)	(0.086)	(0.197)
Treated × Above med.	0.298*	0.678*	0.429**	0.782**
	(0.175)	(0.378)	(0.173)	(0.346)
Constant	-0.232***	-0.169	-0.222***	-0.109
	(0.079)	(0.127)	(0.070)	(0.174)
Observations	862	220	903	226
No. students	369	123	386	127

Notes: This table shows regression on the size of the update (standardized variable) in math and reading on treatment status and an above-median indicator. Columns 1 and 3 present results for the main sample and columns 2 and 4 for the restricted sample (students who complete beliefs surveys two weeks in a row). Standard errors clustered at the student level. *** p<0.01, **p<0.05, * p<0.1.

C Experimental protocol

C.1 Lab-in-the-field belief elicitation

Students take a weekly practice test that can be administered on paper or online. Immediately after the practice test, they receive a paper or online survey with the questions below. A few days later, when they receive the performance report, they answer the belief elicitation (the Quartiles game) once again. As explained in the main text, this is intended to understand how students form posterior beliefs given the information provided: absolute scores in the case of control students, and above-/below-median signal in the case of treated students.

In case of questions, students had access to a FAQ sheet that they could access from the website containing the survey. They were also provided with a phone number and email so they could ask a specific question.

To determine how much cash they would win if they were selected as one of the weekly winners, a computerized dice was thrown in the last page of the performance report. See below for the instructions students had to follow.

Beliefs survey instructions

The following questions are related to today's practice test. Based on your answers, you may win one of the cash prizes. The winners will be chosen based on the distance between the last four digits of their ID and the Loteria de la Cruz Roja jackpot that plays next week. The IDs that are closest to the jackpot will receive cash prizes until 300,000 pesos are awarded. Recall that you must complete all surveys to enter the raffle for six laptops. Winners will be contacted by email.

Q1: How many questions in the practice test do you think you will answer or have answered correctly? For every correct guess, you will receive 5,000 pesos if you are selected in the draw.

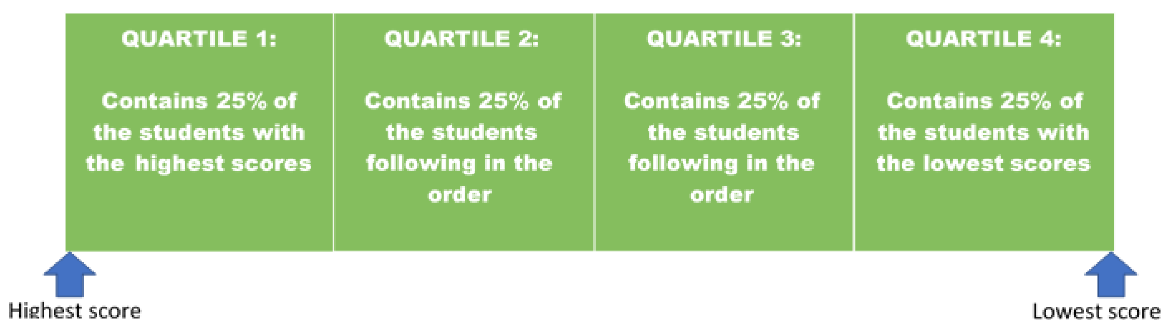
Mathematical logical reasoning: ______ correct out of 40 questions

Reading competency: ______ correct out of 40 questions

Q2: The Quartiles Game

Imagine that you enter a casino and are given 24 tokens to play. You choose to bet them in the "Quartiles Game."

The bet you will make is regarding how you think you performed in the practice test this week. To make your bet, the dealer explains to you that there are four sections in the betting table called "quartiles." Each quartile contains a subgroup of the scores of students who took the practice test ordered from highest to lowest. This is how the table looks like:



Your bet consists in assigning the 24 tokens to the quartiles in the two sections of the practice test (12 tokens to mathematical and logical reasoning and 12 tokens to reading).

Tip: Assigning more tokens to the quartile(s) in which you truly believe your score lies will maximize your chances of winning one of the prizes. Remember that no one besides you will see your answers.

For each section of the practice test, please assign 12 tokens to the quartiles. If you think it is unlikely that your score will be in one or more of the quartiles, please assign zero tokens to those quartiles. Make sure that the sum of your allocations is equal to 12 in each column:

	Mathematical Logical Reasoning	**Reading Competency**
Bet to quartile 1 (group with highest scores)	__ tokens	__ tokens
Bet to quartile 2	__ tokens	__ tokens
Bet to quartile 3	__ tokens	__ tokens
Bet to quartile 4 (group with lowest scores)	__ tokens	__ tokens
Sum of tokens	__ tokens	__ tokens

To determine if you win your bet, the dealer will look at the quartile your score is in and roll a 12-sided die:

If the result of the roll of the dice is equal to or below your bet in the quartile your score is in, you will win 20,000 pesos in case you are selected in the draw. For example, if you bet eight tokens and the result of the dice roll is 6, you will be eligible for a prize.

If the result of the roll of the dice is above your bet, the dealer will roll the dice again. You will win 20,000 pesos if the new result is below the result of the first dice roll and you are selected in the draw. For example, if the result of the first dice roll is 6 and the result of the second is 3.

Q3: Please rate the level of difficulty of each section of the practice test, where 1 is extremely easy and 5 extremely hard.

Q4: Approximately, how many hours did you study last week for each section of the practice test? Include the time you dedicated to solving practice questions and reviewing materials. Do NOT include class time and homework. Choose zero if you did not study for that section of the test last week.

Options: integers from 0 to 8+

Q5: Based on how you feel today, how confident are you that you will gain admission to Universidad de Antioquia?